Evaluating the Research Quality of Education Journals in China: Implications for Increasing Global Impact in Peripheral Countries

Juanjuan Zhao
University of Cincinnati

Gulbahar H. Beckett
Iowa State University

Lihshing Leigh Wang
University of Cincinnati

There has been a rapid growth of academic research and publishing in non-Western countries. However, academic journal articles in these peripheral countries suffer from low citation impact and limited global recognition. This critical review systematically analyzed 1,096 education research journal articles that were published in China in a 10-year span using a multistage stratified cluster and random sampling method and a validated rubric for assessing research quality. Our findings reveal that the vast majority of the articles lacked rigor, with insufficient or nonsystematic literature reviews, incomplete descriptions of research design, and inadequately grounded recommendations for translating research into practice. Acknowledging the differences in publishing cultures in the center-periphery divide, we argue that education research publications in non-Western countries should try to meet Western publishing standards in order to participate in global knowledge production and research vitality. Implications for emerging countries that strive to transform their research scholarship are discussed.

KEYWORDS: research quality, education research, periphery scholarship, content analysis, China

Academic research and publishing have become increasingly global due to academic and economic globalization (Feng, Beckett, & Huang, 2013), as evidenced by increasing participation of developing and non-Anglophone countries in

academic research article submission and output (Hyland, 2015). According to Hyland (2015), although submissions from Japan and the United States increased 127% and 177%, respectively, during 2005 to 2010, submissions from emerging countries experienced much higher growth rates (e.g., India, 443%; China 484%; and Iran and Malaysia, 800%) during the same period. These increases are the result of financial investment in research and scholarship by emerging countries as they aspire to "export" their knowledge to Western countries (Feng et al., 2013; Guilford, 2013) and move up their universities' national and global rankings (Hyland, 2015). Such investment is related to the fact that counties such as China no longer seem to be content with unidirectional learning from Western countries by "importing" knowledge produced there (Beckett & Zhao, 2016; Feng et al., 2013).

However, despite such a rapid growth in submission and a strong desire for participation in global publishing, journal article output from Western countries still dominates the stage, and the dramatic increase in submission from peripheral countries has not resulted in corresponding global recognition or citation impact (Li, 2004; Wang, Wang, & Weldon, 2007). According to Hyland (2015), of the 1,865 journals that were indexed in the 2007 Social Sciences Citation Index, 79.62% (*n* = 1,585) originated in the United States and the United Kingdom, whereas many emerging countries were underrepresented or even experienced decreased indexing rates. There has also been little change in the frequency of citation rates for publications from these emerging countries.

Bridging the Center-Periphery Divide in Knowledge Globalization

Some scholars have explained the above phenomenon from the perspective of a center and periphery dichotomy, accusing the center of prejudice against peripheral scholars and their academic publications and of failing to take the responsibility to improve the center-periphery dialogue, as well as for its economic superiority that provided it an edge over the periphery (Aydinli & Mathews, 2000; Canagarajah, 2002). In contrast with industrialized, mainstream, center ones, periphery stands for developing, nonindustrialized regions and countries, or emerging research centers (Canagarajah, 2002; Salager-Meyer, 2015). Scholarly work produced from the periphery is often seen as disadvantaged due to its limited resources and access to up-to-date technologies and literature, unsatisfactory research environments, and poor work conditions that result from a weaker market economy and industrialization (Altbach, 2002). Among the many disadvantages to the recognition and citation of peripheral scholarship, as Salager-Meyer (2015) argued, are language barriers, as scholars in the center might not read Chinese, Farsi, Russian, or Spanish and therefore are less likely to cite works published in those languages. In summary, the unequal participation and underrepresentation of developing countries in global knowledge construction is often simply associated with the center-periphery divide, which is framed as the poorly equipped periphery being victimized by the prestige of well-resourced and trained center.

Rather than furthering the center-periphery dichotomy by targeting at the centrality of developed countries as the sources of peripherality of academic publications in developing countries, our study is based on the concept and an increasing need to build a bridge between the central and peripheral scholarship and to help

diversify and democratize knowledge production. As Aydinli and Mathews (2000) pointed out, the geographical boundary and limitations to the peripheral scholarship have long been recognized, but they have not become a focus of the current efforts to bridge the gap. The current center and periphery dichotomy, if lacking such an effort, will continue to isolate researchers from both sides and discourage knowledge production and dissemination from the peripheral countries to the center stage (Matsuda, 2013).

The Case of Academic Publishing in Mainland China[1]

Since 1978, when China embarked on its economic reforms, the government has made various efforts to push the internationalization of its academic research and scholarly publication. There were four *Decisions* (national policies that were issued by the Chinese government) on research development in 1985, 1995, 1999, and 2006 that emphasized the learning of science and technology from other countries (Liu, 2008). Understanding the gap in academic research between China and the leading countries, China established the National Science Funding Committee in 1986, introduced Social Sciences Citation Index as an evaluation indicator for related fields in 1987, and began to rank and select core journals in each field from 1992 in an effort to increase both the quality and quantity of its academic research (Feng et al., 2013). Led by the Ministry of Education (2014), universities in China established its own social science citation index, the Chinese Social Sciences Citation Index, in 1997 to evaluate and rank journals in the social sciences, including education (Tam & Chen, 2010).

From the beginning of the 21st century, the Chinese government has focused on the globalization of its academic research in the humanities and the social sciences, especially in terms of promoting its academic achievements and gaining international recognition for Chinese academic research in these fields. According to Guilford (2013), from 2008 to 2012, "China's state research and development spending has grown around 18% each year," exceeding $163 billion in 2012. The country "aims to boost R&D spending to 2.5% of GDP by 2020." Such investment is a result of China's emerging discontent with unidirectional learning, that is, with simply importing knowledge from Western countries, and China's aspiration to contribute to global knowledge (Beckett & Zhao, 2016; Feng et al., 2013). The recent Decision that was issued in 2011 began to emphasize expansion beyond domestic journals, encouraging journal editors to introduce their journals to the international research community and to promote Chinese culture and research achievements (Feng et al., 2013). This new policy shifted the country's focus from import-oriented learning to import-export sharing (Feng et al., 2013).

To better meet the internationalization goal, some Chinese journals have adopted the citation formats and publishing standards of English language journals to make it easier for their voices and knowledge to be heard and spread (Shi, Wang, & Xu, 2005). Western-trained scholars in China have also been bringing in English academic writing conventions to help change Chinese scholars' informal writing style from relying on experience and context to using empirical evidence and thus enhance the internationalization of academic writing and publications in education research (Shi, 2002). According to Sun (2011), since the beginning of the 21st century, a shift has taken place in China from valuing literature

interpretation to valuing empirical research. In applied linguistics research, for instance, it was found that journal articles were moving from summing-up experiences to the rise of methodological awareness, that is, from a nonempirical direction toward an empirical direction in the final decade of the 20th century (Gao, Li, & Lü, 2001).

However, academic publications and primarily journal articles that emanated from mainland China have had low citation rates and little global recognition and visibility (Moiwo & Tao, 2013). Arunachalam (2008) found that China was ranked the lowest in terms of the number of citations per paper among 12 countries including the United States, the United Kingdom, Germany, Japan, France, and Canada. A report that was published by the Chinese Academy of Science showed similar findings, with 80% of the publications by China's scholars falling into either the no-citation or low-citation categories (Ju & Wang, 2008). Moreover, although research produced by U.S. scholars accounted for 30% of all citations, research produced by Chinese scholars accounted for only 4% of the citations (Ware & Mabe, 2012), revealing dramatic discrepancies between mainland China's investment in and production of scholarship and their actual impact measured by citation rates (Hyland, 2015).

Problems in Evaluating Education Research Quality in China

Although the standards for conducting and evaluating education research have changed over time and vary across cultures and paradigms (Atkinson, 1999), quality education research shares some fundamental principles, such as a thorough understanding of the existing literature (Swales, 2004), a coherent and explicit chain of reasoning, and a detailed description of procedures and analysis that allows others to critique, analyze, or replicate studies (National Research Council, 2002). From the perspective of policymakers, Gutiérrez and Penuel (2014) reasoned that making research relevant to practice was another key criterion of rigor for education research. Equitable and consequential research focuses on persistent problems of practice, examining their context of development with attention to ecological resources and constraints, including why, how, and under what conditions programs and policies work. Overall, quality education research should produce realizable and valid knowledge that benefits educational practice and society as a whole.

Academic publications from peripheral communities have been criticized for their poor research quality according to the above standards (Salager-Meyer, 2008; Yang, 2005). Research quality in this study is defined as all of the major components reported in a journal article that are indicative of its research quality, including literature review, problem identification, purpose formulation, design selection, data collection, data analysis, and results interpretation. In the case of China, the few studies that examined the research quality of its educational journal articles support the problems identified above, but most of these are outdated. Zhao et al. (2008) compared articles that were published between 2003 and 2004 in the *American Educational Research Journal* (*AERJ*) in the United States and *Jiao Yu Yan Jiu* (*JYYJ*; Educational Research) in mainland China, both of which are considered to be top education research journals in their countries. One of the significant differences was that *AERJ* published more empirical studies than *JYYJ*, with the latter dominated by conceptual reports that were generated from

personal reflections or theoretical interpretations. The few empirical studies that were published in *JYYJ* had many flaws, such as a lack of systematic and comprehensive literature reviews, absence of a standard research report format, inadequacy in research context description, a deficiency of a prescribed methodology for data collection and analysis, and no explanations of study limitations. Similar findings were reported by Chen (1994), Fan (2000), Jiang (2004), Yang (2006), and Zhang and Lu (2008).

Major knowledge gaps still exist that have prevented the education research field in China from transforming to one that values the scholarly and rigorous pursuit of scientific knowledge. The above-cited reviews of Chinese journal articles were limited in their breadth and the depth of the methodological issues that affected research quality. Most of the critiques focused on the identification of primary research approaches (e.g., empirical and nonempirical) and the scholarship of cited references. In fact, none of the reviews cited above conducted a systematic analysis of journal articles that covered all of the aspects of a research study from problem identification to interpretation of results. Similarly, most of the studies reviewed a limited number of selected journals rather than a large sample of representative journals. More important, these reviews are outdated because they do not reflect the policy movement to reform research quality instituted in 2002. As a result, a comprehensive evaluation of the research quality of published education research journal articles in China in recent years is lacking.

Research Purpose and Significance

To address the above knowledge gaps, we developed an education research quality evaluation rubric based on international publication requirements to examine the characteristics of research quality of education journal articles in China and to identify the weak areas that may have contributed to their limited recognition and citation. This rubric is comprehensive enough to be applicable to all types of research but also flexible in that not all criteria need to be used when evaluating a certain type of research. Instead of focusing on a few selected journals, we examined a wide range of core journals with high impact factors to maximize the generalizability of this study. Specifically, we aimed to achieve the following objectives in this study: (a) develop a valid and reliable rubric for evaluating education research quality that is closely aligned with the internationally established criteria for evaluating education research, (b) examine the research quality of education journal articles in China using the developed evaluation rubric, (c) identify the strengths and weaknesses of published education research journals in China in light of the international standards, (d) study the longitudinal trend of education research quality change over a 10-year period from 2002 to 2011, and (e) draw implications from the current study's findings to inform other peripheral countries in their effort to achieve global recognition.

