Undergraduate geoscience education research: Evolution of an emerging field of discipline-based education research

Leilani A. Arthurs

University of Colorado at Boulder

Key words: discipline-based education research, geoscience education research, science education, undergraduate, college
Total Word Count: 12,451 (all text including cover page and captions)

Author Note

Leilani A. Arthurs, Department of Geological Sciences, University of Colorado at Boulder

The research and recommendations presented do not reflect the views of funding sources.

This work was begun while affiliated with the Department of Earth & Atmospheric Sciences at the University of Nebraska-Lincoln and the Center for Science, Math, & Computer Education at the University of Nebraska-Lincoln. This work was completed while affiliated with the Department of Geological Sciences at the University of Colorado at Boulder.

Correspondence concerning this article should be addressed to Campus Box 399; Boulder, CO 80309-0399.
Email: Leilani.Arthurs@Colorado.edu
Tel. #: 303-735-7752
Fax. #: 303-492-2606

**Undergraduate geoscience education research: Evolution of an emerging field of discipline-based education research**

**Abstract.** Discipline-based education research (DBER) conducted by faculty within geoscience departments can address identified needs in undergraduate geoscience education. This study explores the evolution of undergraduate geoscience education research (GER) from 1985 to 2016, primarily in terms of the types of published research and secondarily in terms of the insights this literature offers on the evolution of GER as a scholarly discipline. Stokes' (1997) quadrant model of research types is used as a theoretical framework for the former and Kuhn's (1970) model of disciplinary paradigm for the latter. An exploratory sequential mixed-methods approach to a systematic literature review of 1,760 articles is utilized. The period 1985-2000 is characterized by proto-research as evidenced by the abundance of instructive and informational education articles rather that research articles. From 2000 to 2011, GER underwent a growth period characterized by the presence of applied, use-inspired, and pure basic research. The period 2011-2016 appears to be a period of relative steady-state conditions in the normalized number of GER publications per year. Existing gaps in knowledge about geoscience education, the evident unfamiliarity with education and social science research methodologies among authors of GER articles, and efforts to build consensus about what GER is and how to conduct it suggest that GER is pre-paradigmatic or at a low paradigm state. That is, GER is an immature discipline as far as the evolution of a discipline goes. A path forward is proposed for the continued evolutionary growth of GER. This study provides new perspectives on the emergence of GER as a discipline that can be used as a basis for studies on cross-disciplinary DBER comparisons.

Keywords: discipline-based education research, geoscience education research, science education, undergraduate, college

## Introduction

The topic of this study is the evolution of research on undergraduate geoscience education from 1985 to 2016. “Geoscience” is an umbrella term that includes disciplines in which the primary subject of study is the Earth, such as: environmental science, geology, meteorology, and oceanography (American Geological Institute, 2009; Shea, 1995). An undergraduate education in geoscience is relevant in light of environmental issues such as freshwater quality and access (Adler, 2015; McDonald et al., 2016), natural resource production and consumption (Wiedmann et al., 2015; York et al., 2015), climate change and related impacts (Intergovernmental Panel on Climate Change, 2015; Stern & Dietz, 2015), and natural hazards and disasters (Sodhi, 2016). Such an education is essential for training future geoscientists (Gonzales & Keane, 2009; Liverman, 2009) and preparing future geoscience educators at all grade levels (Schoon, 1995; Shen et al., 2016). In addition, an undergraduate geoscience education can also promote the cultivation of more scientifically savvy citizens who are capable of applying geoscience knowledge to everyday decision making (Earth Science Literacy Initiative, 2010; Ratcliffe, 1997; Sadler & Zeidler, 2005).

Although an education in geoscience is relevant, the number of undergraduate geoscience degrees conferred per year, 2012-2015, in the US remained relatively constant (Wilson, 2016). This was consistently lower than other disciplines such as engineering and biology from at least 1973-2009 (Gonzales & Keane, 2009). In addition, US geoscience departments have difficulty retaining under-represented minorities (URMs) in particular. For example, geoscience graduates in 2015 comprised ~11% URMs (Wilson, 2016) and the US Census Bureau reported ~30% URMs in the national population (n.d.).

Reasons for low graduation rates can be explained, at least in part, by students' poor experiences in science, technology, engineering, and mathematics (STEM) courses associated with geoscience majors. Seymour & Hewitt (1997) found that experiences with poor teaching weighs heavily in students' decisions to leave science majors, including geoscience majors. A meta-analysis by Freeman et al. (2014) provides compelling evidence that active learning significantly improves student achievement in undergraduate STEM courses. Nevertheless, a challenge for geoscience instructors is translating general research-based instructional strategies (RBISs) to specific disciplinary contexts and instructional settings. RBISs are instructional strategies that promote active engagement and for which data show their positive impacts on student learning outcomes (Henderson & Dancy, 2007).

Discipline-based education research (DBER) and specifically geoscience education research (GER) can advance our understanding of students' and instructors' needs for specific discipline-based pedagogies and departmental/institutional factors that support geoscience instructor effectiveness and student achievement. It is worth noting, however, that there is currently a lack of consensus among the geoscience education community about what DBER and GER mean (St. John and McNeal, 2016). Nevertheless, both DBER and GER are used to connote research on undergraduate education in the National Research Council (NRC) report, *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (Singer et al., 2012), and they are used as such in this study. Additionally, as Singer (2013) states, a common feature of DBER is its "focus on undergraduate teaching and learning within a discipline, using a range of methods with deep grounding in the discipline's priorities, worldview, knowledge, and practices" (p. 769).

Three known literature reviews address research on geoscience education. Of them, only one focuses exclusively on the undergraduate level. None examines the longitudinal evolution of GER as a scholarly field. In his two-page review, Perkins' (2004) categorized articles into two arenas, (a) primary and secondary education and (b) undergraduate education. He described (a) what percentage of articles include assessments and (b) how each article's use of assessments rates on a scale from 0 to 3, with a 0-rating indicating failure to mention assessment at all and a 3-rating indicating the inclusion of a complete and thoughtful assessment. In King's (2008) traditional narrative literature review, he focuses mainly on pre-college geoscience education and discusses seven strands: (a) geoscientific thinking, (b) systems thinking, (c) spatial abilities, (d) geological time, (e) fieldwork, (f) misconceptions, and (g) teacher professional development. The rapid review by Piburn et al. (2011) was commissioned by the Board on Science Education of the National Research Council. It addresses eight distinct strands of research on undergraduate geoscience education: (a) conceptual frameworks, (b) introductory-level courses, (c) field-based courses, (d) affordances and constraints, (e) temporal thinking, (f) spatial visualization, (g) systems thinking, and (h) student affect (i.e., attitudes, values, and/or beliefs).

The present study complements the three aforementioned reviews. It aims to provide a longitudinal perspective on the evolution of research on undergraduate geoscience education from 1985 to 2016. To accomplish the aim of this study, a systematic review was conducted.

## What is a Systematic Review?

Literature reviews are broadly characterized as rapid reviews, traditional narrative reviews, and systematic reviews (HLWIKI International, 2017). However, no universally accepted standards for writing and reviewing literature reviews exist (Gough et al., 2012; Grant & Booth, 2009; Torraco, 2005). Despite the absence of universally accepted standards, there is agreement that a systematic review is a methodology (Dixon-Wood & Sutton, 2004). It is a methodology for utilizing primary/original studies as data to answer specific research questions. Rarely does a systematic review examine all aspects of primary/original studies. For example, quantitative evidence drawn from primary/original studies may be used to answer questions of effectiveness; whereas qualitative evidence may be used to answer questions of meaningfulness. Indeed, *what* aspects of primary/original studies are examined depends on the research question(s) driving a systematic review.

In terms of *how* the studies are examined, the methodology is based upon explicit, reproducible, and transparent methods to identify, critically appraise, and synthesize relevant primary/original studies. “A systematic review is a piece of research – it follows standard methods and stages … [and] … [l]ike other research, [it] requires following specific steps to minimize bias, the introduction of errors, and the possibility of drawing wrong conclusions” (Harden, 2010, p. 2). The methods used in the review process are made explicit so that, in principle, others with access to the same resources could undertake the review and reach generally the same conclusions.

To explore the evolution of GER as a scholarly discipline, this study utilizes the methodology of a systematic literature review and more specifically an integrative literature review as described by Torraco (2005). An integrative literature review “reviews, critiques, and synthesizes representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated” (Torraco, 2005, p 356). Common forms of synthesis include: a research agenda, an alternative model or conceptual framework, and a metatheory (Torraco, 2005). Following Harden’s (2010) recommendation that systematic reviews follow the format of research articles, this systematic review is presented following the format of social science and education research articles as outlined by Creswell (2014) and includes a description of the study's theoretical framework, methodology, methods, results, discussion, and conclusion.

## Research Questions

The evolution of research on geoscience education has not yet been systematically examined. The main question driving this study is: How did research on geoscience education evolve from 1985 to 2016? To constrain the scope of this study, the main question is addressed by answering the following research questions:

(1) What patterns in the types of articles published in the reviewed literature emerge? Also, what trends appear in the rates of GER publications over this time period?

(2) What topics are the subject of GER publications during this 31-year time period?

(3) How was the research undertaken during this period of time?

(4) How rigorous is the research published during this time period?

## Theoretical Framework

Stokes' (1997) quadrant model of research types is used as a theoretical framework to describe different types of published research on geoscience education. Kuhn's (1970) concept of disciplinary paradigm is used to describe the evolution of geoscience education research as a scholarly discipline. Together, the quadrant model and concept of disciplinary paradigm provide lenses through which to analyse and interpret the research articles appraised in this systematic review about how research on geoscience education has evolved from 1985 to 2016.

Stokes' (1997) quadrant model facilitates the conceptualization of different types of geoscience education research. Stokes argues research can be characterized according to the response to two questions, (a) is the research focused on fundamental knowledge for the sake of knowledge? and (b) is the research focused on applications? When the answer to the first question is “yes” and “no” to the second, Stokes categorizes it as pure basic research, naming the quadrant “Bohr’s quadrant.” When the answer to the first question is “no” and “yes” to the second question, he calls it pure applied research (i.e., “Edison’s quadrant”). When the answers to both questions are “yes,” he refers to it as use-inspired pure basic research (i.e., “Pasteur’s quadrant”). If the answers to both questions are “no,” then he refers to this quadrant as wissenschaft (i.e., German word for “science”) and opines it is a blank quadrant that can accommodate some future type of research.

The model of “disciplinary paradigm” that Kuhn (1970) initially developed and that Lodahl and Gordan (1972) later reconceptualised provides a useful framework for examining GER as a discipline. Kuhn (1970) conceptualized scientific disciplines as embracing a disciplinary paradigm when there is a high level of consensus in the theoretical structures and methodological approaches that define the discipline. Lodahl and Gordan (1972) stated, “the high consensus found in high paradigm fields … provides an accepted and shared vocabulary for discussing the content of the field” (p. 61). Kuhn’s paradigm development model asserts:

> [R]esearch activities and outcomes vary as a function of a field’s level of maturity: early-stage disciplines are expected to exhibit lower levels of research productivity.’ … [The model] argues that (a) some disciplines are more advanced than others, and (b) these differences affect the way that research is done. Communities with more developed paradigms have greater structure and predictability (Lodahl and Gordon, 1972), and fewer debates “over legitimate methods, problems, and standards of solution” (Kuhn, 1996, p. 48). In contrast, less mature fields have weaker levels of consensus among researchers. (Boyd et al., 2005, pp. 841-842)

## Methodology

The worldview (i.e., "a set of beliefs that guide action" [Guba, 1990, p. 17] that "researchers bring to inquiry" [Creswell, 2014, p. 35-36]), that grounds the design and conduct of this study is pragmatism. It is concerned with the application or intended consequences of the research study (Creswell, 2014). The intended application of this study is to support reflection on the evolution of research on geoscience education, inform a path for advancing the maturation of GER as a scholarly field, and provide some foundation for future cross-disciplinary comparisons between education research in geoscience and other STEM disciplines. Pragmatism also lends freedom of choice to the researcher in utilizing any and all available approaches that will help to solve a problem or answer a question (Creswell, 2014). As such, this study utilizes the methodology of a systematic review and more specifically an integrative literature review.

Torraco (2005) states most integrative literature reviews address either mature topics or emerging topics. In the former, they can reconceptualise the topic as it continues to develop and, in the latter, they can "lead to an initial or preliminary conceptualization of the topic (i.e., a new model or framework) rather than a reconceptualization of previous models" (Torraco, 2005, p. 357). This review explores the evolution of research on geoscience education. It synthesizes the literature as a conceptual model for describing the types of articles published in the corpus of literature reviewed, a juxtaposition of Stokes' quadrant model for GER studies, research questions for future study, and a proposed research agenda for advancing the field of GER.

