

Prevailing Questions and Methodologies in Biology Education Research: A Longitudinal Analysis of Research in *CBE—Life Sciences Education* and at the Society for the Advancement of Biology Education Research

Stanley M. Lo,^{†*} Grant E. Gardner,^{†*} Joshua Reid,[‡] Velta Napoleon-Fanis,[‡] Penny Carroll,[‡] Emily Smith,[‡] and Brian K. Sato^{§*}

[†]Section of Cell and Developmental Biology, Division of Biological Sciences, and Program in Mathematics and Science Education, University of California, San Diego, La Jolla, CA 92093;

[‡]Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132; [§]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 92697

ABSTRACT

Biology education research (BER) is a growing field, as evidenced by the increasing number of publications in *CBE—Life Sciences Education* (*LSE*) and expanding participation at the Society for the Advancement of Biology Education Research (SABER) annual meetings. To facilitate an introspective and reflective discussion on how research within *LSE* and at SABER has matured, we conducted a content analysis of *LSE* research articles ($n = 339$, from 2002 to 2015) and SABER abstracts ($n = 652$, from 2011 to 2015) to examine three related intraresearch parameters: research questions, study contexts, and methodologies. Qualitative data analysis took a combination of deductive and inductive approaches, followed by statistical analyses to determine the correlations among different parameters. We identified existing research questions, study contexts, and methodologies in *LSE* articles and SABER abstracts and then compared and contrasted these parameters between the two data sources. *LSE* articles were most commonly guided by descriptive research questions, whereas SABER abstracts were most commonly guided by causal research questions. Research published in *LSE* and presented at SABER both prioritize undergraduate classrooms as the study context and quantitative methodologies. In this paper, we examine these research trends longitudinally and discuss implications for the future of BER as a scholarly field.

INTRODUCTION

With the growing emphasis on improving science, technology, engineering, and mathematics (STEM) education, there has been a push to transform student learning experiences and to develop evidence-based, inclusive interventions that serve students from diverse backgrounds and experiences (Boyer, 1998; National Research Council [NRC], 2003, 2009, 2015; America Association for the Advancement of the Science, 2011; National Academies of Sciences, Engineering, and Medicine [NAS], 2011, 2016, 2018; House of Lords, 2012; President's Council of Advisors on Science and Technology, 2012; Marginson *et al.*, 2013). Biology education research (BER) is a new and expanding research field that supports these educational initiatives (NRC, 2012). Compared with related fields of chemistry and physics education research, BER has only recently matured into its own distinct field from the larger parental lineages of discipline-based education research (DBER) and science education research

David Feldon, *Monitoring Editor*

Submitted Aug 27, 2018; Revised Dec 13, 2018; Accepted Dec 17, 2018

CBE Life Sci Educ March 1, 2019 18:ar9

DOI:10.1187/cbe.18-08-0164

*Address correspondence to: Stanley M. Lo (smlo@ucsd.edu), Grant E. Gardner (Grant.Gardner@mtsu.edu), or Brian K. Sato (bsato@uci.edu).

© 2019 S. M. Lo *et al.* CBE—Life Sciences Education © 2019 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>). “ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

(Dirks, 2011; Gul and Sozibilir, 2016). DBER has a “deep grounding in the discipline’s priorities, worldview, knowledge, and practices” (Gul and Sozibilir, 2016, p. 1632), whereas science education research is broadly concerned with educational issues in science classrooms. BER is a subfield of DBER grounded in an understanding of the biological sciences.

BER has adopted and adapted theoretical and methodological traditions from DBER and science education research, as well as other social science and psychology fields, to study issues in the learning and teaching of biology (NRC, 2012). The growth of BER can be seen in the number of articles published in BER journals such as *CBE—Life Sciences Education (LSE)*; 26 publications in 2002 vs. 60 publications in 2015) and the number of presentations at BER conferences such as the Society for the Advancement of Biology Education Research (SABER) meeting (94 presentations in 2011 vs. 192 presentations in 2015). These numbers highlight the increase in BER scholarly participation; as BER continues to expand, it is important to reflect on how the field has potentially matured or changed over time.