We used nondirectional hypotheses in addressing the second and third objectives because no current research that exists provides parameters we can use to compare the statistics to, although we did expect that the overall quality found as a result of this review would be low. The lack of existing literature to inform directional hypotheses also adds to the significance of the current study, because in our exhaustive search, we were unable to find an evaluation rubric similar to the one

we developed for this study. Even though China was selected as a representative case study because of the sharp contrast between the country's internationalization efforts in academic research and its low global visibility, this evaluation rubric can potentially be used for other countries as well, center or peripheral. The comprehensive and flexible design of the evaluation rubric makes it a valuable tool for all types of research across different disciplines, paradigms, and ontologies. Findings from the current study can contribute to promoting international scholarship as well as informing academic research globalization policy and practice.

Method

Research Design

This study used quantitative content analysis to evaluate research quality of educational journal articles in mainland China. Quantitative content analysis is a research technique for a systematic, objective, and quantitative description of a content that allows analysts to make inferences about the characteristics and meanings of the material (Government Accountability Office, 1989; Neuendorf, 2002). It was chosen for the study because of its advantages in producing summary results of a large volume of data using explicit coding instructions, extensive reliability checks, and inferential statistical analysis (Government Accountability Office, 1989). The procedure of quantitative content analysis in the study followed the steps described in Neuendorf (2002): (a) deciding content to be examined (i.e., educational journal articles in China), (b) choosing and defining the variables to be examined (i.e., evaluation criteria for education research), (c) sampling the material to be analyzed (i.e., multistage stratified cluster and random sampling), (d) developing standardized categories and coding rubric (see the appendix), (e) training raters and conducting preliminary rater reliability check, (f) coding the material with final rater reliability check, and (g) analyzing and reporting the results.

Sampling

The articles that were selected for review in this study came from 63 key education journals that were published during the period from 2002 to 2011 in China, excluding Hong Kong and Macau. The selection of key journals was based on the sixth edition of *A Guide to Chinese Key Journals* published by Peking University, a nationally recognized index of key journals across disciplines. The accessible journal population was indexed by the China Academic Journals Full-Text Database, the Chinese Social Sciences Citation Index Database, and Wanfang Data.

The articles were selected through a multistage stratified random cluster and simple random sampling. Stratified sampling was used to provide a greater precision of population estimates and a better representation of the population being studied (Cochran, 1977). The first stage was stratified random cluster sampling, with *journal* and *year* as strata. We then randomly selected one, two, or three issues as a cluster from each journal-by-year stratification cell. The number of issues was proportional to the publication frequency of each journal: For journals that publish quarterly or bimonthly, one issue was sampled; for journals that publish monthly, two issues were sampled; and for journals that publish two or three times a month, three issues were sampled. Among the 73 education journals, there were 3 quarterly and 16 bimonthly journals, 43 monthly journals, and 7 semimonthly and 4

TABLE 1

Journal article sampling results

Publishing frequency	No. of journals	Sampling weight	No. of articles sampled
Quarterly	3	10	30
Bimonthly	16	10	160
Monthly	43	20	860
Semimonthly	7	30	210
Trimonthly	4	30	120
Total target sample	73		1,380
Total final sample	63		1,096

Note. The difference between the "Total target sample" and the "Total final sample" is due to the unavailability of some randomly selected journals or articles.

trimonthly journals. This number of journals yielded 1,380 ($19 \times 10 + 43 \times 20 + 11 \times 30$) articles in total for review. Due to the unavailability of some of the issues and journals, 1,096 articles from 63 journals were ultimately collected (see Table 1). The second stage was a stratified simple random sampling with *journal issue* as the stratum, and we randomly selected one article from each issue.

Previous studies have suggested that article characteristics and methodological approaches tend to be stable within a 5-year span (Goodwin & Goodwin, 1985; Hutchinson & Lovell, 2004). We focused our analysis on more than 1,000 articles over a 10-year span rather than a shorter term or a single period in time because we wanted to detect developments or changes, if any, in the characteristics of education research quality in Chinese publications longitudinally.

Instrumentation

Identification of Established Evaluation Criteria

After an exhaustive search, we were unable to find an existing tool for coding research quality of published articles that meets our purpose. Therefore, we developed our own coding form for the purpose of conducting a content analysis of the articles in our sample. We first determined the criteria for evaluating research quality by acknowledging the fact that different types of research may require different evaluation criteria. Among the many taxonometric schemes for classifying research types, the most useful one was an empirical versus a nonempirical study because the requirements were different along this division line for the obvious reason of a data-driven versus a theory-driven approach to solving a research problem. Acknowledging the debates on the incompatibility of positivistic and postpositivistic ontologies (Guba & Lincoln, 1994; Howe, 2009), we embodied only the common themes across different epistemological frameworks. Our goal was to develop an evaluation rubric that is both comprehensive enough to be applicable to different types of research and flexible enough to be modifiable to different types of research.

Empirical research. Empirical work is defined as studies with systematic collection and analysis of quantitative or qualitative data (Gao et al., 2001; Hutchinson

& Lovell, 2004). The development of evaluation criteria for empirical research journal articles was based on the standards for reporting on empirical social science research by the American Educational Research Association (AERA, 2006), scientific articles guidelines by the European Association of Science Editors (EASE, 2013), reporting standards for research in psychology by the American Psychological Association (APA, 2008), research design for qualitative and quantitative approaches (Creswell, 1994), a checklist of review criteria for research manuscripts (Gastel, 2002), scientific research in education (National Research Council, 2002), a guide to the organization of a research report (Fraenkel & Wallen, 2009), criteria for international journal paper evaluation (Nunn & Adamson, 2007), an instrument for evaluating experimental education research reports (Suydam,1968), a qualitative research checklist (Schostak, 2008), a literature review rubric by Boote and Beile (2005), a rubric for evaluating a psychology research report (Gottfried, Vosmik, & Johnson, 2008), and the elements of a proposal (Pajares, 2007). As this content analysis is a comprehensive review of all of the major factors that are associated with research quality, we included all of the categories in the main body of a journal article, specifically, introduction, literature review, method, results, and discussion, based on the above guidelines. An explanation of the coding categories and their definitions is provided in the subsequent sections.

Introduction: The main goal of this section is to "provide a clear statement of the purpose and scope of the study" (AERA, 2006, p. 34) by describing the problem and research questions (EASE, 2013; Gastel, 2002) and explaining the significance of the inquiry (APA, 2008). Research problems are issues that exist "in the literature, theory, or practice that lead to a need for the study" (Creswell, 1994, p. 50). The function of the problem formulation serves as the first step toward the goal of research to generate new claims, and more important, it justifies the demands for the attention to research (Brewer, 2005). Thus, a rationale should be provided for problem formulation "as it relates to the groups studied (especially with respect to relevant features of the historical, linguistic, social, and cultural backgrounds of the group) where questions about appropriateness of the connections may arise" (AERA, 2006, p. 34).

A journal article should also make clear the contribution of a study by indicating how the research will "refine, revise, or extend existing knowledge in the area under investigation" (Pajares, 2007). Specifically, it demonstrates its theoretical importance (i.e., testing, elaborating, or enriching theoretical perspectives or establishing new theories); states its practical importance by describing the practical concerns, including why they are important and how the current investigation can address the concerns; or suggests applicability and interest to the field (i.e., substantive contribution to the scholarly research in the field; AERA, 2006).

A study is also expected to describe its conceptual or theoretical frameworks, theories, or lines of inquiry used and to explain the rationale with relevant citations to what others have written about (AERA, 2006; Schostak, 2008); nonetheless, it may be handled differently in quantitative versus qualitative studies (Pajares, 2007). According to Creswell (1994), quantitative studies use theory deductively and place theory at the beginning of a study with the purpose to test or verify the theory. "The theory becomes a framework for the entire study, an

organizing model for the research questions or hypotheses for the data collection procedure" (pp. 87–88). In qualitative studies, a theoretical perspective or framework is a philosophical stance that researchers take that informs the methodology and thus provides a context for the process and grounds its logic and criteria (Crotty, 1998).

Literature review: The literature review was included in the Introduction by some of the evaluation standards (i.e., AERA, 2006; APA, 2008). However, as the majority of evaluation standards had it listed as a stand-alone category (i.e., Boote & Beile, 2005; Fraenkel & Wallen, 2009; Gastel, 2002; Nunn & Adamson, 2007; Pajares, 2007; Schostak, 2008), we coded it as a separate category. According to Schostak (2008), there are four types of literature reviews: a review of perspectives, a methodological review, a theoretical review, and a substantive review. Regardless of the review type, the relevance and comprehensiveness of reviewed studies with justified criteria for inclusion and exclusion are always considered to be important (Boote & Beile, 2005; Gastel, 2002; Nunn & Adamson, 2007). Each type of the review needs to identify the main debate within the field of inquiry, be it a perspective, a methodology, a theory, or a belief (Boote & Beile, 2005; Schostak, 2008). By critically analyzing the literature, researchers are expected to distinguish what has been done from what needs to be done (Boote & Beile, 2005; Gottfried et al., 2008) and to explain the relationship with previous research by describing how the current research contributes or challenges theory or knowledge from the previous research (AERA, 2006; APA, 2008). Finally, it is important to integrate and synthesize the literature reviewed and tie this synthesis into the issues that are being investigated in the current study (Gastel, 2002; Schostak, 2008).

Method: The method section includes a description and justification of research design, sampling or participants, instruments, data collection procedure, and analytical methods. The type of research design (e.g., survey, experiment, case study, ethnography) has to be described clearly with an articulation of its appropriateness to the research purpose (APA, 2008; Gastel, 2002; Gottfried et al., 2008; Suydam, 1968). Information about the sample along with justification should include target population, sampling method, sample size, and sample representativeness (APA, 2008; EASE, 2013; Gastel, 2002; Gottfried et al., 2008; Pajares, 2007; Schostak, 2008). Eligibility and exclusion criteria or special arrangements of a sample should be noted and justified (APA, 2008; Gottfried et al., 2008). Procedures for recruiting participants should also be adequately described and justified (AERA, 2006; APA, 2008; EASE, 2013; Pajares, 2007).

Descriptions of the instruments used for data collection should include their development processes (AERA, 2006), and the reliability and validity of the scores that are collected using the instrument should justify its use (Gastel, 2002; Nunn & Adamson, 2007). Data collection describes the types of data that are collected (e.g., surveys, interview, documents, records, or artifacts gathered) and the ways in which they are gathered (e.g., electrical surveys, paper copies, and data sets; AERA, 2006); outlines the procedures for collecting the data, including time and duration (AERA, 2006; Suydam, 1968); and provides the contextual information (settings and locations) of the data that are gathered (AERA, 2006; APA, 2008; Gastel, 2002).