## Methods

### Mixed Methods Approach to Systematic Review of Literature

An exploratory sequential mixed-methods approach was used (Figure 1). It began with a qualitative research phase of exploration and then used the results of that phase in the second quantitative phase (Creswell, 2009, 2014). In the first phase of this study, definitions and coding rubrics for quantitative data analysis were developed. In the second phase of the study, the definitions and rubrics were applied to the selected corpus of literature for review and quantified. The third phase involved mixing all data and interpreting them together.

**Phase 1: qualitative phase of study**

The decision-making process for how to code articles began with four author-posed questions. (a) What is research? (b) What needs in the area of geoscience education can be addressed via GER? (c) What are generally accepted components of social and education science research? (d) How can the rigor of GER articles be evaluated?

Three sources of information were used to define "research." The first, the Code of Federal Regulations, defines research as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge" (US Department of Health and Human Services, 2009, §46.102 Definitions, para. 4). In the second, research is recognized by the presence of a question or hypothesis to be investigated (St. John & McNeal, 2016). In the third, research can be classified into three types: applied research, basic research, and use-inspired research (Stokes, 1997).

Three sources were used to identify and aggregate already identified research needs in geoscience education: (a) *Bridges: Connecting research to education in the Earth sciences* (Committee of Geoscience Education in the Next Millennium, 2000), (b) *A new century for geoscience education research* (Piburn et al., 2011), and (c) *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (Singer et al., 2012). The needs described in these sources were compiled into a list and sorted using a constant comparative analysis (Charmaz, 2014; Corbin & Strauss, 2014). From this analysis, six main themes of needed research topics emerged: teaching methods, assessment tools, student cognition, student affect, minorities and recruitment, and departmental programs and curricula. See Supplementary Material for detailed rubric.

Creswell's (e.g., 2009, 2014) work on framing social and education sciences research form the basis for the coding rubric about *structural components of research design*. The structural components of particular interest included: worldview, theoretical framework, methodology, approach, and methods. See Supplementary Material for more details.

Lastly, rigor was assessed using two rubrics developed by Henderson et al. (2011) that were modified for this study. The first is a rubric about the *level of connections* and the second is about *level of evidence*. See Supplementary Material for detailed rubrics.

To obtain evidence of the trustworthiness of this study's findings, the rubrics were subjected to the external peer review (Shenton, 2004) of a long-time science education researcher external to the project. The peer review indicated that the rubrics are robust and well-grounded in the literature, thus providing evidence of credibility.

**Phase 2: quantitative phase of study**

The corpus of literature selected for this review is composed of peer-reviewed journal articles published in the *Journal of Geoscience Education* (*JGE*) from 1985 to 2016. Peer-reviewed research articles were selected for these reasons: (a) research articles are the very research products of GER this study aims to characterize and (b) they represent the highest quality data for a study aimed at characterizing research because they were peer reviewed. Articles published in the *JGE* were selected because, as Gough et al. (2012) state, systematic reviews aim to "identify a representative sample of studies" (p. 3) and *JGE* articles are considered representative for the following reasons: (a) *JGE* publishes the overwhelming majority of education studies on undergraduate geoscience education (in preparation), (b) *JGE* is the flagship journal for the National Association of Geoscience Teachers, and (c) *JGE* was published continuously during the period of interest (1985-2016). The starting year 1985 was selected because 31 years of data was potentially sufficient for answering the research questions. The year 1985 is also when the National Science Board began a year-long study into problems developing in US undergraduate science, mathematics, and engineering education (National Science Board's Task Committee on Undergraduate Science and Engineering Education, 1986).

A staged review (Torraco, 2005) was conducted. In the first stage, the abstracts for all of the articles published in the *JGE* (1985-2016) were read in published hard copy form to determine whether they dealt with undergraduate geoscience education research and to identify their research question or hypothesis (if present). During this stage, a total of 1,760 articles were reviewed. Of these, 167 articles met the inclusion criteria for the second stage of review (i.e., question- or hypothesis-driven research on undergraduate geoscience education).

In the second stage, rubrics developed in Phase 1 were applied to these 167 articles. A double-coding process (Krefting, 1991) was applied to all articles selected for the second stage of review two months and again five months after the initial coding. With greater than 96% agreement between the three coding iterations, the process yielded little to no discrepancies in the codes assigned to each article, providing evidence of reliability. Although double-coding is an accepted method for checking reliability (Krefting, 1991), interrater comparisons were conducted as an additional check. A second rater coded 12% of the articles, with greater than 95% initial interrater agreement. Discrepancies in codes were discussed and resolved with 100% final interrater agreement. The results of both the double-coding and interrater checks provide evidence of the results' reliability.

### Phase 3: mixing phase of study

In this phase, data were analyzed and interpreted together. The occurrences of coded information were quantified and plotted. To enhance trustworthiness, the results were subjected to a process of triangulation whereby they were compared against other findings available in the literature (Shenton, 2004), such as the previously mentioned earlier literature reviews.

## Results

### Patterns and Trends in Types of Articles

The first research question is: *What patterns in the types of articles published in the reviewed literature emerge? Also, what trends appear in the rates of GER publications from 1985-2017?* During the first stage of review, three patterns of articles were identified: (a) instructive "how-to" articles, (b) informational "we-did-this" articles, and (c) hypothesis- or question-driven research articles. The first two patterns were not determined *a priori* and emerged from the search and review process. The instructive articles contain no research question or hypothesis and no evidence of the effectiveness of the activity described. The informational articles usually do not contain a research question or hypothesis and contain weak to no evidence of the effectiveness of the activity or program that is described. The research articles contain a research question or hypothesis and stronger evidence for the claims made. Examples of articles binned into these three categories are provided in Supplementary Material Table S2. Given the focus of the present study is on research articles, the second stage of review involved further appraisal of only the research articles.

From 1985 to 2016, the publication of GER articles changed noticeably (Figure 2). The period 1985-2000 is characterized by GER articles that represent less than 10% of the total articles published per year. It is more dominated by instructive and informational articles. The period 2000-2011 is characterized as a period of growth, as represented by the positive trend ($R^2$=0.57) in the percentage of GER articles per year. It is characterized by all three contemporary forms of research (applied, basic, and use-inspired). The period 2010-2016 appears to be a period of levelling off in GER publications ($R^2$=0.00), with GER articles representing approximately 20-30% of the total articles published each year in this period. However, much like evaluating the state of peak oil (Maggio & Cacciola, 2012), a retrospective look at this period years from now will provide the necessary information to determine how real this apparent levelling off is. It may very well be part of a continued period of growth.

Looking more closely at the research, the types of GER studies published also changed over time (Supplementary Material Figure S1). The period 1985-2000 is dominated by applied research, with some basic research between 1992 and 1996. In 2001, use-inspired research made its first appearance. The period 2001-2016 is characterized by a mix of all three types of research (i.e., applied, use-inspired, and basic), with basic research being dominant.

**Topics of Research on Geoscience Education**

The second research question is: *What topics are the subject of research on geoscience education from 1985-2016?* Education research in geoscience is needed to facilitate the teaching and learning of geoscience (Manduca et al., 2002). Using the results of Phase 1, articles were coded with six themes of research topics in mind. While the purpose of many literature reviews is to synthesize the state of knowledge on a single or a few specific topics of research (e.g., Brotman & Moore, 2008), the purpose of this review is to provide a longitudinal perspective on the evolution of GER as a discipline over time, and the topics researched represent only a part of the broader story. Thus, a distillation of the six emergent themes of research topics is presented to provide a synopsis of research progress made in these areas rather than to provide a comprehensive review of all six research themes. The research themes with the greatest to the least representation in terms of percentage of GER articles published per year, 1985-2016, are: teaching methods, student cognition, assessment tools, student affect, minorities and recruitment, and departmental curricula and programs. Supplementary Material Figure S2 illustrates the changing presence of these six themes over time.

From about 1994 to 2016, there was a fairly continuous emphasis in GER articles on teaching methods. The impact of teaching methods in a variety of undergraduate educational settings are described in about 40% of the articles in the second stage of review. These articles show the effectiveness of teaching methods, particularly in field, lab, and introductory-level course settings (e.g., Feig, 2010; Giorgis, 2015; Hodder, 2001). A number of articles describe instructional strategies that promote student learning such as: (a) cooperative learning; (b) computer-, technology-, or WEB-based activities; (c) inquiry-based learning; (d) just-in-time teaching, (e) collaborative exams; (f) peer instruction; and (g) lecture tutorials. The articles describe RBISs, thus providing ample examples of teaching that others may incorporate into their own classes. However, the majority of these studies are single case studies with single instructors. Therefore, there is a need to understand how transferable the described RBISs are to other settings with other instructors. This broader need prompts the following research questions:

- Building on the work of MacDonald et al. (2005), how widespread is the implementation of RBISs in undergraduate geoscience courses?
- How replicable are the instructional strategies discussed in the context of single case studies to other instructional settings with different instructors, and what kind of supports do faculty need to implement RBISs with fidelity?

Although these two questions have been asked previously, they are underresearched. A related question that has not yet been investigated in the body of literature reviewed is:

- With what degree of fidelity are RBISs implemented when they are used?

Answering this question is of special importance for understanding why the adoption of RBISs works more or less in different settings with different student populations.

We now know student learning can be enhanced using effective teaching methods (Freeman et al., 2014) and measuring student learning depends on using appropriate assessment tools. Although the development and testing of teaching methods was ubiquitous and fairly continuous (1994-2016), there was a notable lack of similar emphasis on the development and evaluation of assessment tools (Supplementary Material Figure S2a). Articles about assessment tools first arose in 2001. They are described in about 12% of articles in the second stage of review. These tools address a variety of constructs including: factual knowledge, conceptual understanding, attitudes, learning styles, views on the nature of science, stewardship, scientific writing, impact of research experiences, and recruiting effectiveness (e.g., Arthurs et al., 2015; Hanks et al., 2007; Jolley et al., 2012; McNeal et al, 2014). This study shows a relatively small number of assessments have been developed mainly in response to local needs. As such, their development presents potential limitations to their broader validity and reliability in contexts beyond the local setting in which they were developed. This observation raises the following research questions:

- What are the learning objectives for the same course (e.g., physical geology) taught by different instructors and/or at different institutions? To what extent are they similar?
- How can available assessment tools be further developed for wider use to assess learning between the same courses taught by different instructors and/or at different institutions?

- What are the challenges to developing assessment tools that can be used more widely for comparative studies, and how can these challenges be overcome?

These research questions are aligned with previously identified needs for assessment tools and their use (CGENM, 2000) and to "identify and measure appropriate learning objectives" (Singer et al, 2012. p. 120). Science literacy documents (Carley et al., 2013; Earth Science Literacy Initiative, 2010; Johnson et al., 2009; US Global Change Research Program, 2009) and a report called *Future of Undergraduate Geoscience Education* (Mosher et al., 2014) may be useful resources to draw on for the development of shared learning objectives and assessment tools. Readers can access information on some assessment tools at a website released in summer 2017 called *Geoscience Education Researcher Toolbox*.

The development of assessment tools that measure student learning should also be informed by and benefit from basic research in areas of, for example, student cognition and student affect. Research on student cognition was generally ongoing since about 2004 (Supplementary Material Figure S2b). Student cognition is the primary subject of study of 32% of the articles reviewed in the second stage of review. These studies target students' thinking on a variety of geoscience concepts, spatial skills, argumentation and reasoning skills, systems thinking skills, geologic time, and uncertainty (e.g., Black, 2005; Kusnick, 2002; Sibley, 2009). Most, however, were conducted in introductory-level geoscience courses at a single location with nominal cultural diversity if any at all. The following research questions emerge from these observations:

- What types of alternate conceptions do students have about Earth processes and phenomena discussed in not only introductory-level courses but also in advanced undergraduate courses?

- To what extent do students' alternate conceptions transcend cultural and geographic backgrounds? To what extent are these alternate conceptions more specific to certain demographic considerations, such as cultural affiliations and/or geographic locations?
- How can identified alternate conceptions be effectively incorporated into geoscience course instruction to facilitate metacognitive skill development and conceptual change from more novice-like ways of thinking to more expert-like ways of thinking?

These questions are aligned with previously highlighted needs to understand the details of students' prior knowledge and alternate conceptions (CGENM, 2000; Singer et al., 2012) and understand "the nature and development of expertise in a discipline" (Singer et al, 2012, p. 186). The appraised literature addresses these needs in the context of relatively few conceptual areas and narrowly defined parts of those areas. Thus, relative to the breadth and depth of geoscience concepts as well as the prior knowledge and alternate conceptions that students bring to the geoscience learning environment, the research theme of student cognition and these specific research questions remain highly relevant in advancing a research agenda for GER.

While student cognition is involved in learning, so too is student affect (i.e., feelings, emotions, and beliefs). However, the importance of student affect has only more recently been recognized. In this review, only about 3% of the articles in the second stage of review had a primary focus on student affect (Supplementary Material Figure S2b). These articles address students' levels of motivation in introductory-level geoscience courses, motivation to study for introductory geology course exams, the experiences that led to persistence in field geology, and the role of fun in learning and research (e.g. Dykas et al., 2016; Gilbert et al., 2012; Jarrett et al., 2010). Although little is known about student affect as it pertains to learning geoscience, van der

Hoeven Kraft et al. (2011) consider available information and propose a new model for teaching geoscience that attends to the role of student affect in learning.

This review finds student affect is understudied. Existing studies show students commonly hold negative attitudes towards geoscience. Given that we now know student learning is intimately connected to student affect (Entwistle & Ramsden, 2015; Fink, 2013; Mega et al., 2014), research on student affect is all the more important. Observations made while reviewing the literature produced the following questions to help guide research on student affect:

- How can students' perceptions of geoscience as a single course, as an undergraduate major, and as a field of potential employment be improved?
- Where and how do students with positive attitudes and feelings toward geoscience develop those attitudes? How can such attitudes be cultivated in geoscience courses?
- How can students' attitudes and feelings toward geoscience and learning geoscience be incorporated into geoscience course instruction and curricula in pragmatic terms?

Although research in the area of student affect is relatively sparse (1985-2016), there were also few studies about minorities and recruitment (Supplementary Material Figure 2c). About half of them characterize the factors that lead URMs and/or non-URMs to choose a geoscience major or not (e.g., Levine et al., 2007). The remaining articles describe interventions aimed at recruiting and/or retaining students in geoscience (e.g., April, 1994). This study reveals very few studies contribute to our understanding of the experiences URMs have in the geoscience pipeline to and from the college level, what attracts/repels them, and what they need to thrive in the pipeline. Also, the category of "URM" represents an aggregate of very different populations with different backgrounds and potentially different needs. Observations made while reviewing the literature support a place for the following questions in a research agenda for GER:

- What are strategies for increasing URM students' awareness of geoscience as a potential undergraduate major and field of future employment? How can undergraduate geoscience programs utilize local resources to enhance recruitment of URMs into geoscience majors?
- What types of socio-political and cultural awareness of different URM populations are useful for drawing connections between URM students and geoscience, which may aid in attracting them and helping them to thrive?

There were similarly few articles about departmental curricula and programs. They address degrees proffered, rankings, and the possibility of developing national curricular norms (e.g., Drummond & Markin, 2008; Ulanski, 1995). Overall, this review finds very few studies address geoscience departments' curricula and programs and much remains to be learned. Given geoscience departments play a critical role in promoting active learning in geoscience, the absence of studies on them, spur the following research questions:

- To what extent do geoscience departments encourage, support, and reward the use of RBISs to facilitate active learning in their courses?
- What patterns or models of leadership exist in geoscience departments with respect to valuing and proactively supporting RBISs in course instruction?
- From departments with a record of successfully transforming departmental culture to not only value but also proactively support RBISs in course instruction, what lessons can be learned that are potentially applicable to other departments?

The present review of the literature reveals different degrees of progress in addressing already identified needs in geoscience education that fall under the six themes of research topics. The review also reveals several notable gaps in the research. For example, although the cognitive and affective domains are discussed, the psychomotor domain is not. In addition, almost all of the articles do not address URMs in the learning environment. Finally, although student learning is a focus of the majority of the articles, instructor and department chair learning is absent. These observations prompt the following research questions for inclusion in an agenda for GER:

- How do considerations of the psychomotor domain influence students' learning about geoscience in classroom, laboratory, and field settings?
- What patterns of socio-cultural learning environments exist in geoscience courses, with and without URMs?
- Building on the professional development work of Manduca et al. (2017), what do instructors and departmental heads need in order to create more inclusive and effective geoscience learning communities in their classrooms and departments?

**How GER was Undertaken**

The third research question is: *How was the research on undergraduate geoscience education undertaken from 1985 to 2016?* Recall that Creswell's (2014) components of social and education sciences research articles are used to characterize how research on geoscience education is undertaken, particularly with respect to worldview, theoretical framework, methodology, approach, and methods. None of the articles appraised discuss a worldview. The overwhelming majority of articles do not explicitly specify a theoretical framework or methodology. Only about 12% of the articles explicitly state a methodology. Although not explicit, most studies used the methodology of a single case study. This is consistent with Piburn and colleagues' (2011) finding of the same. Although generally not explicit, the methodological approach could be derived from an examination of the methods described.

The earliest published GER articles utilized approaches quantitative in nature. In 1992, mixed methods approaches were introduced. From 1992 to 2001, GER studies were conducted using quantitative approaches or mixed methods approaches. In 2002, GER studies began being conducted using qualitative methodological approaches (Supplementary Material Figure S3).

In terms of GER methods, these too have changed over time (Supplementary Material Figure S4) and vary. This finding is consistent with earlier observations that "[t]he methods DBER scholars use are as diverse as the research questions they investigate" (Singer et al., 2012, p. 50). Methods discussed in GER studies include, for example: tests, surveys, coursework, observations, reviews, video recording, social network analysis, and GPS tracking.

In terms of the settings in which GER studies are undertaken, about 3% of studies were conducted in the context of large-enrolment courses, about 53% in the context of introductory- or lower-level courses, and ~15% in upper-level courses. In terms of the subjects of study, about 62% of studies specify some aspect of participant demographics. There is a notable absence of and/or lack of reporting of URMS in geoscience courses. Finally, Singer et al. (2012) noted the need for GER studies that are multi-instructor, multi-institutional, and longitudinal in nature. Of the studies reviewed about 10% were multi-instructor, less than 20% were multi-institutional, and less than 1% were longitudinal. Thus, the need for such studies remains.

**Rigor of GER**

The fourth research question is: *How rigorous is the research on geoscience education published 1985-2016?* Rigor was evaluated in terms of connections made to other literature and evidence provided to support claims made. Rigor was highly variable in terms of both level of connection and level of evidence from 1985 to 2000 (Supplementary Material Figure S5). During 2000-2011, there was an apparent overall improvement in the level of connections while the level of evidence remained somewhat constant. During 2011-2016, both the level of connections and the level of evidence were better than in 200-2011 overall.

These results expand on Perkin's findings from a review of "slightly more than 300 articles published in the *JGE* [1998-2004] to determine what percentage of articles included good project assessments" (2004, p. 113). His two-page review revealed that 72% of the articles about undergraduate geoscience education either "failed to mention anything about how the project affected student/participant learning (21%) … [or] …included comments or assertions about learning but gave no evidence in support (51%)" (Perkins, 2004, p. 133). The present study shows *JGE* articles on GER published in the same period (1998-2004) are characterized by averages between Level 1.5 and Level 2.5 for their level of evidence, suggesting articles are adequate to strong in their treatment of evidence that supports their claims.

The discrepancy between Perkins' findings and the present study is explained by differences in article selection criteria. Perkins reviewed all *JGE* articles in the time period of interest for his study. In contrast, this study reviews only *JGE* articles that meet the criteria for being a research article. Research articles are more likely to use evidence to support claims than the instructive and informational types of articles published in the *JGE*.

## Discussion

### Types of Research on Geoscience Education

Stokes' quadrant model of scientific research is juxtaposed with examples of different types of GER studies in Figure 3. GER shown in the quadrant model do not represent exhaustive examples for each research type. Instead, they serve to conceptualize how Stokes' model is applied to GER. Stokes' types of research are mapped alongside the three aforementioned emergent categories of *JGE* articles (i.e., research, instructive, and informational) and their relative positions on a spectrum for the strength of evidence for claims made (Figure 4).

The German concept of "wissenschaft" involves thorough descriptions such as, for example, detailed drawings and descriptions that early naturalists made of flora and fauna, with no real intention to expand scientific knowledge or to develop applications from these drawings and descriptions (H. Doebel, personal communication, September 22, 2016). In the context of GER, it is proposed that Stokes' notion of wissenschaft represents *proto-research* that is exemplified by instructive and informational articles (Supplementary Material Table S2). In parallel terms, whereas the paradigm of natural science research has roots in the practices of naturalists, the paradigm of GER has roots in the activities of instructional practitioners who instruct readers how to do certain activities (instructive articles) and who inform readers about courses or programs they carried out (informational articles).

**How Research on Geoscience Education was Designed**

Creswell, an expert on the conduct of social and education sciences research, states social and education sciences research should acknowledge the researchers' worldview, specify a theoretical framework, characterize the methodology and approach, and describe the methods (2014). The notable absence of social and education science world views, methodologies, theoretical frameworks, approaches, and methods in the appraised articles suggests the need for greater familiarity with them among GER scholars. This finding also corroborates the finding that "[m]any DBER studies either do not situate themselves in a broader theoretical frame, or do not explicitly define that frame" (Singer et al., 2012, p. 52).

The corpus of literature reviewed reveals an already very rich collection of single case studies. This is not to say that there is no longer need for single case studies. The point is there exists a collection of single case studies that be drawn upon in studies that are, for example, larger-scale and/or or meta-analytical in nature.

**Placing Observed Trends in a Historical Context**

The inevitable question of what explains the observed longitudinal trends in research on geoscience education cannot be definitively answered. Nevertheless, these findings can be at least partially positioned in a broader historical context by describing factors that potentially help to explain them. Possible factors include, but are not limited to: (i) changes in the *JGE* Editors' plans for the journal, (ii) community interest in research on geoscience education, and (iii) the existence of other publication venues. Many of the references used to inform this discussion are archived on a webpage called *Twenty five years of progress in geoscience education*.

The period 1985-2000, under Shea's *JGE* editorial leadership, is characterized by an abundance of instructive and informational articles rather than research articles. The dominance of such articles during this time is consistent with the interest that the nation had in geoscience education. For example, during this time, at least 11 different reports (disseminated as publications, conference proceedings, or webpages) outline problems in undergraduate science education in the US and urge reforms in undergraduate science education. It is likely that the *JGE*'s instructive and informational articles helped to meet the needs stated in these reports.

The observed growth in research on geoscience education in 2000-2011 (Figure 2) took place under the helm of two different *JGE* editors, Drummond (2001-2008) and Libarkin (2009-2011). At least two notable reports released a few years just before this time period likely helped plant the seeds for increased attention to research on geoscience education. The first is titled, *Geoscience education: A recommended strategy* (NSF, 1997) and the second is titled, *Bridges: Connecting research and education in the Earth sciences* (Mogk, 2000). The reports produced by working groups immediately before or during the period of growth in geoscience education research reflect a level of community interest than may help to explain the observed growth.

The apparent relative steady-state conditions in growth from 2011to 2016 occurred during St. John's time as *JGE* Editor. During this period, stakeholders of geoscience education exhibited continued interest in GER, including in the *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics* report produced by the President's Council of Advisors for Science and Technology in 2012, the National Research Council's 2012 *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (Singer et al., 2012), *Rising Above the gathering storm: Energizing and employing America for a brighter economic future* (National Academy of Sciences et al., 2007), and *Rising above the gathering storm, revisited: Rapidly approaching category 5* (National Academy of Sciences et al., 2010).

During the period of overall growth, there is a dip in 2012 (Figure 2). The dip coincides with generally higher levels of rigor (Supplementary Material Figure S5), which may be attributed to changes in *JGE* manuscript guidelines (Libarkin et al., 2009; St. John et al., 2013; St. John et al., 2016). While certainly not accounting for all possible competing publishing venues, it is worth noting that the dip in 2012 coincides with the publication of a book titled, *Earth and mind II: A synthesis of research on thinking and learning in the geosciences* (Kastens & Manduca, eds., 2012). The precursor to this volume (Manduca & Mogk, eds., 2006) also coincides with a low point in growth in 2006 (Figure 2). While certainly not definitive, it is possible changes in manuscript guidelines and competing publication venues partly explain the apparent dip observed between 2012 and 2016.

**Status of GER as a Discipline**

Kuhn (1970, 1996) conceptualized a mature discipline as characterized by a high level of consensus in theoretical structures and methodological approaches that define the discipline. The present study reveals an overall absence of and/or very low level of consensus on theoretical frameworks and methodological approaches for conducting research on geoscience education. This paucity provides evidence for the relative immaturity of GER as a scholarly field in terms of Kuhn's model of paradigm development. Additionally, although research progress is being made, much remains to be understood in geoscience education. Finally, the apparent improvement in the levels of connections and of evidence over the 31-year period is interpreted as evidence of the maturation process of this relatively new area of DBER.

Lodahl and Gordan (1972) expanded on Kuhn's ideas and noted that high paradigm disciplines have an "accepted and shared vocabulary for discussing the content [of that discipline]" (p. 61). For example, the name of a discipline should carry the same essential meaning among members of that discipline. Although the meanings for DBER and GER for the purposes of this study are taken from the NRC's report (Singer et al., 2012), there is nevertheless an active discussion in the geoscience education community about what constitutes GER (St. John and McNeal, 2016). This lack of consensus provides additional evidence of the pre- or low-paradigmatic state of GER as a discipline.