Data analysis is judged by whether a research article sufficiently describes the analytical techniques and the procedure of analysis (AERA, 2006; Gastel, 2002; Nunn & Adamson, 2007) and if it makes clear how the analysis addresses the research questions or conforms to the research design (AERA, 2006; Gastel, 2002; Pajares, 2007; Schostak, 2008). For quantitative studies, missing data, data cleaning, and outliers handling, or changes in data analysis models, that is, information that concerns data treatment, problems with statistical assumptions or data distribution, and any considerations that arise in data collection and processing or data analysis that can affect the validity of the statistical analysis or inferences should be reported (AERA, 2006; APA, 2008).

Findings: Results should be presented effectively and in a manner that is easy to understand (Gastel, 2002). Results should be complete, with a sufficient and appropriate amount of data presented (Gastel, 2002). Researchers should describe the results or findings that are pertinent to each of the research hypotheses or questions (AERA, 2006).

Discussion: The discussion includes an interpretation of the findings and an explanation of the patterns in the data with evidence and concrete examples (AERA, 2006; APA, 2008; Gottfried et al., 2008; Nunn & Adamson, 2007; Schostak, 2008). The discussion explains how the claims and interpretation address the research problems and research questions (AERA, 2006; APA, 2008; EASE, 2013), and relates the findings or arguments to broader problems in the field by demonstrating how the conclusions connect to support, elaborate, or challenge those in previous studies (AERA, 2006; APA, 2008; Schostak, 2008). In addition, the discussion of the study's limitations should be considered (APA, 2008; EASE, 2013; Gastel, 2002; Pajares, 2007) in terms of the extent to which the results are conclusive and can be generalized (Gottfried et al., 2008) and what the unsolved problems and weakness of the current research are (APA, 2008; Fraenkel & Wallen, 2009; Pajares, 2007). A journal article concludes with implications for educational theory, research, and/or practice (AERA, 2006; Gastel, 2002) and suggestions for further research (APA, 2008; Fraenkel & Wallen, 2009; Gastel, 2002; Schostak, 2008).

Nonempirical work. Theoretical exposition, commentaries, reviews, and position papers fall into the category of nonempirical work as they are characterized with no "description of data collection or data analysis procedures" (Hutchinson & Lovell, 2004, p. 389). The evaluation criteria for nonempirical studies are based on the standards for reporting on humanities-oriented research in AERA publications (AERA, 2009), reporting standards for research in psychology (APA, 2008), the APA publication manual (APA, 2010), the standards and criteria for review articles in *Review of Educational Research* (n.d.), research articles in *Educational Researcher* (n.d.), scholarly paper review criteria from the Association for the Study of Higher Education (ASHE, n.d.), and academic writing for graduate students (Swales & Feak, 1994). A review of nonempirical study evaluation criteria reveals a strong emphasis on the introduction, literature review, and discussion sections.

Introduction: The introduction includes the purposes and the questions of the inquiry (AERA, 2009; ASHE, n.d.); the philosophical, theoretical, or practical

arguments, or the problem statement (APA, 2010); and the significance of the topic (AERA, 2009; ASHE, n.d.; *Review of Educational Research*, n.d.). According to AERA (2009) standards, an inquiry should be significant to the scholarly community by addressing issues that have been neglected in the literature, filling in gaps in current knowledge, or raising important questions about the extant knowledge. An inquiry that poses analytical questions, synthesizes divergent bodies of literature, or elaborates new theoretical or conceptual frameworks, regardless of its forms, should also make clear its scholarly contributions by demonstrating its theoretical importance, its practical importance, or its contribution to scholarly research in the field (AERA, 2009; ASHE, n.d.; *Review of Educational Research*, n.d.). In addition, "the perspective, scholarly tradition, school, and/or conceptual framework and the methods employed" (AERA, 2009, p. 484) should also be explicitly stated.

Literature review: For the literature review, the authors are expected to review all of the relevant literature on a topic (AERA, 2009; *Review of Educational Research*, n.d.), "particularly with respect to identifying its perspective and aims" (AERA, 2009, p. 485). Through the review of the literature, the main ideas, theories, perspectives, or methodologies of previous research studies should be identified (APA, 2010). The authors should also go beyond the literature interpretation by critically analyzing theories and methods—"pointing out the flaws or demonstrating the advantage of one theory over another" (APA, 2010, p.10) and distinguishing literature gaps—what has been done and what remains unresolved (APA, 2010; *Educational Researcher*, n.d.; *Review of Educational Research*, n.d.). In reviewing the literature, it is important for a current inquiry to identify relationships with previous studies by demonstrating how it joins and advances or challenges the existing literature (AERA, 2009). Last, the literature review should be well integrated into the overall research issues (AERA, 2009; Swales & Feak, 1994). According to AERA (2009), the literature review in nonempirical research does not have to be in a particular section, but it can be interwoven in the discussion.

Discussion: Similar to the criteria for empirical studies, the discussion and conclusion sections require "interpretations and portrayals of education phenomenon that are credible, persuasive, and/or effective interrogatory" (AERA, 2009, p. 485), supported by evidence, observational data, documentation, or other types of sources (AERA, 2009; ASHE, n.d.). They require the establishment of knowledge claims and arguments that pertain to the educational issue that is being studied (AERA, 2009). A nonempirical paper should also describe how the issue is conceptualized within the literature and how it is related to a larger context by drawing on the previous literature in the field to support and elaborate the conclusions or arguments (AERA, 2009; ASHE, n.d.; *Review of Educational Research*, n.d.). Findings from previous literature can also serve as objections or counterexamples to current conclusions. According to AERA (2009), authors should demonstrate a "critical self-awareness" (p. 485) of their own perspectives by acknowledging and discussing counterarguments. Moreover, implications for practical or theoretical issues (ASHE, n.d.; *Educational Researcher*, n.d.; *Review of Educational Research*, n.d.) and suggestions for unresolved questions and future directions (APA, 2010; *Educational Researcher*, n.d.) should be discussed.

Development of the Coding Form

Because all of the evaluation criteria for nonempirical studies (i.e., introduction, literature review, and discussion) were included in the criteria for empirical studies, we created a single coding form that contains all of the evaluation criteria for both types (see the appendix). The categories that are applicable to empirical studies but not to nonempirical inquiries were coded as *not applicable* (N/A) in this study. The coding form in our study consists of the following categories: introduction, literature review, method, results, and discussion. The introduction is further divided into problem, significance, and theoretical framework. The subcategories under literature review are coverage and synthesis. Method contains research design, sampling, instrumentation, data collection procedure, and data analysis. Findings are about the presentation of the results, and discussion includes results, discussion, limitations, and implications.

Determining the number of points on a rating scale is the key to the development of the coding form. To inform the development of the coding form, we examined previous studies that evaluated the research quality of education research. To identify these review forms, we conducted repeated electronic searches of major academic databases in education, including ERIC, Educational Research Complete, Education Full Text, and Educational Administration Abstracts. A dichotomous scale (i.e., *existence* vs. *nonexistence*) was used in most of the research reviewed for this study, either alone or with other continuing scales (e.g., Moyer, Finney, & Swearingen, 2002; Price et al., 2005; Randolph, Bednarik, & Myller, 2005; Randolph, Julnes, Lehman, & Sutinen, 2008). A 3-point scale was also used in many of the studies. For example, Hutchinson and Lovell (2004) used *thorough*, *minimal*, and *no* as the rating categories to indicate the extent of adequacy. Reichow, Volkmar, and Cicchetti (2008) used a three-level rubric, *high*, *acceptable*, and *unacceptable*, for primary critical indicators for research validity and a binary coding scheme for secondary indicators. Stokes and Miller (1975) employed a mix of a binary scale (*yes* or *no*), a 3-point scale (*poor*, *adequate*, *good*), and a 5-point scale (*very weak*, *weak*, *adequate*, *strong*, *very strong*) to evaluate methodological adequacy of research reports.

In the present study, we adopted a 3-point scale (2 = *strong*, 1 = *weak*, 0 = *not present*) to capture the variation in the degree to which a criterion is met. In cases in which the criterion was not relevant, it was coded as *not applicable* or N/A. For each of the criteria that were relevant to a particular study, we coded "0" if the author did not mention anything at all about any aspect of a criterion, "1" (*weak*) if an article implicitly or vaguely stated some aspects of a criterion, or "2" (*strong*) if an article explicitly and sufficiently described all aspects of a criterion.

Regarding the weighting of the various criteria, most of the studies we reviewed in the extant literature used equal weight, that is, each criterion carried the same weight in deriving a composite index of the research quality of an article (e.g., Hutchinson & Lovell, 2004; Price et al., 2005; Randolph et al., 2005; Randolph et al., 2008; Stokes & Miller, 1975). A few studies used unequal weights that were based on subjective judgment of the importance of each criterion (e.g., Miller & Wilbourne, 2002; Morley et al., 1996; Moyer et al., 2002; Reichow et al., 2008). For example, in the methodological analysis of clinical trials of treatments for

alcohol use disorders, Miller and Wilbourne (2002) weighted 12 dimensions of methodological quality from one to four based on their significance.

In evaluating the methodological quality of alcoholism treatment outcome studies, Morley et al. (1996) weighted 19 evaluation factors differently, also based on their perceived contribution to the accurate estimation of treatment effects. The assigned points for each item ranged from 0.5 to 9.0. Reichow et al. (2008) divided coding categories into primary indicators and secondary indicators. The former category included elements of research design that were deemed to be critical to demonstrate the validity of a study. The primary indicators were measured on a trichotomous ordinal scale, *high quality*, *acceptable quality*, and *unacceptable quality*, whereas the secondary indicators were rated dichotomously, *evidence* or *no evidence*. Our review indicated that the assigned weights were generally arbitrary and reflected only the perception of the importance of each criterion. Thus, we decided to go with the equal weight scheme.

Reliability Checks and Revision of Coding Rubrics

During the development of the coding form, a field trial was conducted to assess the reliability of the rubric. SPSS was used to calculate intercoder reliability (Lombard, Snyder-Duch, & Bracken, 2010). Fifteen articles were randomly selected and coded independently by two researchers. Prior to the trial, neither of the researchers discussed or exchanged thoughts about the coding categories. The main purpose of this field trial was to test how consistently each researcher interpreted the coding categories so as to revise the rubric later on. An overall kappa value of .30 was obtained, a low level of agreement. The two raters then read through the coding form together and discussed any questions about the coding categories and the discrepancies. When inconsistencies or ambiguities were found, the coding categories were modified to remedy those inconsistencies or ambiguities. For example, one coding category of data treatment (e.g., missing data, data cleaning or outliers handling) was reworded into a broader scope to evaluate both qualitative and quantitative studies.

Both of the raters also felt that ratings were difficult to assign when one category contained more than one aspect, especially as some categories included more aspects than others. Therefore, the two raters went through the whole coding form, broke down categories into more specific subcategories, and redefined them to reach consensus. Most of the coding categories were further divided into smaller and more specific ones, generating 50 subcategories from the original 34 that were tested in the first trial. For instance, under "Literature Review," there used to be one category for coverage, that is, "literature review is relevant and comprehensive with justified criteria for inclusion and exclusion from review." This category was then expanded into three categories: (a) "the literature review is relevant," (b) "the literature review is comprehensive," and (c) "the literature review has justified criteria for inclusion and exclusion from review."