According to Pfeffer (1993), Lodahl and Gordan's (1972) operationalization of paradigm development refers to disciplinary norms and agreements about which research questions will advance knowledge in a field and the methods appropriate for addressing those questions. Complimentary to this, Cole (1983) states research progress within a discipline occurs in times when the disciplinary community adheres to a paradigm and builds upon the work of others. Editorials and columns published in the *JGE* provide evidence of this journal's role in actively establishing norms for the GER community. The cumulative evidence supports the idea that GER is either a pre-paradigmatic or low-paradigm discipline.

**Future of GER and DBER**

This study provides a means for reflection on the past of GER that naturally leads to reflecting on the future of GER in particular and DBER more broadly. It finds that GER is currently pre-paradigmatic or at a low paradigm state in terms of its growth evolution as a scholarly discipline. Thus, reflection on the future of GER involves a vision for moving GER towards a high paradigm state and supporting its growth towards a more established discipline. There are several possible avenues for doing this, but the avenue focused on here builds primarily on the discussion of the appraised products of GER (i.e., peer-reviewed research articles). In particular, the avenue is one of continued research that addresses identified needs for improving undergraduate geoscience education to well prepare future geoscientists, Earth science educators, and scientifically savvy citizens. This avenue is defined by what topics to research, how to conduct the research, and how to report the research.

Others such as the CGEM (2000), Piburn and colleagues (2011), and Singer and colleagues (2012) have previously identified needed research topics to help inform improvements in geoscience education in particular and/or STEM education more generally. Within the six themes of research topics, this study finds there is need for continued research on topics for which research now exists. It finds that some identified topics have not yet begun to be researched. Finally, it reveals there are needed research topics that have not been previously identified in the literature reviewed. These topics would fall under the themes of: "student psychomotor," "socio-cultural interactions," and "instructor development." GER's nascent stage of development lends itself to a broad array of research topics that remain largely open and that could benefit from a coherent agenda for research on geoscience education.

To advance the evolution of GER as a discipline, researchers can also actively attend to how they conduct their research. In particular, they can attend to the structural components of research design by being grounded in a worldview and theoretical framework. GER scholars can also be explicit about the methodology used and why it was selected. Doing so conforms to the practices of social and education sciences research (Creswell, 2009, 2014). It also facilitates conversations that can help establish "an accepted and shared vocabulary for discussing the content of the field" (Lodahl and Gordan, 1972, p. 61), a quality of high paradigm disciplines.

In addition, GER scholars can continue their upward trajectory in improving the rigor in their reporting by attending to the level of connections made to other research and to the level of evidence used to support their claims. This study finds the overall rigor of reporting varied over time and generally improved (Supplementary Material Figure S5). While improvements were observed, it is worth noting relatively low bars were set for what constitutes the highest levels (see Supplementary Material for rubric details). This is consistent with the original rubrics by Henderson and colleagues (2011), and it is literally possible to "raise the bar." The rigor of reported GER studies is expected to increase in the future. Thus, for literature reviews that include the past studies reviewed herein (1985-2016) *and* studies conducted after 2016 (e.g., 2017-2048), raising the bar in the rubrics may be useful for further discriminating between the levels in order to avoid anticipated clumping at what is currently set for the highest levels.

Thus far, the discussion about a vision for how to move the discipline of GER towards a higher paradigm state focused on what individual researchers or small groups of researchers could do independently (i.e., deciding what topics to research, how to conduct the research, and how to report the research). The vision, however, also includes disciplinary cultural norms embraced within the disciplines of GER specifically and DBER more broadly. Drawing upon the literature appraised and other literature read for this study, the envisioned disciplinary cultural norms for conducting research in these emerging disciplines include: systematicity, collaboration, and inclusion.

Systematicity in research would be evidenced by, for example: (a) the sequence in which research topics are addressed, (b) new research explicitly building on past research (e.g., levels of connections), and (c) large-scale projects being informed by results of local or small-scale studies. For instance, in terms of sequencing, basic research to define geoscience learning objectives can be used to narrow the scope of basic research to investigate cognitive, affective, psychomotor, and socio-cultural hindrances to achieving these defined learning objectives. Learning objectives can, in turn, inform use-inspired research such as the development of assessment tools to measure learning. These strands of research are aligned with the need for DBER to “help identify and measure appropriate learning objectives and instructional approaches that advance students towards those objectives” (Singer et al., 2012, p. 2).

Continuing with our example of systematicity, deeper understandings about student learning and instructional practice obtained through the aforementioned research strands can then provide a basis for applied research in the evaluation of education/training pathways for current and future geoscience faculty to learn about RBISs. Either in sequence or in tandem to this strand of research, basic research can be pursued to identify obstacles to the effective implementation of RBISs and course design principles in geoscience courses. The results of the two aforementioned strands of research can then aid in the design of use-inspired and applied research to develop and evaluate ways that support geoscience instructors and departments in effective implementation of course design principles and RBISs. This systematic approach is illustrated in a conceptual model for a proposed research agenda for GER (Supplementary Material Figure S6).

The proposed GER agenda is one that can support a higher level of systematicity and greater coherence to the nascent field of GER. In pragmatic terms, however, the areas of research outlined in the agenda can be pursued independent of the "steps" in the model (Supplementary Material Figure S6). Thus, the proposed research agenda is flexible enough to address existing calls to action in undergraduate geoscience education in particular and undergraduate STEM education in general (Handelsman & Brown, 2016; President's Council of Advisors on Science and Technology, 2012) by highlighting different areas of research that might be approached sequentially (or not) and/or simultaneously. The proposed research agenda is a tool that can help guide the evolutionary growth of GER. In addition, it is a tool that can aid in cross-disciplinary studies between DBER disciplines as, for example, an overall framework for cross-disciplinary comparisons and/or to help narrow the scope of cross-disciplinary comparisons.

Collaboration is the second envisioned disciplinary norm. Collaboration in research is pictured within the geoscience education community and in partnership with non-DBER disciplines and other DBER disciplines. Collaboration within the geoscience education community would be evidenced through, for example, multi-instructor studies within the same department, multi-department studies within the same institution, and multi-institution studies. These types of studies would support the GER community of scholars in expanding upon the single case studies that characterize the majority of GER studies appraised. Also, such studies would meet an identified need for "cohort studies" (St. John and McNeal, 2016, para. 4).

A collaborative approach with non-DBER disciplines, such as cognitive science and cognitive psychology, can facilitate meeting the need for GER studies to be more deeply grounded in theoretical frameworks and be more connected to the extant research of complimentary fields (Arthurs, 2018; Singer et al., 2012). A collaborative approach between

GER and other DBER disciplines can (a) provide opportunities for GER scholars to learn from more established sister-DBER disciplines, such as Physics Education Research (Singer et al, 2012); (b) broaden and deepen our understanding of STEM concepts and skills that transcend discipline (e.g., spatial, temporal, and systems thinking are not unique to geoscience and are relevant in astronomy, biology, and chemistry); and (c) provide an evidentiary basis for what may be unique about teaching and learning in different STEM disciplines. Furthermore, they could (d) promote multi-disciplinary DBER conversations and interactions that would facilitate the growth of not only the GER discipline but also facilitate a more catholic DBER discipline with shared vocabularies, norms, practices, methodologies, and methods. Finally, such studies could (e) yield larger data sets and lead to more generalizable conclusions, which could in turn generate new multi-disciplinary DBER research questions.

Inclusion is the final envisioned disciplinary norm. Here, inclusion refers to *inclusive engagement* in the research enterprise. Notably, the notion of inclusion here includes but extends beyond traditional associations with the term "diversity." Specifically, the notion of inclusive engagement is inspired by the work of education and learning sciences researchers who use the design-based methodology (Barab, 2006) for their studies. In such studies, research is conducted in naturalistic settings (i.e., in actual teaching and learning environments instead of labs) created by the researcher. However, it is possible for researchers to partner with other practitioners to do this. They can work together to develop, implement, and/or evaluate aspects of basic, use-inspired, and applied research. Importantly, the vision of inclusive engagement here also involves practitioners in the reporting of that research. That is, partnering practitioners may contribute to the larger GER and DBER enterprises as co-authors or co-generators of new knowledge and new perspectives. These kinds of inclusive engagement may promote a sense of ownership in pedagogical research and promote buy-in of the findings. Furthermore, these practitioners work in institutions of higher education and may receive departmental recognition for their involvement in and/or co-authorship of GER studies. Inclusive engagement of practitioners in these ways is a possible change mechanism for transforming undergraduate STEM education that has not yet been the subject of study.

In summary, reflection on the future of GER and DBER leads to a vision for conducting GER that: (a) attends to identified needs in geoscience education, (b) addresses the structural components of research design characteristic of social and education sciences research, and (c) is rigorous in its reporting of GER studies by making connections to other research and providing evidence that supports claims. This vision also includes cultural norms for the disciplines of GER specifically and DBER more broadly. The norms are ones that embrace the conduct of research in ways that are systematic, collaborative, and inclusive. Approaching the conduct of research in all these ways is one avenue for further advancing the evolutionary growth of GER (and DBER) toward a higher paradigm state.

## Limitations

Potential limitations of this study include (i) the methodology (literature review) used to explore the evolution of GER over time and (ii) the relatively short time period (31 years) used to examine the longitudinal evolution of a discipline. In terms of the methodology, certainly, the study of the evolution of a discipline can entail much more than a literature review of the research articles that the discipline produces. For example, the maturity of a discipline could be characterized by the number of institutions that proffer advanced degrees in the discipline or the number and types of funding sources that support research in that discipline (Pfeffer, 1993). Nevertheless, a review of GER articles utilizes the very products emerging from the discipline and, therefore, provides a credible grounding from which to evaluate the evolution of GER studies and of GER as a discipline. With respect to the time period selected for the study (1985-2016), 31 years may be insufficient to evaluate the evolution of every discipline. Nevertheless, this time period captures at least two distinctive phases (one of relative non-existence and one of growth) and one tentative steady-state phase (which may later prove to be part of the growth phase) in the development of GER (Figure 2). This suggests the 31-year interval is a critical time in GER's evolution and it is sufficient to answer this study's research questions.

## Conclusions

This study provides evidence that the field of GER is a relatively nascent discipline and there remains tremendous potential to address identified needs in undergraduate geoscience education. New perspectives developed through this study include: (a) categories for the types of articles on geoscience education published in the *JGE* from 1985 to 2016, (b) juxtaposition of Stokes' (1997) quadrant model of research types with different types of articles or studies on geoscience education (Figure 3), (c) superposition of Stokes' (1997) types of research onto the categories of articles published in the *JGE* (Figure 4), and (d) interpretation of the state of the evolution of GER as an emerging discipline from the perspective of disciplinary paradigm development (Kuhn, 1970, 1996; Lodahl & Gordan, 1972). Consistent with the syntheses produced via integrative reviews (Torraco, 2005), the synthesis of the reviewed literature here takes the forms of (a) an account of longitudinal changes in the research on geoscience education and (b) a proposed research agenda. Although publishing venues for discipline-based education research (DBER) tend to be discipline specific (Singer, 2013), such as GER articles published in *JGE*, Singer (2013) states that the *Journal for Research in Science Teaching* (*JRST*) "offers an opportunity for cross-fertilization of ideas between disciplines" (p. 769) Although *JRST* contains a dearth of GER articles (Lewis & Baker, 2010), the findings of this study can provide support for future studies not only in the area of GER, but also more broadly in the DBER undertakings of other disciplines. It is, for example, conceivable that the rubrics and proposed research agenda could be applied to studies whose unit of analysis extends beyond a single DBER discipline and involves cross-disciplinary comparisons between other disciplines with DBER efforts.

**References**

Adler, R. W. (2015). US Environmental Protection Agency's new waters of the United States rule: Connecting law and science. Freshwater Science, 34(4), 1595-1600.

American Geological Institute. (2009, February). Status of the geoscience workforce 2009. Retrieved from http://www.americangeosciences.org/workforce/reports

April, R. H. (1994). An NSF-Funded Curriculum Initiative Designed to Attract Minority Students to Geology. Journal of Geological Education, 42(5), 447-451.

Arthurs, L., Hsia, J. F., & Schweinle, W. (2015). The oceanography concept inventory: A semicustomizable assessment for measuring student understanding of oceanography. Journal of Geoscience Education, 63(4), 310-322.

Arthurs, L. (2018). How explicit is the cognitive science foundation of geoscience education research? A study of syntactical units in JGE articles. Journal of Geoscience Education, 66(1), 77-91

Black, A. A. (2005). Spatial ability and earth science conceptual understanding. Journal of Geoscience Education, 53(4), 402-414.