Similarly, the "Problem Statement" was originally described as "poses statement of the problem or gap and supports with rationale," which points to problematic phenomena in the literature, observed puzzling events in reality, or problematic theories that are challenged by new hypotheses, including questions that indicate gaps in the scope or certainty of our knowledge and its justification.

This category was expanded into three categories: (a) "poses statement of the problem," (b) "describes a research gap," and (c) "supports with rationale." Literature or research gaps were separated from "Problem Statement" to assess if an article describes something unknown or unsatisfactory, or if there is a lack of information in the current body of literature such as never-been-researched areas, inconclusive findings, or inappropriate methodologies. These refinements allowed the researchers to examine the foci of the Chinese journal articles in terms of how they delineated and justified their studies.

As a result, a more specific set of operational definitions was created (see the appendix), especially for those categories that had led to greatest interrater inconsistency. Using another random sample of 15 articles, this revised rubric resulted in kappa values ranging between .83 and 1.00. The established criteria for the interrater agreement suggests that reliability is good (.60-.74) to excellent (.75–1.00; Cicchetti, 2001) and substantial (.61-.80) to almost perfect (.81–1.00; Landis & Koch, 1977), suggesting that the results from the second trial fell in the excellent and almost perfect range. This revised coding form was then used to complete the review of the rest of the journal articles that were selected for the study. When the coding of the total sample was finished, a final reliability was calculated on a randomly selected sample of 50 articles that were drawn from the sample of 1,096 articles. According to Neuendorf (2002), a reliability subsample of between 50 and 200 units is appropriate for estimating levels of interrater agreement. In this case, 50 articles were selected and coded independently by the two raters to estimate interrater reliability. The Cohen's kappa values ranged from .82 (excellent/ perfect) to 1.00 (excellent/perfect) with an average of .98.

Coding Procedure

Our coding procedures were fairly standard for quantitative content analysis (e.g., Gall, Gall, & Borg, 2003). For each article, the rater rated the quality of each coding category on the coding form. For example, after entering the basic descriptive information for the article (i.e., journal title, publication year, and article title), the rater determined into which type of inquiry the article fell. If the study was empirical, all of the coding categories on the coding form would apply. If the articles were nonempirical, then the criteria under the method, results, and limitations sections were deemed not applicable. A separate coding form was used for each journal article.

Data Analysis

Other than Cohen's kappa reliability coefficient described above, the data analytical methods that were employed in this study were descriptive statistics, analyses of variance (ANOVA), and *t* test. For Objective 2 (examining the research quality of educational journal articles in China), we calculated the means and frequencies that allowed us to compare the ratings to the three standards: *nonexistent, weak*, and *strong*. For Objective 3 (identifying the strengths and weaknesses of published education research journals in China in light of the international standards), we employed one-way ANOVA to compare the mean score differences across the five main categories: introduction, literature review, method, results, and discussion sections. The independent variable was Research Quality

Category with five levels (introduction, literature review, method, results, and discussion), and the dependent variable is mean composite scores within each category collapsed across subcategories. This analysis allows us to identify the strengths and weaknesses in those five categories. We used a dependent-samples mixed-model approach to compare the category means because the same set of journals were repeatedly rated across categories and because it is more robust to violations of distributional assumptions. No coding errors, missing data, or outliers were found.

For Objective 4 (studying the longitudinal trend of quality change from 2002 to 2011), we used an independent-samples t test to identify changes in research quality between the first 5 years and the latter 5 years and to determine if any significant changes have taken place over the 10-year span. The independent variable is Year Span with two levels (2002–2006 and 2007–2011), and the dependent variable is mean composite score collapsed across subcategories. The rationale for this dichotomization was based on the previous research that suggested that article characteristics and methodological approaches tend to be stable within a 5-year span (Goodwin & Goodwin, 1985; Hutchinson & Lovell, 2004). There were no missing data because unavailable journals and articles that did not meet the inclusion criteria had already been excluded from the coding procedure. Assumption of homogeneity of variances was checked to be significant ($F = 16.51$, $p < .05$), meaning that the variances of the two 5-year spans were unequal, so we reported results from the t test for unequal variances.

Results

### *Developing a Rubric That Yields Valid and Reliable Scores*

The content validity of the evaluation rubric was established by its close alignment with the international standards through an extensive literature review. We included all major publication standards and created a comprehensive list of categories that are commonly adopted in the center community. We also recognized the variation across different types and paradigms of research by allowing the rubric to be modifiable to meet the needs of a particular type or paradigm. As noted in the method section, Cohen's kappa values were in the excellent to perfect range. This outcome provides confidence in the reliability of the coding of studies in this article. It also suggests that with proper training, the rubric allows the raters to reach a consensual understanding of what those categories mean and how to code the category values.

Research Quality of Educational Journal Articles in China

The overall mean score collapsed across categories and subcategories of the 1,096 articles was 0.25 ($SD = 0.59$) on the 3-point scale (0 = *nonexistent*, 1 = *weak*, 2 = *strong*), which means that these articles weakly or even barely met the standards. The mean scores of the five main categories ranged from 0.20 for the introduction and literature review to 0.87 for results (see Table 2). These mean values suggest that the results category appears to be relatively strong compared to the other categories but still below the *weak* standard. Eighty-six percent ($n = 940$) of the articles fell between 0.1 and 0.4, for both empirical and nonempirical studies, as shown in Figure 1. These low mean scores of the reviewed articles led

TABLE 2

Descriptive statistics of major evaluation categories

Groups	N	M	SD	SE	95% Confidence interval	Minimum	Maximum
Introduction	10,960	0.20	0.52	0.005	[0.19, 0.21]	0	2
Literature review	8,768	0.20	0.50	0.005	[0.19, 0.21]	0	2
Method	3,580	0.28	0.62	0.010	[0.26, 0.30]	0	2
Results	358	0.87	0.94	0.050	[0.77, 0.97]	0	2
Discussion	8,209	0.33	0.69	0.008	[0.31, 0.34]	0	2
Total	31,875	0.25	0.59	0.003	[0.24, 0.25]	0	2
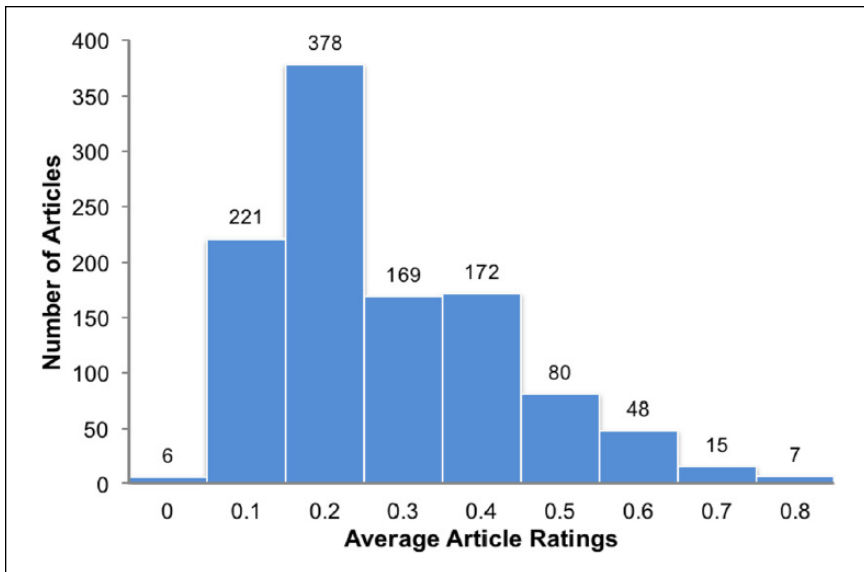
FIGURE 1. *Distribution of overall mean scores.*

to a further examination of the underlying causes for the overall low research quality of education journal articles in China.

Strengths and Weaknesses of Publications in Chinese Education Research Journals

A one-way dependent-samples ANOVA was used to compare the means of the five major categories in the coding rubric. The means of the five categories differed significantly, $F(4, 3641) = 411.629$, $p = 0.00$. The effect size η^2 is 0.31, indicating a large substantive difference among the category means. Except between the introduction ($M = 0.20$, 95% confidence interval [CI; 0.19, 0.21]) and the literature review ($M = 0.20$, 95% CI [0.19, 0.21]), all of the other nine pairs of

comparisons were statistically significant at $p < .05$. Both the introduction and the literature review received significantly lower ratings than the other three categories. The results section ($M = 0.87$, 95% CI [0.77, 0.97]) had significantly higher ratings than those of the other four categories. Results for each of the five categories are reported below.

Introduction

The overall mean of the 10 coding criteria in the introduction was 0.2, none of which was above 1. Research problems ($M = 0.36$) and purpose ($M = 0.37$) were two more highly rated categories, compared to significance ($M = 0.04$) and theoretical framework ($M = 0.02$; Table 3). Examining the research problem sections indicated that these articles identified problems mostly from practical real-world issues rather than from literature gaps. Of the 1,096 articles, 38% ($n = 413$) posed real-world problems either explicitly or implicitly, whereas only 7% ($n = 78$) indicated a literature gap (e.g., a never-been-researched area, inconclusive findings, or an inappropriate methodology). Only 33% ($n = 359$) described the problems or gaps further by providing background information and relevant literature to support such statements. Fifty-six percent ($n = 611$) did not describe the study's purpose, and 90% ($n = 984$) did not report the research questions to be addressed.

In keeping with research problems (i.e., focus only on practical issues), these articles tended to focus more on practical significance rather than theoretical or scholarly academic contributions. Only 2 out of the 1,096 articles that were reviewed indicated the theoretical importance of their inquiry. One percent of the articles ($n = 13$) described the substantive contribution and applicability to the scholarly research in the field, and 10% ($n = 114$) described the practical significance of the inquiry in the introduction. Two percent ($n = 24$) mentioned the theoretical framework that was used in their study, but only three articles provided further information regarding why such a framework was necessary.

Literature Review

In the literature review category, the *relevance* of the literature received an average rating of 0.67, which was quite weak, but it was the highest rated criteria among the eight coding items in this section. Approximately half of the articles ($n = 515$) reviewed literature that was highly relevant or somewhat relevant, and the other half did not even review the existing literature at all (see Table 3). The comprehensiveness of the literature review scored only 0.30, which suggests that few articles cited sufficient references that spoke to their research purposes. These articles also rarely reviewed what other researchers had found on the same topic. Only 11% ($n = 117$) identified the main debate within the field of inquiry by summarizing perspectives, theories, or methodologies that had been used in the previous literature, but the other 89% ($n = 979$) either had no literature review or only cited a few references without thoroughly or comprehensively examining the topic. In fact, only 17 out of the 1,096 articles explained their study's relationship with previous studies by demonstrating how their inquiries advanced or challenged the existing body of literature (see "Relation with previous research" in Table 3).