Barab, S. (2006). Design-based research: A methodological toolkit for the learning scientist. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 153-169). New York, NY: Cambridge University Press.

Boyd, B. K., Finkelstein, S., & Gove, S. (2005). How advanced is the strategy paradigm? The role of particularism and universalism in shaping research outcomes. Strategic Management Journal, 26(9), 841-854.

Brotman, J. S., & Moore, F. M. (2008). Girls and science: A review of four themes in the science education literature. Journal of research in science teaching, 45(9), 971-1002.

Carley, S., Chen, R., Halversen, C., Jacobson, M., Livingston, C., Matsumoto G., Payne, D., Paytan, A., Schoedinger, S., Strang, C., Tran, L. U., Tuddenham, P., Whitley, L., & Wilson. S. (2013, March). Ocean literacy: The essential principles and fundamental concepts of ocean sciences for learners of all ages Version 2. Retrieved from http://www.coexploration.org/oceanliteracy/documents/OceanLitChart.pdf

Charmaz, K. (2014). Constructing grounded theory. Sage.

Cole, S. (1983). The hierarchy of the sciences?. American Journal of Sociology, 111-139.

Committee of Geoscience Education in the Next Millennium. (2000, April 28). Bridges: Connecting Research and Education in the Earth System Sciences. Retrieved from http://serc.carleton.edu/research_education/bridges.html

Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.
Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. Washington, DC: Sage Publications Inc.
Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches. Washington, DC: Sage Publications Inc.
Dixon-Wood, M., & Sutton, A. (2004). Systematic review. In Michael S. Lewis-Beck, A. Bryman, & Tim Futing Liao (Eds.), The SAGE Encyclopedia of Social Science Research Methods. (pp. 1111-1112). Thousand Oaks, CA: Sage Publications, Inc.
Drummond, C. N., & Markin, J. M. (2008). An analysis of the bachelor of science in geology degree as offered in the United States. Journal of Geoscience Education, 56(2), 113-119.
Dykas, M. J., & Valentino, D. W. (2016). Predicting Performance in an Advanced Undergraduate Geological Field Camp Experience. Journal of Geoscience Education, 64(4), 314-322.
Earth Science Literacy Initiative. (2010). Earth science literacy principles: The big ideas and supporting concepts of Earth science. Retrieved from www. earthscienceliteracy.org/es_literacy_6may10_. pdf.
Entwistle, N., & Ramsden, P. (2015). Understanding student learning (Routledge Revivals). Routledge.
Feig, A. D. (2010). Technology, accuracy and scientific thought in field camp: An ethnographic study. Journal of Geoscience Education, 58(4), 241-251.
Fink, L. D. (2013). Creating significant learning experiences: An integrated approach to designing college courses. John Wiley & Sons.
Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410-8415.
Gilbert, L. A., Stempien, J., McConnell, D. A., Budd, D. A., van der Hoeven Kraft, K. J., Bykerk-Kauffman, A., ... & Wirth, K. R. (2012). Not just “rocks for jocks”: Who are introductory geology students and why are they here?. Journal of Geoscience Education, 60(4), 360-371.
Giorgis, S. (2015). Google Earth mapping exercises for structural geology students—A promising intervention for improving penetrative visualization ability. Journal of Geoscience Education, 63(2), 140-146.
Gonzales, L. M., & Keane, C. M. (2009). Who will fill the geoscience workforce supply gap? Environmental science & technology, 44(2), 550-555.
Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. Systematic reviews, 1(1), 28.
Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries Journal, 26(2), 91-108.
Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), The paradigm dialog (pp. 17–30). Newbury Park, CA: Sage.
Handelsman, J., & Brown, Q. (2016, August 17). A call to action: Incorporating active STEM learning strategies into K-12 and higher education. Retrieved from https://www.whitehouse.gov/blog/2016/08/17/call-action-incorporating-active-stem-learning-strategies-k-12-and-higher-education

Hanks, C., Levine, R., Gonzalez, R., Wartes, D., & Fowell, S. (2007). Survey development for measuring the near-term effectiveness of a program to recruit minority geoscientists. Journal of Geoscience Education, 55(3), 244-250.
Harden, A. (2010). Mixed-methods systematic reviews: integrating quantitative and qualitative findings. Focus, 2010, 1-8.
Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. Physical Review Special Topics-Physics Education Research, 3(2), 020102.
Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. Journal of Research in Science Teaching, 48(8), 952-984.
HLWIKI International. (2017). Rapid reviews. Retrieved from http://hlwiki.slais.ubc.ca/index.php/Rapid_reviews
Hodder, P. W. (2001). “Earthquake!”—A cooperative learning experience. Journal of Geoscience Education, 49(3), 280-285.
Intergovernmental Panel on Climate Change. (2015). Climate Change 2014: Mitigation of Climate Change (Vol. 3). Cambridge University Press.
Jarrett, O. S., & Burnley, P. (2010). Lessons on the role of fun/playfulness from a geology undergraduate summer research program. Journal of Geoscience Education, 58(2), 110-120.
Johnson, R., Snow, J., Abshire, W., Buhr, S., Cullen, H., Denning, S., Holland, M., Manduca, C., Pennington, P., Schoedinger, S., Schultz, P., & Marshall, J., (2009). Essential principles and fundamental concepts for atmospheric science literacy. Retrieved from http://gcoos.tamu.edu/documents/AtmosphericLiteracy.pdf
Jolley, A., Lane, E., Kennedy, B., & Frappé-Sénéclauze, T. P. (2012). SPESS: A new instrument for measuring student perceptions in Earth and ocean science. Journal of Geoscience Education, 60(1), 83-91.
Kastens, K. A., & Manduca, C. A. (Eds.) (2012). Earth and mind II: A synthesis of research on thinking and learning in the geosciences. Geological Society of America Special Paper 486. Boulder, CO: Geological Society of America.
King, K. (2008). Geoscience education: An overview. Studies in Science Education, 44(2), 187-222.
Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. American Journal of Occupational Therapy, 45(3), 214-222.
Kuhn, T. S. (1970). The structure of scientific revolutions. (2nd edn.). Chicago, IL: University of Chicago Press.
Kuhn, T. S. (1996). The Structure of Scientific Revolutions (3rd edn.). Chicago, IL: University of Chicago Press.
Kusnick, J. (2002). Growing pebbles and conceptual prisms–understanding the source of student misconceptions about rock formation. Journal of Geoscience Education, 50(1), 31-39.
Levine, R., González, R., Cole, S., Fuhrman, M., & Le Floch, K. C. (2007). The geoscience pipeline: A conceptual framework. Journal of Geoscience Education, 55(6), 458-468.
Lewis, E.B., & Baker, D.R. (2010). A call for a new geoscience education research agenda. Journal of Research in Science Teaching, 47(2), 121-129.
Libarkin, J. C., Elkins, J. T., & St. John, K. (2009). Editorial: The evolution of JGE: Responding to our community's needs. Journal of Geoscience Education, 57(3), 165-167.

Liverman, D. (2009). Communicating geological hazards: Educating, training and assisting geoscientists in communication skills. In Geophysical Hazards (pp. 41-55). Springer Netherlands.
Lodahl, J. B., & Gordon, G. (1972). The structure of scientific fields and the functioning university graduate departments. American Sociological Review, 37, 57-72.
Macdonald, R. H., Manduca, C. A., Mogk, D. W., & Tewksbury, B. J. (2005). Teaching methods in undergraduate geoscience courses: Results of the 2004 On the Cutting Edge survey of US faculty. Journal of Geoscience Education, 53(3), 237-252.
Maggio, G., & Cacciola, G. (2012). When will oil, natural gas, and coal peak?. Fuel, 98, 111-123.
Manduca, C. A., Mogk, D. W., & Stillings, N. (2002, July). Bringing research on learning to the geosciences. Johnson Foundation, Wingspread Conference Center.
Manduca, C. A., & Mogk, D. W. (Eds.) (2006). Earth and Mind: How Geologists Think and Learn about the Earth. Geological Society of America Special Paper 413. Boulder, CO: Geological Society of America.
Manduca, C. A., Iverson, E. R., Luxenberg, M., Macdonald, R. H., McConnell, D. A., Mogk, D. W., & Tewksbury, B. J. (2017). Improving undergraduate STEM education: The efficacy of discipline-based professional development. *Science Advances*, *3*(2), e1600193.
Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. Journal of Educational Psychology, 106(1), 121.
McDonald, R. I., Weber, K. F., Padowski, J., Boucher, T., & Shemie, D. (2016). Estimating watershed degradation over the last century and its impact on water-treatment costs for the world's large cities. Proceedings of the National Academy of Sciences, 113(32), 9117-9122.
McNeal, K. S., Walker, S. L., & Rutherford, D. (2014). Assessment of 6-to 20-grade educators' climate knowledge and perceptions: results from the climate stewardship survey. Journal of Geoscience Education, 62(4), 645-654.
Mogk, D. (n.d.) Twenty five years of progress in geoscience education. Retrieved from https://serc.carleton.edu/NAGTWorkshops/GeoEd_Progress.html
Mogk, D. W. (2000). Bridges: Connecting research and education in the Earth sciences. Retrieved from https://serc.carleton.edu/research_education/bridges.html
Mosher, S., Bralower, T., Huntoon, J., Lea, P., McConnell, D., Miller, K. … White, L. (2014). Future of Undergraduate Geoscience Education. Retrieved from http://www.jsg.utexas.edu/events/files/Future_Undergrad_Geoscience_Summit_report.pdf
National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2007). Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. Washington, DC: The National Academies Press. https://doi.org/10.17226/11463
National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2010). Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5. Washington, DC: The National Academies Press. https://doi.org/10.17226/12999
National Science Board's Task Committee on Undergraduate Science and Engineering Education. (1986). Undergraduate, science, mathematics, and engineering education. Washington, DC: National Science Board.

National Science Foundation. (1997). Geoscience education: A recommended strategy (NSF 97-171). Retrieved from https://www.nsf.gov/pubs/1997/nsf97171/nsf97171.txt

Perkins, D. (2004). Scholarship of teaching and learning, assessment, and the Journal of Geoscience Education. Journal of Geoscience Education, 52(1), 113-114.

Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. Academy of Management Review, 18(4), 599-620.

Piburn, M. D., van der Hoeven Kraft, K., & Pacheco, H. (2011). A new century for geoscience education research. Retrieved from http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072590.pdf

President's Council of Advisors on Science and Technology. (2012, November 30). Transformation and opportunity: The future of the US research enterprise. Retrieved from https://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports

Ratcliffe, M. (1997). Pupil decision-making about socio-scientific issues within the science curriculum. International Journal of Science Education, 19(2), 167-182.

Sadler, T. D., & Zeidler, D. L. (2005). Patterns of informal reasoning in the context of socioscientific decision making. Journal of Research in Science Teaching, 42(1), 112-138.

Schoon, K. J. (1995). The origin and extent of alternative conceptions in the earth and space sciences: A survey of pre-service elementary teachers. Journal of Elementary Science Education, 7(2), 27-46.

Seymour, E., Hewitt, N. A. (1997). Talking about leaving: Why undergraduates leave the sciences. Westview Press.

Shea, J. H. (1995). Forty-five years of the Journal of Geological Education. Journal of Geological Education, 43(5), 450-450.

Shen, K. M., Lee, M. H., Tsai, C. C., & Chang, C. Y. (2016). Undergraduate students' Earth science learning: Relationships among conceptions, approaches, and learning self-efficacy in Taiwan. International Journal of Science Education, 38(9), 1527-1547.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. Education for Information, 22(2), 63-75.

Sibley, D. F. (2009). A cognitive framework for reasoning with scientific models. Journal of Geoscience Education, 57(4), 255-263.

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (Eds.). (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. Washington, DC: National Academies Press.

Singer, S. R. (2013). Advancing research on undergraduate science learning. Journal of Research in Science Teaching, 50(6), 768-772.

Sodhi, M. S. (2016). Natural disasters, the economy and population vulnerability as a vicious cycle with exogenous hazards. Journal of Operations Management, 45, 101-113.

St. John, K., Dickerson, D., & McNeal, K. S. (2013). Editorial: Guide to aspiring authors. Journal of Geoscience Education, 61(3), 253-255.

St. John, K., Petcovic, H., Stokes, A., Arthurs, L., Callahan, C., Feig, A., ... & Nagy-Shadman, E. (2016). Editorial: Un-packaging manuscript preparation and review guidelines for curriculum and instruction and research papers. Journal of Geoscience Education, 64(1), 1-4.

St. John, K., & McNeal, K. (2016, September 1). One approach for characterizing the strength of evidence of geoscience education research (GER) community claims. Retrieved from http://nagt.org/nagt/profdev/workshops/geoed_research/pyramid.html

Stern, P. C., & Dietz, T. (2015). IPCC: Social scientists are ready. Nature, 521(7551), 161-161.

Stokes, D. E. (1997). Pasteur's quadrant: Basic science and technological innovation. Washington, DC: Brookings Institution Press.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356-367.

Ulanski, S. L. (1995). Curriculum reform in undergraduate geology programs. Journal of Geological Education, 43(1), 43-46.

US Census Bureau. (n.d.). Quick facts: United States. Retrieved from https://www.census.gov/quickfacts/table/PST045215/00

US Department of Health and Human Services. (2009). Code of federal regulations. Title 45 public welfare. Department of Health and Human Services. Part 46: Protection of human subjects.

US Global Change Research Program. (2009, March). Climate literacy: The essential principles of climate science Version 2. Retrieved from http://cpo.noaa.gov/sites/cpo/Documents/pdf/ClimateLiteracyPoster-8_5x11_Final4-11LR.pdf

van der Hoeven Kraft, K. J., Srogi, L., Husman, J., Semken, S., & Fuhrman, M. (2011). Engaging students to learn through the affective domain: A new framework for teaching in the geosciences. Journal of Geoscience Education, 59(2), 71-84.

Wiedmann, T. O., Schandl, H., Lenzen, M., Moran, D., Suh, S., West, J., & Kanemoto, K. (2015). The material footprint of nations. Proceedings of the National Academy of Sciences, 112(20), 6271-6276.

Wiggins, G. P., & McTighe, J. (2005). Understanding by design. Ascd.

Wilson, C. (2016). Status of recent geoscience undergraduates. Retrieved from http://www.americangeosciences.org/workforce/reports

York, R., Rosa, E. A., & Dietz, T. (2015). A tale of contrasting trends: Three measures of the ecological footprint in China, India, Japan, and the United States, 1961-2003. Journal of World-Systems Research, 15(2), 134-146.

Supplementary Information linked to the online version of the paper at Wiley-Blackwell: Supplementary Material.

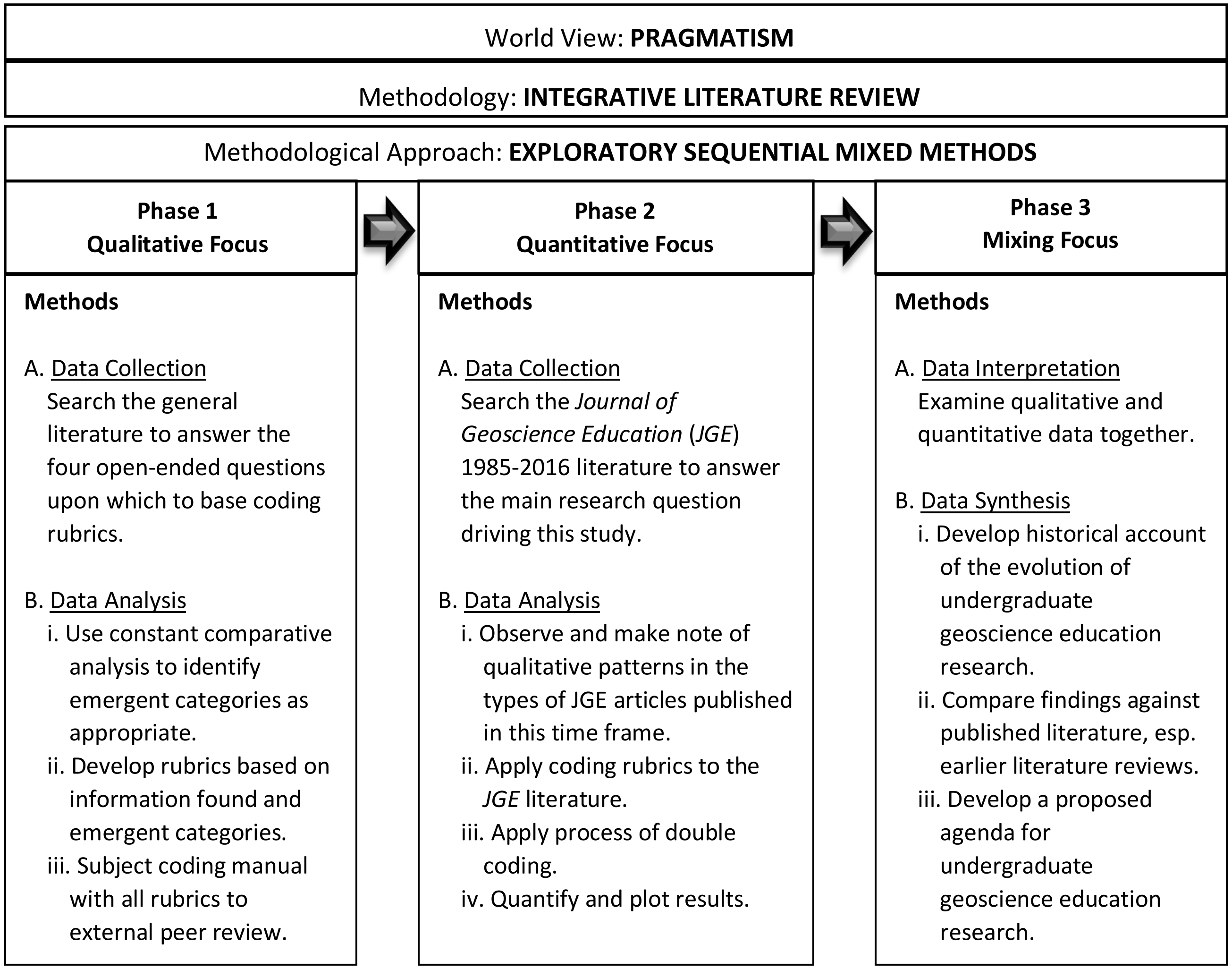


*Figure 1*. Schematic of this study's design.

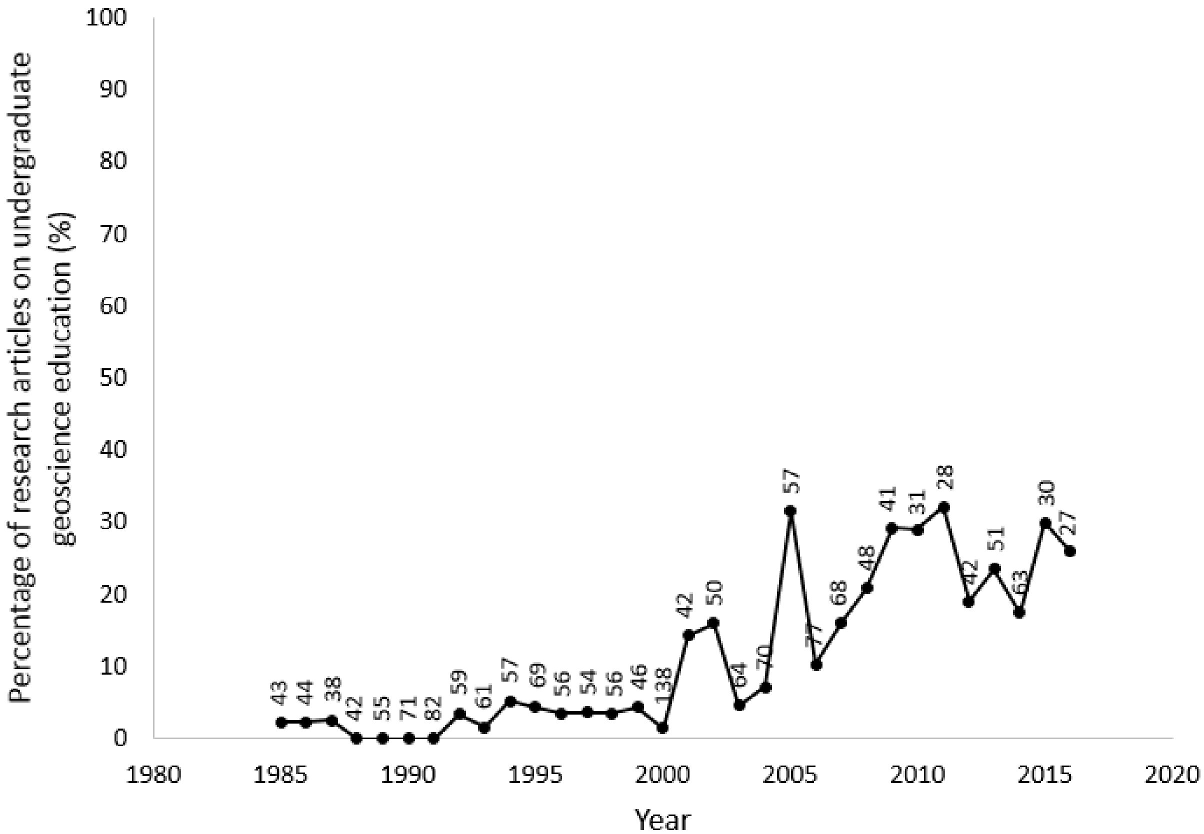


*Figure 2*. Number of research articles on geoscience education per year are normalized as the percent of total JGE articles on undergraduate geoscience education research, 1985-2016. Inset data labels show total JGE articles per year.

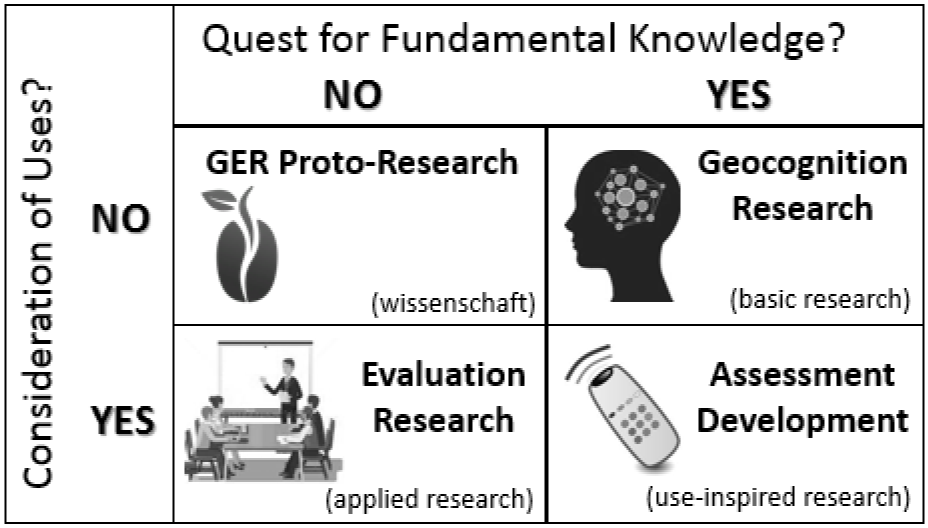


*Figure 3*. Juxtaposition of examples of research on geoscience education onto Stokes' quadrant model for research.

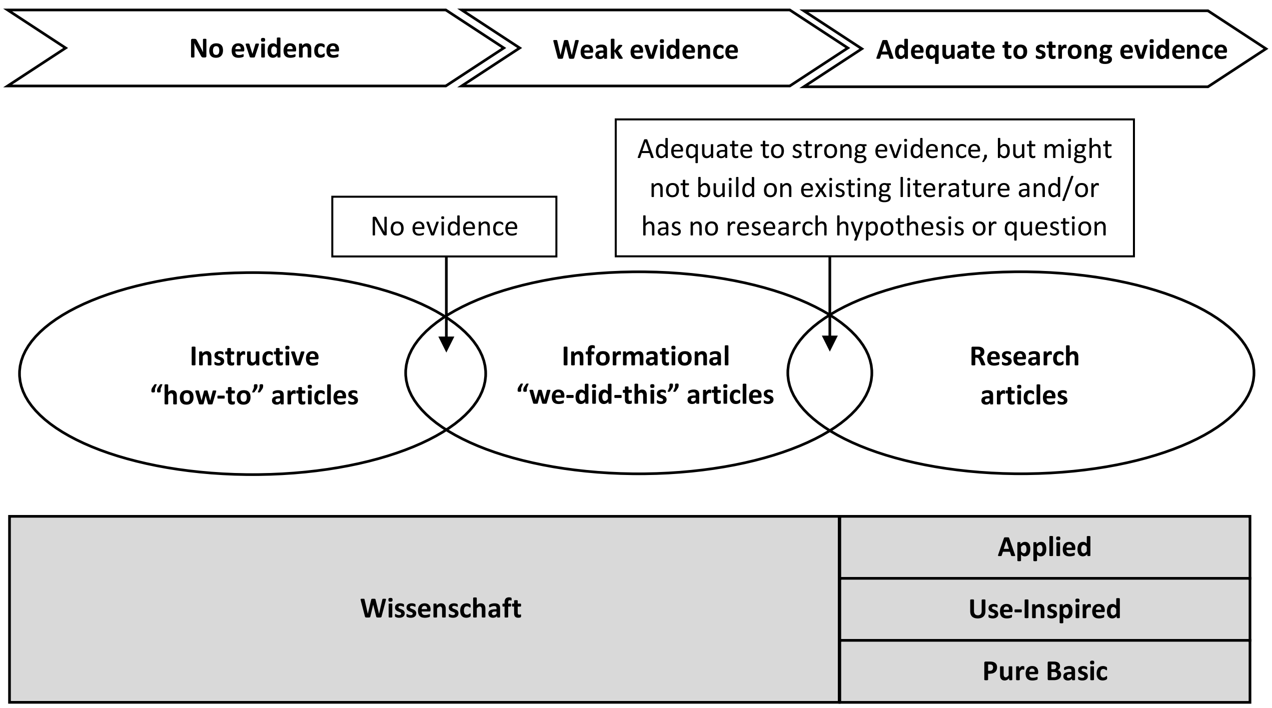


*Figure 4*. Conceptual model showing the relationships between the strength of evidence for claims made (arrows), the three emergent categories of JGE articles (ellipses), and the four quadrants in Stokes' 1997 model of research (grey rectangles).