In addition, none of the articles that were reviewed mentioned inclusion or exclusion criteria for a literature review of any type. Twenty-seven articles (2%) provided critical examinations of the advantages or disadvantages of previous

TABLE 3

Article scoring and means for introduction, literature review, results, and discussion

Categories	Coding criteria	M	SD	0	1	2
Introduction						
Problem (M = 0.36)	Problem statement	0.54	0.76	684 (62%)	233 (21%)	180 (17%)
	Research gap	0.12	0.46	1,018 (93%)	21 (2%)	57 (5%)
	Rationale	0.43	0.67	737 (67%)	250 (23%)	109 (10%)
Purpose (M = 0.37)	Purpose	0.58	0.72	611 (56%)	335 (31%)	150 (14%)
	Research questions	0.15	0.48	984 (90%)	51 (5%)	55 (5%)
Significance (M = 0.04)	Theoretical importance	0	0.04	1,094 (100%)	2	0
	Practical importance	0.12	0.37	982 (90%)	98 (9%)	16 (1%)
	Applicability	0.01	0.13	1,083 (99%)	11 (1%)	2
Theoretical framework (M = 0.02)	Theoretical framework	0.03	0.23	1,072 (98%)	13 (1%)	11 (1%)
	Justification of theoretical framework	0	0.07	1,093 (100%)	2	1
Literature review						
Coverage	Relevance	0.67	0.79	581 (53%)	297 (27%)	217 (20%)
	Comprehensiveness	0.30	0.54	806 (74%)	245 (22%)	45 (4%)
	Inclusion and exclusion criteria	0.00	0.00	1,096 (100%)	0	0
	Main ideas	0.15	0.46	979 (89%)	70 (7%)	47 (4%)
	Critical examination	0.03	0.21	1,069 (98%)	20 (2%)	7
Synthesis	Contextualization of literature gaps	0.03	0.23	1,078 (98%)	5	13 (1%)
	Relation with previous research	0.02	0.15	1,079 (98%)	14 (1%)	3
	Integration and synthesis	0.37	0.58	749 (68%)	288 (26%)	59 (5%)
Method						
Design	Types of design	0.21	0.58	155 (87%)	9 (5%)	15 (8%)
	Appropriateness	0.04	0.3	175 (98%)	0	4 (2%)
Sampling and participants	Participants/sample	1.08	0.68	35 (20%)	95 (53%)	49 (27%)
	Eligibility and criterion	0.13	0.42	162 (91%)	11 (6%)	6 (3%)
	Criterion justification	0.03	0.21	174 (97%)	4 (2%)	1 (1%)
	Sample size justification	0.01	0.07	178 (99%)	1 (1%)	0
	Sampling procedure	0.18	0.5	163 (91%)	14 (8%)	2 (1%)
	Procedure justification	0.01	0.15	178 (99%)	0	1 (1%)

(continued)

TABLE 3 (continued)

Categories	Coding criteria	M	SD	0	1	2
Instrumentation/measures	Instruments or method	0.71	0.83	95 (53%)	41 (23%)	43 (24%)
	Reliability and validity	0.20	0.55	157 (88%)	9 (5%)	13 (7%)
	Data collection types	1.15	0.79	44 (24%)	64 (36%)	71 (40%)
Data collection	Collection methods	0.38	0.74	137 (76%)	14 (8%)	28 (16%)
	Collection procedure	0.40	0.72	130 (73%)	25 (14%)	24 (13%)
	Contextual info	0.49	0.71	134 (75%)	22 (12%)	23 (13%)
Data analysis	Analytical methods	0.44	0.79	134 (75%)	11 (6%)	34 (19%)
	Analytical procedure	0.11	0.44	167 (93%)	4 (2%)	8 (5%)
	Address research questions	0.02	0.15	175 (98%)	4 (2%)	0
	Confirm to research design	0.00	0	179 (100%)	0	0
	Unintended circumstances	0.01	0.17	177 (99%)	1 (1%)	1 (1%)
	Validity, reliability, or trustworthiness	0.02	0.21	177 (99%)	0	2 (1%)
Results	Effective presentation	1.72	0.52	6 (3%)	38 (21%)	135 (75%)
	Pertinence to research questions	0.02	0.21	177 (99%)	0	2 (1%)
Discussion						
Discussion	Findings interpretation	0.79	0.80	494 (45%)	339 (31%)	263 (24%)
	Interpretation to research problem	0.01	0.10	1,092 (100%)	2	2
	Interpretation to research questions	0.00	0.00	1,096 (100%)	0	0
	Connection with literature	0.07	0.28	1,026 (94%)	63 (6%)	7
Limitations	Conclusiveness and generalizability	0.04	0.23	172 (96%)	6 (3%)	1
	Unsolved problems	0.02	0.21	177 (99%)	0	2 (1%)
	Weakness	0.08	0.39	171 (96%)	1	7 (4%)
	Theoretical implications	0.01	0.14	1,087 (99%)	5	4
Implications	Practical implications	1.48	0.80	215 (20%)	136 (12%)	745 (68%)
	Implications for (future) research	0.08	0.34	1,037 (95%)	34 (3%)	25 (2%)

Note. Sections that are highlighted in gray were calculated based on the number of empirical studies (*n* = 179). The rest of the categories were calculated based on the 1,096 sample articles. In the first row, 0, 1, 2 represent the numbers and percentages of articles that were rated 0, 1, 2.

studies. Of the 78 articles that indicated research gaps in their introductions, only 18 specified what had been done and what needed to be done based on a literature review. The mean score of the integration and synthesis of literature reviews for the 1,096 sample articles was 0.37, 5% ($n = 59$) of which synthesized the review well to tie into the issues that were being investigated; 26% ($n = 288$) integrated a few examples from the literature into the main debate.

Method

The method section was the second lowest rated category after the introduction and literature review. Of the 1,096 articles that were coded in this study, 84% ($n = 907$) were nonempirical studies that included theoretical expositions, commentaries, reviews, and position papers, whereas 16% ($n = 179$) represented empirical studies that were evaluated based on 20 coding criteria in the method section. Eight of the 20 coding criteria obtained mean scores below 0.05 (see Table 3). Nine received average ratings between 0.1 and 0.5 and 3 between 0.8 and 1.2. Low scores on the coding criteria highlighted methodological ambiguity and incomplete descriptions of research design in these articles. For example, the articulation of appropriateness of research methodology averaged 0.04, as only four studies explained how a research methodology that was selected served their research purposes and addressed their research questions. The majority (87%; $n = 155$) did not relate their studies with any type of inquiry approach but simply started the method section with data collection.

In addition, these studies tended to only list what they did (e.g., the schools that they visited and the participants whom they recruited) without elaborating or justifying their choices. For example, information about the sample was rated with an average of 1.08, which ranked second place next to "Data collection types" with an average of 1.15. However, no rationale for the sample size or the numbers of participants was given except in one article. Ninety percent of the articles ($n = 161$) had no inclusion or exclusion criteria for selecting samples or participants, not to mention justification of any criterion (which only five articles did). Another problem with the method section was the underreporting of data analyses. The descriptions of data analyses averaged 0.44, with 74% ($n = 133$) missing data analytical methods and 93% ($n = 166$) without data analytical procedures. Many studies confused analytical methods with tools such as SPSS and Nvivo, thus listing software as the data analysis method. None of the articles that were reviewed mentioned how data analysis methods and procedures conformed to their research designs. Other elements in the method section that were underreported included information about intended or unintended circumstances, or missing data that might have affected data analysis, and a discussion of validity and trustworthiness of data analysis due to data treatment (which only two studies did for each item).

Results

The results section, with an average of 0.87, was the highest rated category but still did not reach the *weak* rating. This section was rated based on (a) the effectiveness of the presentation of results and (b) the descriptions of the results that were pertinent to each research question or hypothesis, which were applicable only to empirical studies in this review (see Table 3 Results categories). The criteria for effectiveness of results scored at 1.72, almost reaching *strong* on our

3-point scale. Seventy-five percent (*n* = 135) of the articles were able to report complete results with sufficient and appropriate amounts of data, and another 21% (*n* = 38) reported results with some amount of data. The mean rating was brought down by the low score (*m* = 0.02) on the description of results that were pertinent to research questions or hypotheses, as the vast majority of the articles did not have research questions.

Discussion

In the discussion section, interpretation of findings was rated with a mean score of 0.79, which was the second-highest rated criterion next to practical implications (see Table 3). Only 24% (*n* = 263) of the 1,096 articles interpreted findings or education phenomena effectively using empirical evidence, observed data, or examples. Thirty-one percent (*n* = 339) of the articles discussed their findings or education phenomena somewhat effectively with empirical evidence, observed data, or examples. Discussions that were related to research problems and questions were two extremely weak aspects of these publications. Only four of the articles explained how their claims and interpretations addressed research problems or the issues that were being investigated. However, none of the papers provided a statement about how their interpretations addressed their research questions, as they usually did not state research questions. In terms of connection with the existing literature, only 6% (*n* = 70) related discussions with previous studies either by demonstrating how their findings and arguments supported or challenged previous studies, whereas 94% (*n* = 1,026) did not cite any type of literature in their discussion.

Empirical studies were coded based on acknowledgement of limitations, namely, conclusiveness, unsolved problems, and weaknesses, which received mean scores of 0.02, 0.08, and 0.01, respectively. Only 4% (*n* = 5) stated to what extent their findings and discussions were conclusive and generalizable, 1% (*n* = 2) indicated unsolved problems, and 4% (*n* = 8) noted other types of weakness in their studies either explicitly or implicitly. With respect to implications, practical implications (*M* = 1.48) was rated the highest category among three types of implications and all of the coding criteria in the discussion and conclusion sections. Eighty percent (*n* = 881) of the articles drew practical implications by suggesting applications of their findings to educational practice or policy making. This finding echoes the relatively high ratings of the problem statements from real-world issues (*M* = 0.54; in comparison with *research gaps*) and practical significance (*M* = 0.12; compared with significance to theory and scholarly research) in the introduction section. All of these three categories are more concerned with practical aspects of education than theoretical focuses or issues that were derived from literature (e.g., addressing literature gaps). Emphasis on theoretical implications and suggestions for research, on the other hand, scored at an average of 0.01 and 0.08 respectively. Only 9 out of 179 (5%) research articles drew implications for educational theory, and 59 articles (33%) provided suggestions for further research.

Change in Education Research Quality From 2002 to 2011

Overall, mean scores by years were calculated to identify changes in research quality over the 10-year span. As shown in Figure 2, there has been a slight
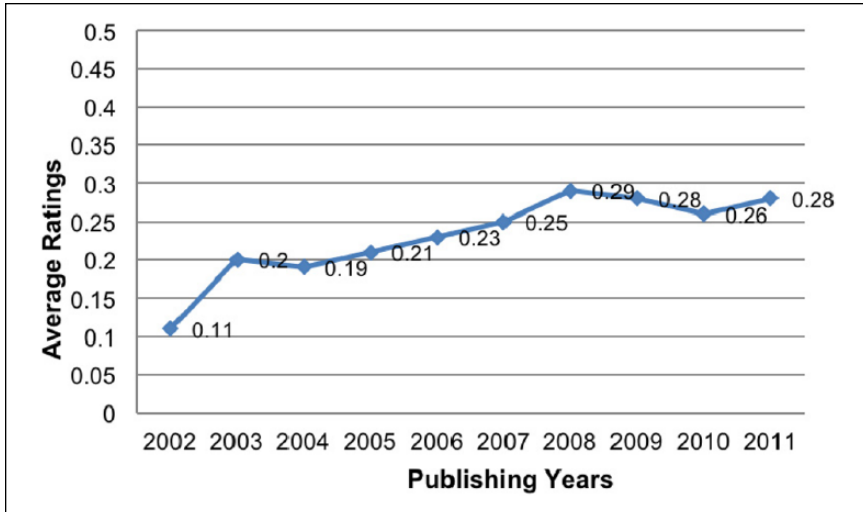
FIGURE 2. *Average rating changes between 2002 and 2011.*

increase in ratings from an average of 0.11 in 2002 to 0.28 in 2011, a growth of approximately 0.02 annually, which suggested a small improvement in research quality by international publishing standards. As shown in the graph, the sharpest increase occurred between 2002 and 2003, where the average rating reached 0.2 from 0.11 and stayed above 0.2, except for a slight decrease in 2004. To compare mean differences between the first 5 years and the latter 5 years, an independent-samples t test with unequal variances was conducted.