## Supplementary Material

### Rubric for Topics of Research

The emergent themes form the basis for the coding rubric about *topics of research*, and it includes: (a) teaching methods, (b) assessment tools, (c) student cognition, (d) student affect, (e) departmental curricula and programs, and (f) minorities and recruitment. With respect to the "student affect" category, articles are coded for student affect only when the primary focus of study is actually student affect (i.e., articles that included opinion surveys about an activity do not, for the purposes of the present review, constitute studies about affect). Articles that were about other topics fell into a "miscellaneous" category and were not further discussed in this study. Table S1 provides details about the rubric for coding the topics of research.

Table S1
*Rubric about topics of research.*

| Category Name | Explanation |
|---|---|
| Teaching methods | The category "teaching methods" is constrained to mean what the instructor does during class time to teach students, especially the opportunities they create to enable active learning. Most education articles focus on non-lecture-based methods of teaching and focus on methods for which there exists an evidence or research basis that shows it promotes student learning. These are what Henderson and Dancy (2007) call "research-based instructional strategies" (RBISs). Examples of RBISs include: peer instruction, just-in-time teaching, and inquiry-based learning. |
| Assessment tools | The category "assessment tools" refers to instruments developed to measure student learning in, for example, the cognitive and affective domains. These are instruments that are developed using one or more psychometrically-grounded methods for developing valid and reliable assessment instruments. |
| Student cognition | The term "cognition" means mental processes involved in learning new knowledge, developing mental skills, and/or applying new knowledge and new skills to a variety of different contexts for problem solving and/or decision making. Subjects of geoscience education research in the area of student cognition include, for example: misconceptions, spatial thinking skills, and understanding deep time. |
| Student affect | The term "affect" is a psychological term that refers to emotions or feelings about/toward something. It is often associated with values and beliefs. Subjects of geoscience education research in the area of student affect include, for example: perceived value of studying geoscience, motivation for studying geoscience, and affinity for the outdoors. |
| Minorities and recruitment | Here, "minorities" is defined broadly and refers to underrepresented minorities in terms of sex, gender, race, ethnicity, disability, age, and marital/partner status. "Recruitment" here refers to the process of adding individuals to an existing population (e.g., population of atmospheric sciences majors, population of professional geologists, etc.) |
| Departmental curricula and programs | Here, "departmental curricula" refers to the curricula for the undergraduate degrees and minors proffered by a department and "departmental programs" refers to other programs a department might host, such as undergraduate summer research programs. |

**Rubric for Structural Components of Research Design**

Creswell's (e.g., 2009, 2014) work on framing social and education sciences research articles form the basis for the coding rubric about *structural components of research design*. The rubric includes: (i) worldview, (ii) methodology, (iii) approach, and (iv) methods. The results, discussion, and conclusion sections are not included in this rubric because they appear in some form in all GER articles reviewed. Thus, the research articles analysed for this study were examined in terms of whether or not they included a worldview, methodology, approach, and methods. If they were included, then what was included was also examined.

**Rubric for Level of Connections and Rubric for Level of Evidence**

The degree to which an article connects to other studies as well as the quality of evidence provided to support the article's claims are used as two measures of rigor. Two rubrics developed by Henderson et al. (2011) were modified for this study. The first is a rubric about the *level of connections* and the second is about *level of evidence*. Henderson et al. (2011) developed a rubric with four categories of connection: (a) Level 1 means that "the article [is] strongly connected to the … literature" (p. 966), (b) Level 2 means that "the article [is] weakly connected to the … literature" (p. 966), (c) Level 3 means that the article "cite[s] some … literature, but [does] not make connections between the literature cited and the [subject matter] studied" (p. 966), and (d) Level 4 articles "do not cite any … literature" (p. 966).

For the present study, a rubric for the *level of connections* utilizes the aforementioned criteria but also modifies them, to provide a more quantitative basis for assigning the Levels. For this study, (a) Level 1 articles connect their findings to at least four other articles in the literature that are not self-citations, (b) Level 2 articles connect their findings to two or three articles in the literature that are not self-citations, (c) Level 3 articles cite other literature (e.g., in the introductory sections of the article) but do not connect their findings to other literature, and (d) Level 4 articles do not cite other non-self-cited literature. Modifications were made to provide exact cut offs that would enhance the consistency with which the rubric is applied to articles.

Henderson et al.'s (2011) also developed a rubric for use of evidence to support claims and it is also comprised of four categories: (a) Strong articles have "well-explained methods and make clear connections between the claims and supporting evidenc"' (p. 971); (b) Adequate articles present "evidence to support claims of success, but the evidence or methods are not fully explained or fully convincing" (p.971); (c) Poor articles present "anecdotal, vague, and/or undefined evidence to support claims made" (p. 971); and (d) None means "no evidence to support their claims of success or failure" are presented.

For the present study, a rubric for the *level of evidence* utilizes these for categories and criteria but restructures them in terms of levels so that they can be compared side-by-side with the level of connections for each article. Thus, for this study the rubric for the *level of evidence* is as follows: (a) Level 1 articles meet the criteria Henderson et al. set for "strong" articles, (b) Level 2 articles meet the criteria Henderson et al. set for "adequate" articles, (c) Level 3 articles meet the criteria defined by Henderson et al. for "poor" articles, and (d) Level 4 articles meet the criteria set by Henderson et al. for "none."

It is worth noting that the criteria that define the different levels in the rubric about connections can be quantitatively modified to raise or lower the standards for each level. For example, the criteria for Level 1 connections involves citing at least four other articles in the discussion of a study's results. Arguably, this is a relatively low bar and one that researchers using this rubric in the future might adjust. Regardless of the quantitative cut offs, application of the rubrics related to research rigor do indicate a positive trend in the overall rigor of GER studies published in the *JGE* from 1985 to 2016. These rubrics were applied to the Discussion sections, where results are discussed in the context of other studies.

**Examples of Articles for Each Pattern of Articles Identified in the First Stage of Review**

Table S2

*Articles categorized as instructive, informational, and research articles during the first stage of the review. Instructive "how-to" articles and informational "we did this" articles were not included in the second stage. Examples are listed in chronological order.*

| |
|---|
| Instructive "how-to" articles |
| Lutz, T. M. (2001). Enhancing students' understanding of risk and geologic hazards using a dartboard model. *Journal of Geoscience Education, 49*(4), 339-345. |
| Harpp, K. S., Koleszar, A. M., & Geist, D. J. (2005). Volcanoes in the classroom: A simulation of an eruption column. *Journal of Geoscience Education, 53*(2), 173-175. |
| Halfen, A. F., White, T., Slocum, T., Hirmas, D. R., McDermott, D., Atchley, P., ... & Gilbreath, A. (2014). A new stereoscopic (3D) media database and teaching strategy for use in large-lecture introductory geoscience courses. *Journal of Geoscience Education, 62*(3), 515-531. |
| Informational "we-did-this" articles |
| Bank, C. G. (2006). Reading and writing taught in a sophomore course on plate tectonics. *Journal of Geoscience Education, 54*(1), 25-30. |
| Todd, C. E. D., & Goeke, E. R. (2012). Incorporating student-led field trips and learner-centered teaching in a capstone geology course. *Journal of Geoscience Education, 60*(3), 268-276. |
| Kelley, D. F., Sumrall, J. L., & Sumrall, J. B. (2015). Student-designed mapping project as part of a geology field camp. *Journal of Geoscience Education*, *63*(3), 198-209. |
| Research articles |
| Drummond, C. N., & Markin, J. M. (2008). An analysis of the Bachelor of Science in Geology degree as offered in the United States. *Journal of Geoscience Education, 56*(2), 113-119. |
| Nelson, K. G., Huysken, K., & Kilibarda, Z. (2010). Assessing the impact of geoscience laboratories on student learning: Who benefits from introductory labs?. *Journal of Geoscience Education, 58*(1), 43-50. |
| Lou, Y., Blanchard, P., & Kennedy, E. (2015). Development and validation of a science inquiry skills assessment. *Journal of Geoscience Education, 63*(1), 73-85. |

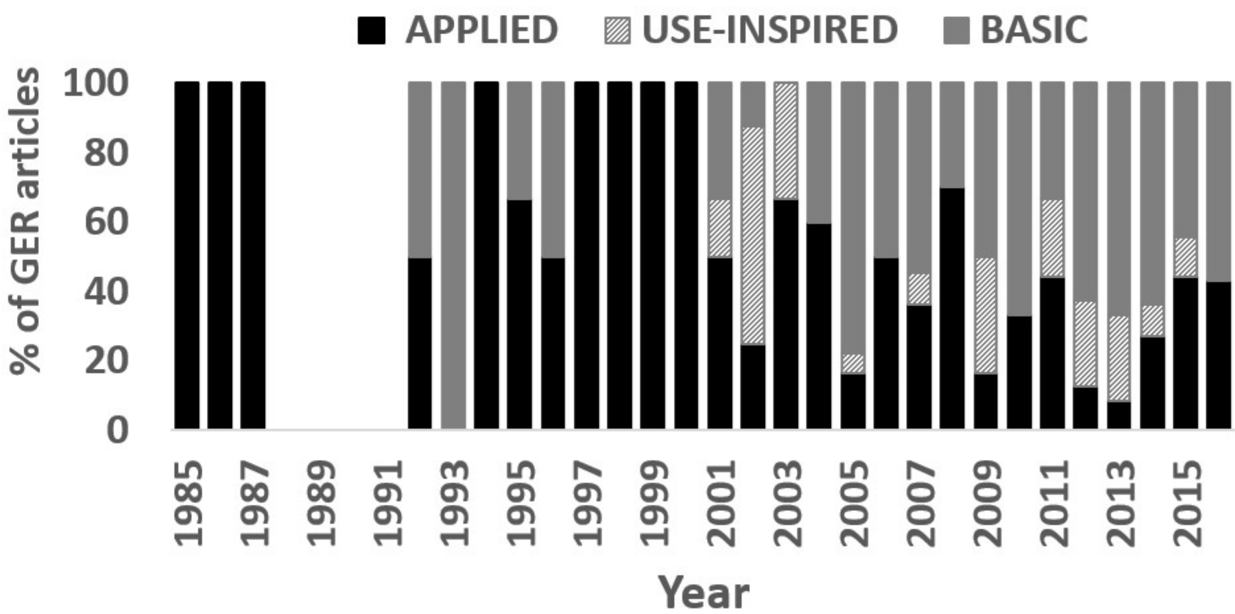


*Figure S1*. Percentage of each type of research represented in GER articles published each year.