There was a significant difference in the mean scores between 2002 and 2006 ($M = 0.21$, $SD = 0.14$) and 2007 and 2011 ($M = 0.27$, $SD = 0.16$), $t(1,032.24) = -7.20$, $p = 0.00$. The effect size dcorrected was 0.40, suggesting a small to moderate substantive mean difference between the two five-year spans. These results suggest that the research quality improved slightly over the 10-year span. A further examination of the five categories revealed that the most notable increase over the 10-year span was the results category, the means of which improved from 0.75 to 0.91 between 2002 and 2011 (see Figure 3). Between the two coding items in the results, the effectiveness of the results presentations showed a higher growth rate, from 1.63 to 1.77, whereas descriptions of the results that were pertinent to the research questions or hypotheses increased from 0 to 0.04 in the same period. Except for the method section, all of the other categories, including the introduction, literature review, and discussion, experienced a minor increase by approximately 0.1 from 2002 to 2011.

Discussion

The current review study was conducted in response to the need to explore potential reasons for limited global recognition and low citation rate of scholarly work produced in peripheral emerging countries despite their effort to increase its
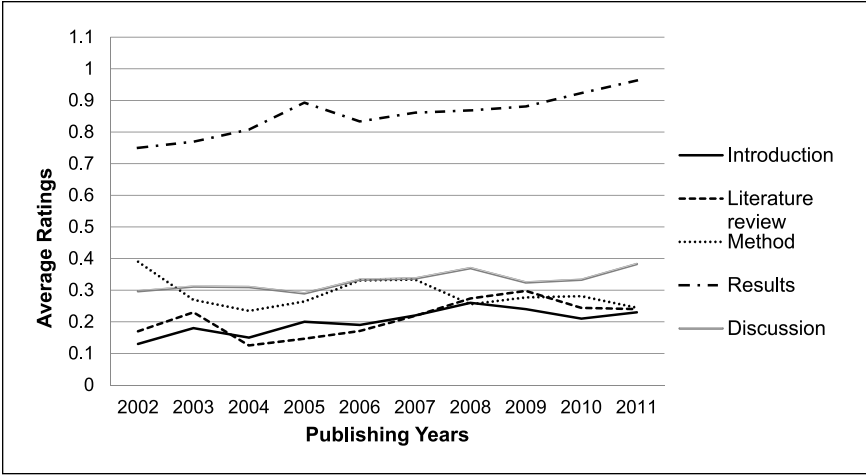
FIGURE 3. *Average rating change by groups.*

knowledge production and dissemination. We selected China as a case of the emerging peripheral countries to explore the characteristics of research quality in its published educational journal articles to help shed some light on the issue. Our systematic review of 1,096 articles shows several major characteristics as potential contributing factors to their low global visibility and recognition: (a) a shortage of empirical research, (b) a decontextualized (local situatedness that does not relate to previous work in other contexts) and atheoretical (limited power for generalization) approach to framing studies, and (c) low research quality that is seriously affected by nonsystematic literature reviews, lack of reference to previous research, methodological ambiguity, and incomplete descriptions of research design, among others.

Our findings are consistent with previous studies (e.g., Yang, 2006; Zhao et al., 2008; Zheng & Cui, 2001) that also identified the shortage of empirical research as a key feature of journals articles published in China. A holistic and dialectical epistemological tradition that relies on personal experience, reflections, and wisdom in the East (Nisbett, 2003) may have *partly* deterred Chinese education researchers from conducting empirical studies (Zhao et al., 2008). However, the material and social reality at the periphery may be more powerful determinants in preventing this transformation. Canagarajah (2003) found that other peripheral scholars were also more devoted to an "essayistic" (p. 204) style of academic writing in which they could draw informal observations, casual conversation, and sheer experience as evidence. Conceptual papers are thus popular in peripheral communities despite the fact that they might be less likely to be accepted by Western standards.

In addition, the promotion and evaluation requirement for faculty in China may be another factor that sustains the popularity of nonempirical discourse, as most universities in China evaluate faculty's scholarship based on the number of their publications (Shi, 2002) regardless of the types of publications and their research quality (Li, 2004). This comment is intended not to suggest that empirical studies

are always of better quality than nonempirical papers but simply to highlight the emphasis on nonempirical discourse, which may require less time and effort to publish than empirical investigations. Apparently, knowledge production and publishing conventions are contextual, shaped by material, historical, and social conditions that govern a community's life and experience (Canagarajah, 2002); nevertheless, publishing conventions and promotion mechanisms may need to be negotiated and modified to help Chinese publications meet international publishing standards and enhance their participation in the global research community.

Our review also suggests that education researchers in China appear to be more concerned with practical aspects of educational studies, which concurs with previous studies (Kang, 2002; Ross; 2000). Specifically, the articles we reviewed focused on practical and local issues and were atheoretical in nature. This finding echoes the problems identified in educational policy research that appeared to be "superficial and far-fetched" (Yang, 2006, p. 215) for lacking theoretical framing and empirical evidence. Although local and practical issues in China can be of interests to the global research community, they need to be investigated by connecting them to similar studies conducted in other countries, relating them to theories used to inform similar research problems, and drawing international implications.

As reported in the results section, a vast majority of education research journal articles published in China only weakly or even barely meet the international standards. This outcome indicates a discrepancy between the country's movement toward internationalization of its knowledge production and the reality of their current research practice in education. Despite various efforts made by the government, universities, and journal editors, major research quality issues still exist that hinder the country from achieving its goal in the promotion and dissemination of its academic knowledge.

Specifically, insufficient and nonsystematic literature reviews and a lack of reference to previous work are salient features identified in the review, which directly affected the low scoring of introduction, literature review, and discussion sections of the articles reviewed. A systematic literature review is crucial for education researchers to understand the subject matter and knowledge that has been accumulated by previous researchers, with which they could participate in an ongoing debate with their contribution (Yang, 2005). However, the articles reviewed seldom situated their inquiry in the context of the existing body of literature. This contrasts significantly with scientific papers in the center where scholars strive to create an intellectual space by situating their inquiry in the context of existing literature and stating a literature gap and ways that their study can fill the gap (Swales, 2004).

Rather than building on previous studies or strengthening their arguments with reference to existing literature, authors of the articles we reviewed also explained their results/findings based on their personal reflections and observations. This phenomenon is reflected mostly in nonempirical discourse that adopts an "informal dialogic approach" (Hayhoe & Pan, 2001, p. 3), in which the authors appear to discuss issues casually. This suggests an informal attitude toward academic publications and ineffective use of resources. Such work that lacks sufficient review of and reference to existing literature scholarship fails to demonstrate its contribution to or expansion of the existing knowledge base. Furthermore, scholars may repeat

similar information that is already reported in previous publications or disseminate untested knowledge based only on the authors' personal experiences and reflections, which can potentially mislead educational practice and policymaking.

Methodological ambiguity and incomplete descriptions of research design were also prevailing issues in the empirical studies, which dramatically lowered their ratings. As Koro-Ljungberg, Yendol-Hoppey, Smith, and Hayes (2009) pointed out, methodological ambiguity can negatively affect research design because

> When researchers do not make as explicit as possible their epistemologies, theoretical perspectives, justification/argumentation systems, and methodologies, as well as the alignment of their research designs within the decision junctures that guide research processes, their research designs can appear random, uninformed, inconsistent, unjustified, and/or poorly reported. (p. 688)

This is the case with Chinese publications that were analyzed in the current study. The method section was usually poorly reported, with no systematic or consistent steps. Many articles contained no data collection procedures and rationale for choosing a particular method. Underreporting of data analysis, including analytical methods, analytical procedures, information about circumstances or missing data that may affect data analysis, and discussion of reliability, validity, or trustworthiness of data analysis, was common in these articles.

These findings replicate those by Zhao et al. (2008), who found that Chinese journal articles usually do not follow a prescribed methodology of data collection and analysis. Such research practices go against modern research ethics that are concerned with systematic procedures in knowledge production and, most essentially, with the power of replicability (Appadurai, 2000). It is thus very unlikely for such articles to be accepted or recognized by scholars in the center.

One question is why the Chinese scholars who published the articles reviewed for the current study write the way they do. As Feng et al. (2013) noted, it could be due to rhetorical background of Chinese scholars who are not trained to write in ways that meet global publishing standards. In fact, there is often a sense of isolation for academics on the periphery, which poses an impediment to the improvement of research quality and to their engagement in global knowledge community (Hyland, 2015). They often feel out of the loop on the current developments in their field due to limited funding, access to up-to-date technologies, a lack of awareness of what constitutes scientific research, and unfamiliarity with the broad (and unwritten) "rules of the game" (Gosden, 1992, p. 133). The incomplete descriptions in the method sections and methodological ambiguity could be examples that show the potential lack of sufficient knowledge of scientific research. The confusion of analytical methods with software tools may derive directly from limited knowledge and training in scientific investigation and reporting. However, as Hyland (2015) argued, despite the fact that some of the social and economic barriers limited the participation of peripheral scholarship in the center, these factors needed not be determining ones. Academic writing above all is a literary practice that can be learned as needed. Peripheral scholars can learn and craft their academic writing to meet global publication standards.

It should be noted, however, that our study also shows an encouraging trend of slight increase of research quality between 2002 and 2011, an indication that progress toward publishing in accordance with international standards was made and may likely continue. This increase in quality aligns with previous studies on the shift that was taking place in China from sharing personal experiences to valuing evidence-based research (Gao et al., 2001; Sun, 2011). This finding could suggest that the efforts in adopting international publication standards have started to show fruition, signaling some small progress being made toward achieving China's goal to compete internationally by "exporting" its knowledge and gaining global respect for its scholarship.

Implications From the Current Study

The findings of our study have a number of important implications for education research and publication practice in China and potentially other peripheral emerging countries that are experiencing similar problems. First, Chinese education researchers can improve the global visibility of their research and publication by conducting more original research, contextualized within the existing international research literature by systematically and comprehensively reviewing prior academic work, discussing their relevance to the research problem under consideration, and drawing broader implications of their findings to international contexts. A systematic review of international research can help researchers in China and other emerging countries to identify research gaps that can gain international recognition for their contribution to extending the existing knowledge base.

Second, education researchers in China and other peripheral countries whose work may not have been cited due to its atheoretical nature can incorporate theories into their studies that have informed existing international research. Atheoretical work, found to be prevalent in China, for example, has limited impact and generalizability and therefore does not attract citation. Education researchers can employ existing theories from abroad to inform local work, compare the findings with similar studies that have been reported internationally, or enrich the existing theories with new concepts or theories generated from research conducted domestically. They should pay closer attention to the development of indigenous knowledge and philosophical traditions and help accumulate, expand, and disseminate these concepts and ideas to a wider global readership, as indigenous knowledge is valuable to the vitality and democratization of global knowledge construction.

Third, education researchers in China and other peripheral countries can improve their publications by providing a more detailed and complete method and design as well as reporting and discussing results with more contextually supporting evidence. Such improvements can help make a real impact on education research and practice as well as gain recognition and attract more citations. Specific initiatives include providing instructions and courses that help peripheral researchers increase their awareness and understanding of education research publication standards, as well as encouraging collaborations and scholarly exchange between domestic and overseas scholars and partners (Lillis & Curry, 2006).

We call for a stronger peripheral research community that implements the above recommendations by conducting and reporting research systematically and rigorously using the research quality rubric such as the one developed and used to

evaluate the education research articles reviewed for the current study. Transforming research quality takes the whole research community as well as policy and funding support for training and instructions that governments provide to their researchers. It should be noted, however, that improving research quality does not necessarily mean imposing Western hegemony in academic research by following only Western paradigms and academic writing conventions.

Rather, we argue that improving education research quality in China and other emerging countries is a means to better disseminate and share knowledge and contribute to the vitality of global knowledge, while celebrating and valuing indigenous knowledge. In fact, we make these suggestions because, as discussed in the Background section of this article, the peripheral research community, including China, aspires to improve the visibility of their scholarship and contribute to global knowledge production. We hope the issues emerged from our findings serve as an inspiration for other peripheral countries in their pursuit of global recognition for their research and desire to contribute to global research knowledge.

Limitations and Future Research

Several limitations of the current study suggest directions for future research. First, as the current review focused only on education research journals in China, we encourage future research that examines research quality issues in other peripheral countries to see if similar issues may contribute to the low global impact of their academic research scholarship. Future research could replicate the current study by using the coding rubric we developed and used for this study. Such studies can make significant contribution to global knowledge construction and dissemination by providing critical evidence and insight through examination of research quality in the global community.

Second, this study included education research journal articles that were published in 2002 to 2011 because most of the previous review studies examined research published before 2002. Since China's *Decision* to promote academic scholarship started in 2002 and previous studies have suggested that article characteristics and methodological approaches tend to be stable within a 5-year span (Goodwin & Goodwin, 1985; Hutchinson & Lovell, 2004), we decided to study a 10-year span after 2002. Future research can replicate this study by studying another 10-year span from 2012 to 2021 to monitor the longitudinal progress of education research quality in China.

Third, the content validity of the evaluation rubric we developed was not empirically tested by experts in education research in either center or periphery. Future research in this direction is necessary if the rubric is to meet international standards that can be accepted by both the center and periphery. Fourth, our extensive search for international standards that are compatible with different research paradigms and epistemologies proved to be a daunting task due to the inherently different world views and philosophical lenses through which education research is evaluated (Guba & Lincoln, 1994; Howe, 2009). Future research that involves evaluators of education research and scholarship from both positivistic and postpositivistic ontologies can shed light on whether we can move in the direction of forging transparadigmatic partnership in improving education research quality.

APPENDIX

Coding form

Sections	Categories	Subcategories	Nonpresent, 0	Weak, 1	Strong, 2
Introduction	Problem statement	1. Poses a statement of the problem			
		2. Describes a research gap			
		3. Supports with rationale			
	Purpose	1. Explains the purpose of the study			
		2. States research questions (for qualitative study) or hypothesis that includes variables going to be measured and studied (for quantitative)			
	Significance	1. Demonstrates theoretical importance of study			
		2. Shows practical importance			
		3. Suggests originality, applicability, and interest to the field			
	Theoretical framework	1. Describes conceptual or theoretical framework used in the study			
		2. Justifies the conceptual or theoretical framework			
Literature review	Coverage	1. Literature review is relevant			
		2. Literature review is comprehensive			
		3. Has justified criteria for inclusion and exclusion from review			
	Synthesis	1. Identifies main ideas, perspectives (theories), or methodologies used in the field			
		2. Critically examines their advantages or disadvantages			
		3. Distinguishes what has been done to what needs to be done			
		4. Explains relations with previous studies by demonstrating how the current research joins and advances or challenges the existing literature			
		5. Integrates and synthesizes the review to tie into the issues being investigated in the current study			

(continued)

Sections	Categories	Subcategories	Nonpresent, 0	Weak, 1	Strong, 2
Method	Research design and method theory	1. Describes types of research design, method, or methodology			
		2. Articulates its appropriateness: how research design/method relate to research questions or hypothesis			
	Sampling/participants	1. Provides information about participants or samples sufficient for the purpose of the study			
		2. Notes eligibility and exclusion criterion or special arrangements			
		3. Justifies eligibility and exclusion criterion or special arrangements			
		4. Justifies appropriateness of samples size or number of participants to research questions			
		5. Describes procedures for selecting participants or samples			
		6. Justifies procedures for selecting participants or samples			
	Instrumentation/measures	1. Describes instruments or method employed and their purpose in the study			
		2. Explains reliability and validity of the instruments or measures			
	Data collection	1. States types of data collected			
		2. Describes the ways in which data were gathered or identified			
		3. Outlines data collection procedures, including time and duration			
		4. Provides the context information (settings or locations) of data gathered			
	Data analysis	1. Describes analytic method/techniques			
		2. Outlines procedure of data analysis			
		3. Clearly describes how analysis procedures address research questions or problem			
		4. Makes it clear how analysis procedures conform to research design			
		5. Includes information about intended or unintended circumstances that may affect analysis and inferences			
		6. Discusses reliability, validity, or trustworthiness (e.g., potential sources of bias and the effects due to data treatment)			

611

APPENDIX (continued)

Sections	Categories	Subcategories	Nonpresent, 0	Weak, 1	Strong, 2
Results	Results/findings	1. Presents results effectively (reports complete results with sufficient and appropriate amount of data presented)			
		2. Describes findings/results pertinent to each research hypothesis or question			
Discussion	Discussion	1. Interprets the findings and explains patterns in the data (document data for nonempirical study) and relations among variables with evidence and concrete examples			
		2. Explains how claims and interpretation address the research problem/issue			
		3. Explains how claims and interpretations address research questions			
		4. Relates the findings/arguments to the broader problem in the field by demonstrating how the conclusions connect to support, elaborate, or challenge those in previous studies			
	Limitations	1. Considers to what extent the results/findings are conclusive and can be generalized			
		2. Indicates unsolved problems			
		3. Notes the weaknesses of the study			
	Implications	1. Emphasizes implications for theory			
		2. Draws implications for practice			
		3. Discusses implications for (further) research			

Note. Categories highlighted in gray do not apply to nonempirical work.

Note

[1]In this study, mainland China excludes special administrative regions such as Hong Kong and Macau, which have their own research and publishing standards for historical reasons.

References

Altbach, P. G. (2002). Centers and peripheries in the academic profession: The special challenges of developing countries. In P. Altbach (Ed.), *The decline of the guru: The academic profession in developing and middle-income countries* (pp. 1–22). Chestnut Hill, MA: Boston College Center for International Higher Education.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications: American Educational Research Association. *Educational Researcher*, *35*(6), 33–40. doi:10.3102/0013189X035006033

American Educational Research Association. (2009). Standards for reporting on humanities-oriented research in AERA publications: American Educational Research Association. *Educational Researcher*, *38*, 481–486. doi:10.3102/0013189x09341833

American Psychological Association. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*, 839–851. doi:10.1037/0003-066x.63.9.839

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Appadurai, A. (2000). Grassroots globalization and the research imagination. In A. Appadurai (Ed.), *Globalization* (pp. 1–21). Durham, NC: Duke University Press.

Arunachalam, S. (2008). The science race continues in Asia. *Current Science*, *94*, 848–849.

Association for the Study of Higher Education. (n.d.). *Scholarly paper review criteria*. Retrieved from http://www.ashe.ws/?page=345

Atkinson, D. (1999). TESOL and culture. *TESOL Quarterly*, *33*, 625–654. doi:10.2307/3587880

Aydinli, E., & Mathews, J. (2000). Are the core and periphery irreconcilable? The curious world of publishing in contemporary international relations. *International Studies Perspectives*, *1*, 289–303. doi:10.1111/1528-3577.00028

Beckett, G. H., & Zhao, J. (2016). Deweyan student-centered pedagogy and Confucian epistemology: Dilemmatic pragmatism and neo-patriotism? In C. P. Chou & J. Spangler (Eds.), *Chinese education models in a global age: Transforming practice into theory* (pp. 265–277). Singapore: Springer.

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, *34*(6), 3–15. doi:10.3102/0013189x034006003

Brewer, J. (2005). Three formulating research problems. In J. Brewer & A. Hunter (Eds.), *Foundations of multimethod research: Synthesizing styles* (pp.39–57). Thousand Oaks, CA: Sage.

Canagarajah, A. S. (2002). *A geopolitics of academic writing*. Pittsburgh, PA: University of Pittsburgh Press.

Canagarajah, A. S. (2003). A somewhat legitimate and very peripheral participation. In C. P. Casanave & S. Vandrick (Eds.), *Writing for scholarly publications: Behind the scenes in language education* (pp. 197–210). Mahwah, NJ: Lawrence Erlbaum.

Chen, S. (1994). Research trends in mainland Chinese comparative education. *Comparative Education Review*, *38*, 233–252. doi:10.1086/447243

Cicchetti, D. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, *23*, 695–700. doi:10.1076/jcen.23.5.695.1249

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Crotty, M. J. (1998). *The foundations of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: Sage.

Educational Researcher. (n.d.). *Aims and scopes*. Retrieved from http://www.sagepub.com/journalsProdDesc.nav?prodId=Journal201856

European Association of Science Editors. (2013). *EASE guidelines for authors and translators of scientific articles to be published in English*. Retrieved from http://www.ease.org.uk/sites/default/files/ease_guidelines-june2013-english.pdf

Fan, B. S. (2000). *Dangdai zhongguo shehui kexue baokan wenxian fenxi: 1978–1995* [An analysis of contemporary China journal and newspaper social science literature: 1978–1995]. Shanghai, China: East China Normal University Press.

Feng, H., Beckett, G. H., & Huang, D. (2013). From "import" to "import-export" oriented internationalization: The impact of national policy on scholarly publication in China. *Language Policy*, *12*, 251–272. doi:10.1007/s10993-013-9285-8

Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education* (7th ed.). New York, NY: McGraw-Hill.

Gao, Y., Li, L., & Lü, J. (2001). Trends in research methods in applied linguistics: China and the west. *English for Specific Purposes*, *20*(1), 1–14. doi:10.1016/s0889-4906(99)00015-0

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston, MA: Allyn & Bacon.