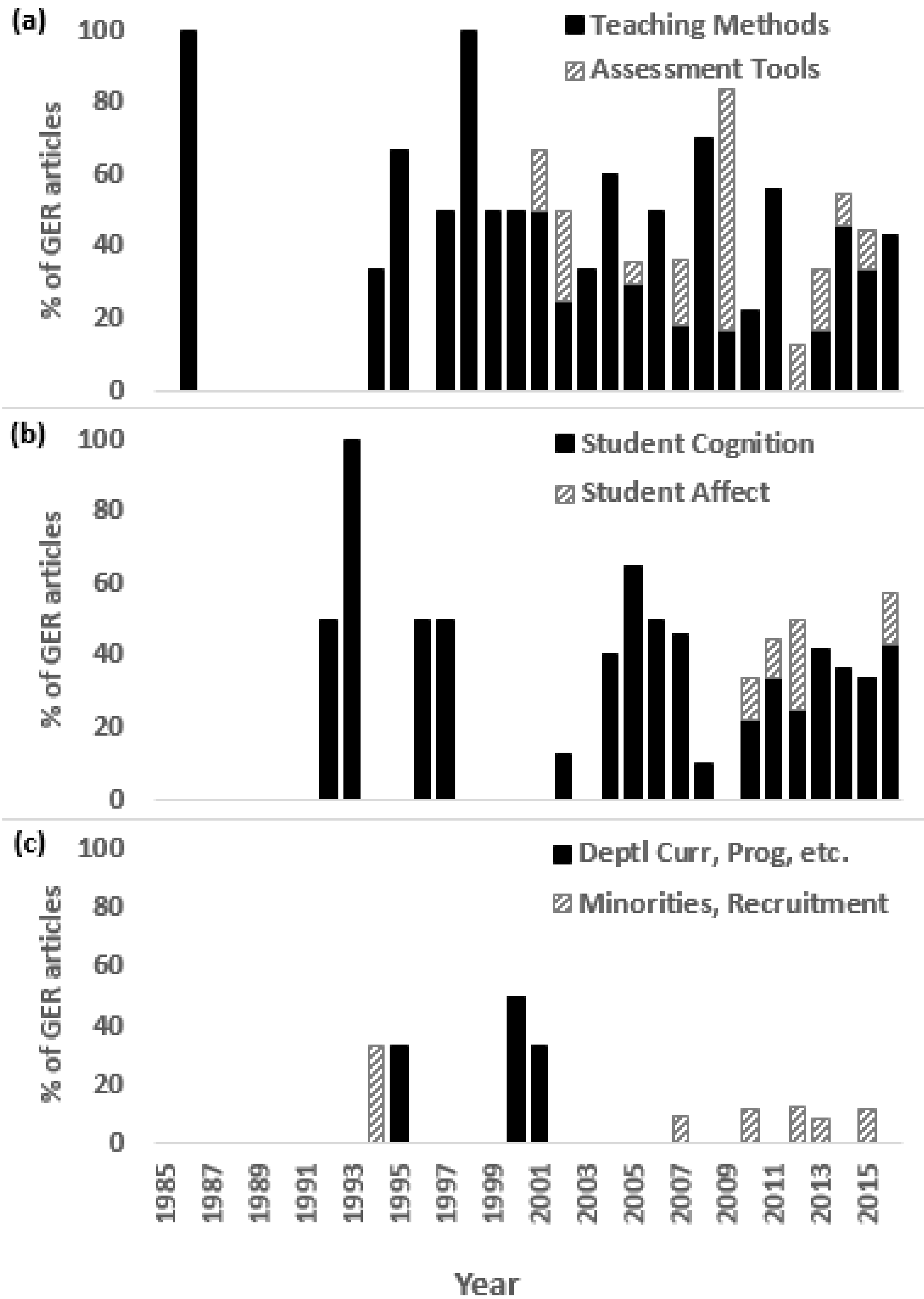


*Figure S2*. Six themes of research topics represented in GER articles each year. A miscellaneous category of topics is not shown, thus each year's sum may not equal 100%. Themes are paired in Figure S2 to highlight differences in two closely related themes.

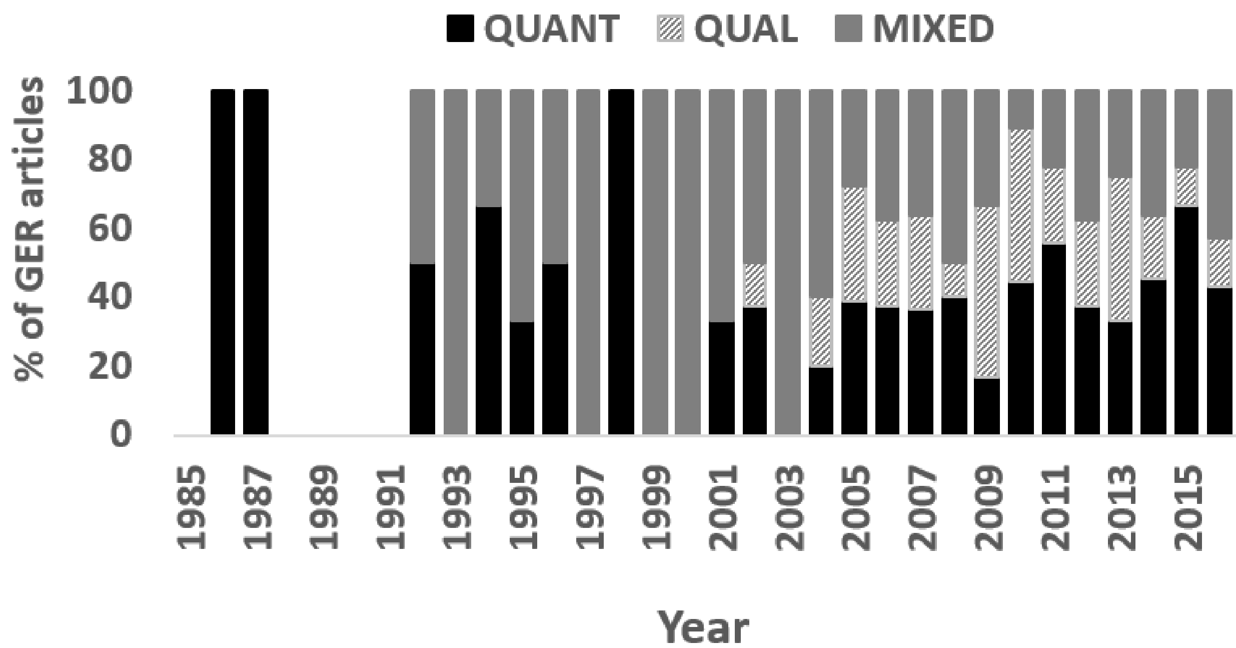


*Figure S3*. Methodological approaches (quantitative, qualitative, and mixed) used to undertake GER each year.

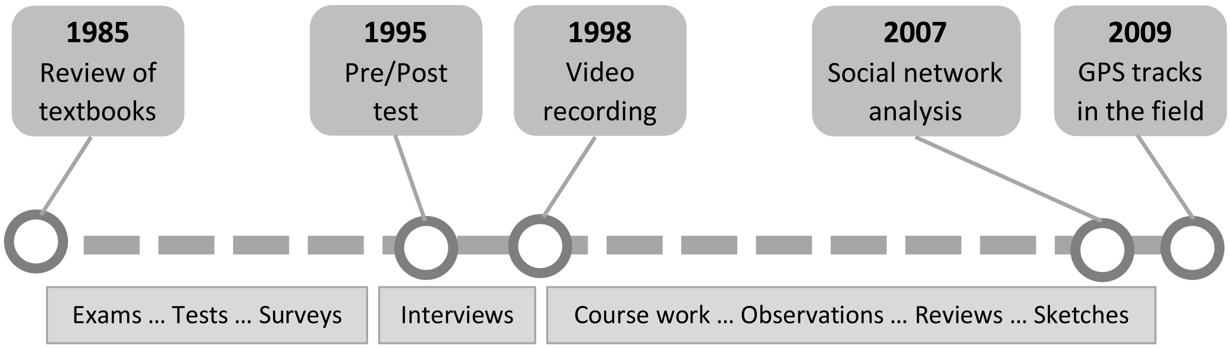


*Figure S4*. Approximate timeline of when methods are first used in the reviewed GER articles.

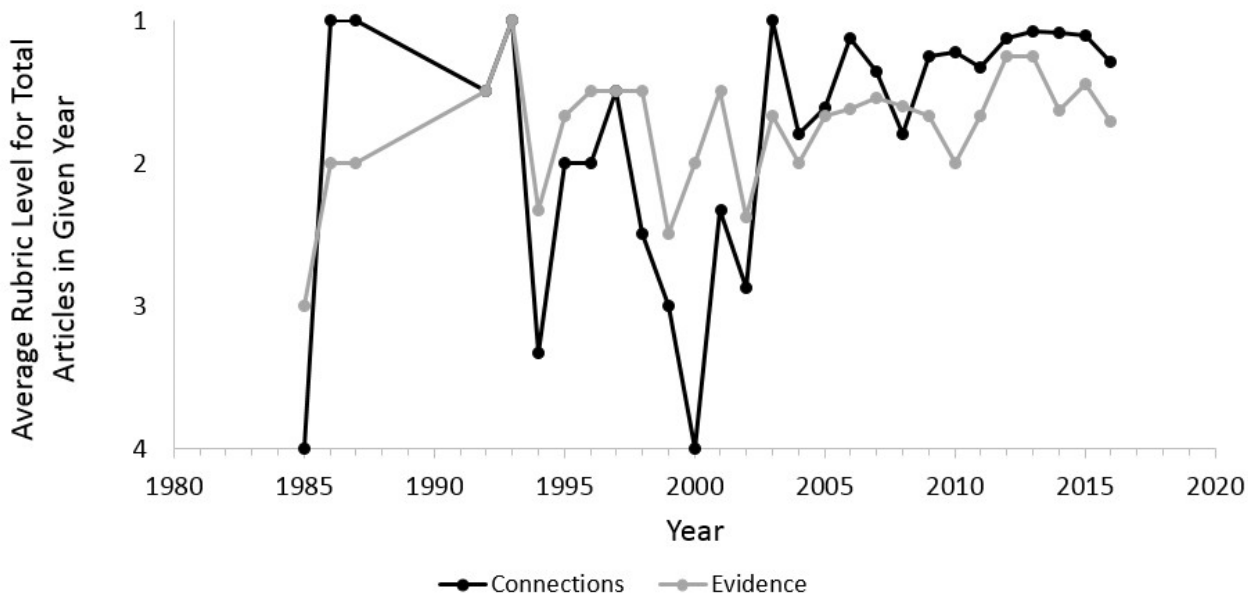


*Figure S5*. Research rigor as illustrated by average level for the connections and average level of evidence in GER articles in a given year. Level 1 indicates the highest level of connections or evidence, and Level 4 indicates the lowest level of connections or evidence.

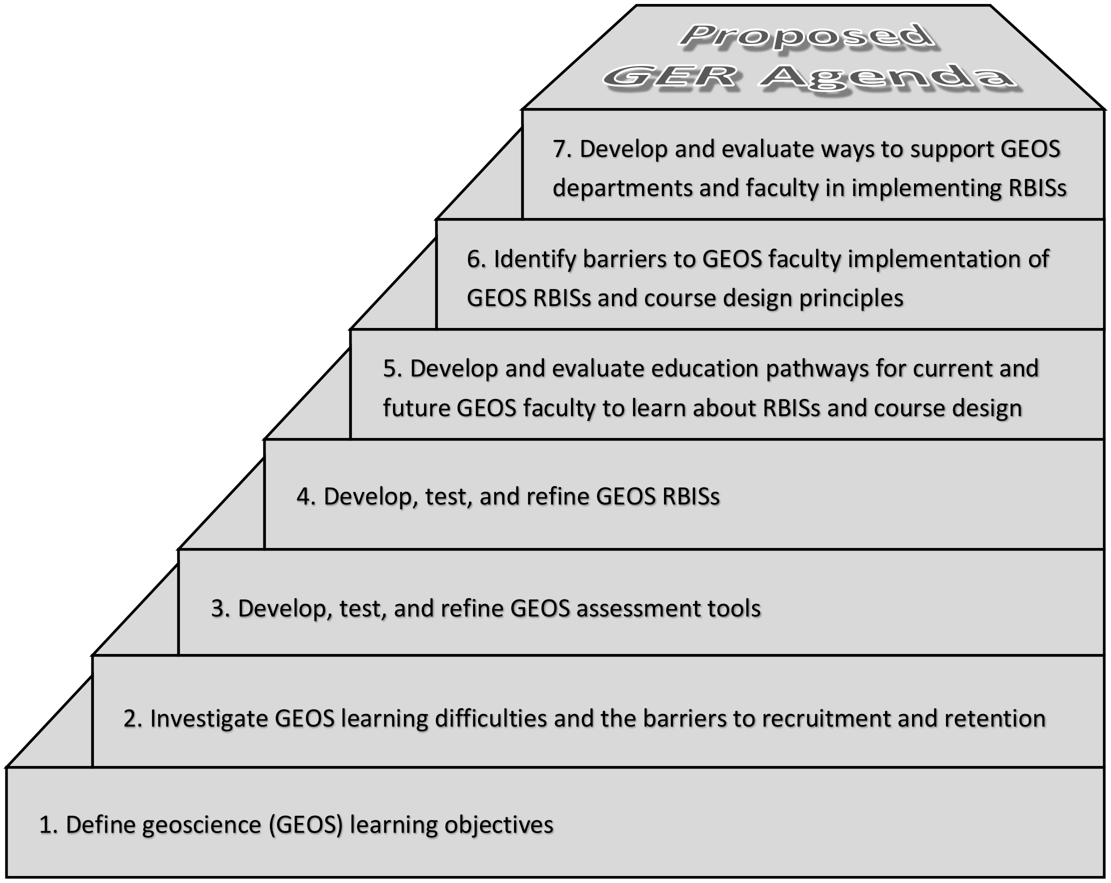


*Figure S6.* Conceptual model for proposed undergraduate geoscience education research agenda.