Gastel, B. (2002). Guide published for peer reviewers of research manuscripts. *Science Editor*, *25*, 46–48.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979–1983: The preparation of graduate students to read the educational research literature. *Educational Researcher*, *14*(2), 5–11. doi:10.3102/0013189x014002005

Gosden, H. (1992). Research writing and NNSs: From the editors. *Journal of Second Language Writing*, *1*, 123–139.

Gottfried, G. M., Vosmik, J. R., & Johnson, K. E. (2008). *A rubric for evaluating a psychology research report*. Washington, DC: Office of Teaching Resources in Psychology, Society for the Teaching of Psychology. Retrieved from http://devscilabs.com/portal/files/Rubric%20article%20draft_July%2015.pdf

Government Accountability Office. (1989). *Content analysis: A methodology for structuring and analyzing written material*. Retrieved from http://archive.gao.gov/d48t13/138426.pdf

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.

Guilford, G. (2013, October 17). China is investing more than ever in science, but it's not paying off. *Quartz*. Retrieved from https://qz.com/136012/china-is-investing-more-than-ever-in-science-but-its-not-paying-off/

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, *43*, 19–23. doi:10.3102/0013189x13520289

Hayhoe, R., & Pan, J. (Eds.), *Knowledge across cultures: A contribution to dialogue among civilizations*. Hong Kong, SAR: University of Hong Kong, Comparative Education Research Center.

Howe, K. R. (2009). Positivist dogmas, rhetoric, and education science question. *Educational Researcher*, *38*, 428–440.

Hutchinson, S. R., & Lovell, C. D. (2004). A review of methodological characteristics of research published in key journals in higher education: Implications for graduate research training. *Research in Higher Education*, *45*, 383–402. doi:10.1023/B:RIHE.0 000027392.94172.d2

Hyland, K. (2015). *Academic publishing: Issues and challenges in the construction of knowledge*. Oxford, England: Oxford University Press.

Jiang, K. (2004). Hanyang kexue jingshen: Guanyu jiaoyu yanjiu fangfalun de shengsi [Building scientific spirit: Reflections on the methodology in educational research]. *Journal of Peking University–Philosophy and Social Sciences*, *41*(1), 63–70.

Ju, N., & Wang, J. (2008). Hexin qikan "banmianfei" zhenxiang [The truth about the publication fees charged by core journals]. *Southern Weekly*. Retrieved from http://www.infzm.com/content/trs/raw/38654

Kang, Y. (2002). *Jiaoyu zhidu de neishengxing* [Endogenousness of the education institutions]. *Theory and Practice of Education*, *22*(2), 11–16.

Koro-Ljungberg, M., Yendol-Hoppey, D., Smith, J. J., & Hayes, S. B. (2009). (E)pistemological awareness, instantiation of methods, and uninformed methodological ambiguity in qualitative research projects. *Educational Researcher*, *38*, 687–699. doi:10.3102/0013189x09351980

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

Li, L. (2004). Analysis of factors restricting the development of science and technology journals in China. *Publishing Research Quarterly*, *19*(4), 45–53. doi:10.1007/s12109-004-0007-0

Lillis, T., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication*, *23*, 3–35. doi:10.1177/0741088305283754

Liu, L. (2008). The evolution of China's science and technology policy, 1975–2007. In *OECD review of innovation policy: China* (pp. 381–393). Retrieved from http://climatesolver.org/sites/default/files/pdf/0809.pdf

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2010). *Practical resources for assessing and reporting intercoder reliability in content analysis research projects*. Retrieved from http://astro.temple.edu/~lombard/reliability

Matsuda, P. K. (2013, June). *English only? Exploring the importance of academic writing across languages*. Paper presented at the 7th conference of the European Association for the Teaching of Academic Writing, Budapest, Hungary. Retrieved from http://www.youtube.com/watch?v=3W295MGGgyo

Miller, W. R., & Wilbourne, P. L. (2002). Mesa grande: A methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction*, *97*, 265–277. doi:10.1046/j.1360-0443.2002.00019.x

Ministry of Education. (2014). *Jiaoyubu guanyu yinfa "guoji Hezuo lianhe shiyanshi jihua"* [Ministry of Education plans to launch "the international cooperative united

lab project"]. Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s3338/201401/163183.html

Moiwo, J., & Tao, F. (2013). The changing dynamics in citation index publication position China in a race with the USA for global leadership. *Scientometrics*, *95*, 1031–1051.

Morley, J. A., Finney, J. W., Monahan, S. C., & Floyd, A. S. (1996). Alcoholism treatment outcome studies, 1980–1992: Methodological characteristics and quality. *Addictive Behaviors*, *21*, 429–443. doi:10.1016/0306-4603(95)00073-9

Moyer, A., Finney, J. W., & Swearingen, C. E. (2002). Methodological characteristics and quality of alcohol treatment outcome studies, 1970–98: An expanded evaluation. *Addiction*, *97*, 253–263. doi:10.1046/j.1360-0443.2002.00017.x

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academies Press.

Neuendorf, K. A. (2002). *The content analysis handbook*. Thousand Oaks, CA: Sage.

Nisbett, R. (2003). *The geography of thought: How Asians and westerners think differently . . . and why*. New York, NY: Free Press.

Nunn, R., & Adamson, J. (2007). Towards the development of alternative criteria for international journal paper evaluation. *Asian EFL Journal*, *9*, 205–228.

Pajares, P. (2007). *Elements of a proposal*. Retrieved from http://www.uky.edu/~eushe2/Pajares/proposal.html

Price, E. G., Beach, M. C., Gary, T. L., Robinson, K. A., Gozu, A., Palacio, A., . . . Cooper, L. A. (2005). A systematic review of the methodological rigor of studies evaluating cultural competence training of health professionals. *Academic Medicine*, *80*, 578–586. doi:10.1097/00001888-200506000-00013

Randolph, J. J., Bednarik, R., & Myller, N. (2005). A methodological review of the articles published in the proceedings of Koli calling 2001-2004. In T. Salakoski, T. Mäntylä, & M. Laakso (Eds.), *Koli calling 2005: Proceedings of the 5th annual Finnish/Baltic Sea conference on computer science education* (pp. 103–109). Turku, Finland: Turku Center for Computing.

Randolph, J. J., Julnes, G., Lehman, S., & Sutinen, E. (2008). A methodological review of computer science education research. *Journal of Information Technology Education*, *7*, 135–162.

Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*, *38*, 1311–1319. doi:10.1007/s10803-007-0517-7

Review of Educational Research. (n.d.). *Aims and scopes*. Retrieved from http://www.sagepub.com/journalsProdDesc.nav?prodId=Journal201854#tabview=aimsAndScope

Ross, H. (2000). In the moment-discourses of power, narratives of relationship: Framing ethnography of Chinese schooling, 1981–1997. In J. Liu, H. Ross, & D. Kelly (Eds.), *The ethnographic eye: Interpretive studies of education in China* (pp.123–152). New York, NY: Falmer Press.

Salager-Meyer, F. (2008). Scientific publishing in developing countries: Challenges for the future. *Journal of English for Academic Purposes*, *7*, 121–132. doi:10.1016/j.jeap.2008.03.009

Salager-Meyer, F. (2015). *Peripheral scholarly journals: From locality to globality*. Retrieved from http://www.redalyc.org/html/2870/287042542002/

Schostak, J. (2008). *Qualitative research assignment/dissertation/thesis checklist*. Retrieved from http://www.enquirylearning.net/ELU/Issues/Research/dissertation-check.html

Shi, L. (2002). How Western-trained Chinese TESOL professionals publish in their home environment. *TESOL Quarterly*, *36*, 625–634.

Shi, L., Wang, W., & Xu, J. (2005). Publication culture of foreign language education journals in China. *TESOL Quarterly*, *39*, 765–776. doi:10.2307/3588537

Stokes, C. S., & Miller, M. K. (1975). A methodological review of research in rural sociology since 1965. *Rural Sociology*, *40*, 411–434.

Sun, M. (2011). Educational research in mainland China: Current situation and developmental trends. *Comparative Education*, *47*, 315–325. doi:10.1080/03050068.2011.586764

Suydam, M. N. (1968). An instrument for evaluating experimental educational research reports. *Journal of Educational Research*, *61*, 200–203. doi:10.1080/00220671.1968.10883645

Swales, J. M. (2004). *Research genre: Explorations and applications*. Cambridge, England: Cambridge University Press.

Swales, J. M., & Feak, C. B. (1994). *Academic writing for graduate students*. Ann Arbor: University of Michigan Press.

Tam, K. Y. B., & Chen, M. L. (2010). Examining scholarship in China's academe: An exploratory study. *Higher Education*, *60*, 69–84. doi:10.1007/s10734-009-9288-8

Wang, S., Wang, H., & Weldon, P. R. (2007). Bibliometric analysis of English-language academic journals of China and their internationalization. *Scientometrics*, *73*, 331–343. doi:10.1007/s11192-007-1775-z

Ware, M., & Mabe, M. (2012). *The STM report: An overview of scientific and scholarly journal publishing*. The Hague, Netherlands: International Association of Scientific, Technical, and Medical Publishers. Retrieved from http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1008&context=scholcom

Yang, R. (2005). Internationalisation, indigenisation and educational research in China. *Australian Journal of Education*, *49*, 66–88. doi:10.1177/000494410504900104

Yang, R. (2006). What counts as "scholarship"? Problematising educational policy research in China. *Globalisation, Societies and Education*, *4*, 207–221. doi:10.1080/14767720600752577

Zhang, J., & Lu, A. (2008). Review of teachers' mental health research in China since 1994. *Frontiers of Education in China*, *3*, 623–638. doi:10.1007/s11516-008-0041-3

Zhao, Y., Zhang, G., Yang, W., Kirkland, D., Han, X., & Zhang, J. (2008). A comparative study of educational research in China and the United States. *Asia Pacific Journal of Education*, *28*, 1–17. doi:10.1080/02188790701849826

Zheng, R., & Cui, L. (2001). *Ershinian lai woguo jiaoyu yanjiu fangfa de huiggu yu fansi* [A review of educational research methods in the last two decades]. *Educational Research*, *6*, 17–21.

Authors

JUANJUAN ZHAO earned her doctoral degree in educational studies from the University of Cincinnati, 615 Teachers College, Cincinnati, OH 45221, USA; email: *juanerzhao@gmail.com*. Her research interests include research methodology, second language studies, and international and comparative education. She has published many studies in those areas.

GULBAHAR H. BECKETT is a professor at Iowa State University, Ames, IA, USA; email: *beckett@iastate.edu*. Her research interests include academic discourse, qualitative, mixed-methods, and meta-synthesis research methods. She has published widely on language and general education issues concerning cross-cultural studies and globalization.

LIHSHING LEIGH WANG is a professor of psychometrics and statistics in the Quantitative and Mixed-Methods Research Methodologies PhD Program at the University of Cincinnati, OH, USA; email: *leigh.wang@uc.edu*. Her research interests include psychometric analysis, measurement and assessment, multivariate and multilevel modeling. She has published extensively on common fallacies in quantitative methodology, critical synthesis of large-scale assessment, and multilevel psychometric modeling.