In this paper, we present a study on the development of BER as a research field over the past 15 years. These types of studies have precedence in science education research and related fields. Some are quantitative meta-analyses focused on different STEM disciplines (Bowen, 2000; Prince, 2004; Freeman et al., 2014; Taylor et al., 2016) or active-learning strategies (Vernon and Blake, 1993; Springer et al., 1999; Gijbels et al., 2005; Ruiz-Primo et al., 2011). Other studies examined changes in research trends over time in different disciplines such as science education or learning sciences (Lee et al., 2009; Chang et al., 2010; Koh et al., 2014; O’Toole et al., 2018), specific areas of study such as conceptual change or identity (Amin, 2015; Chiu et al., 2016; Darragh, 2016), or contrasting geographic locations (Topsakal et al., 2012; Gul and Sozibilir, 2015; Chiu et al., 2016). These broad studies seek to provide an introspective lens for researchers to reflect on the state and development of their fields.

The history of BER has been investigated to a certain extent. One study tracked the origin and development of theoretical frameworks in BER (DeHaan, 2011). Another study analyzed BER literature from 1990 to 2010 and identified three main strands of research: analysis of student learning, analysis of student attitudes, and construction of validated concept inventories or research instruments (Dirks, 2011). This work highlighted particular populations and topics that were sparsely examined in BER, issues with experimental design and data collection, and considerations for future research (Dirks, 2011). A third study analyzed articles in eight major academic journals from 1997 to 2014 and found that learning, teaching, and attitudes were the most common topics and that undergraduate and high school students were the most studied populations (Gul and Sozibilir, 2016). Notably, *LSE* was not included in this analysis, as the majority of the data came from European journals such as the *Journal of Biological Education* and *International Journal of Science Education*. Together, these studies provide complementary but not necessarily coherent examinations of BER as a research field.

A Framework to Analyze BER Studies

To further conceptualize the current state of BER as a field, we drew upon a framework from *Defining an Identity: The Evolu-*

tion of Science Education as a Field of Research (Fensham, 2004). In this framework, the relative maturity of a research field can be evaluated using three criteria: 1) structural criteria, 2) outcome criteria, and 3) intraresearch criteria (Fensham, 2004). Structural criteria are institutional supports for a research field, such as the existence of journals (e.g., *LSE*), conferences and professional associations (e.g., SABER), academic appointments, and programs for training the next generation of researchers. Although BER has successful journals, professional associations, and conferences, academic appointments are just now becoming more common, and programs specifically for training BER scholars are few (Aikens et al., 2016). Based on these observations, BER appears to be a developing field from a structural perspective using Fensham’s (2004) framework.

The outcome criteria refer to the implications a research field has on practice. As seen in a number of meta-analyses, instructional practices in STEM education that are aligned with evidence from DBER and science education research have positive impacts on student outcomes (Springer et al., 1999; Ruiz-Primo et al., 2011; Freeman et al., 2014). While the learning and teaching of biology was included in these meta-analyses, other DBER disciplines are more developed and have in-depth analyses on the connection between research and practice (Vernon and Blake, 1993; Bowen, 2000; Prince, 2004; Gijbels et al., 2005). From an outcome perspective, BER again appears to be a developing field, suggesting the importance of continuing introspection and reflection.

The intraresearch criteria refer to the “substance and methodologies of research itself” (Fensham, 2004, p. 4) and include the scientific background and research questions that drive the scholarship; development of conceptual knowledge, theoretical frameworks, and research methodologies in the field; and progression of the field over time. To date, other researchers have explored the historical contexts and theoretical perspectives of BER (DeHaan, 2011) and conducted broad content analyses related to the field (Dirks, 2011; Gul and Sozibilir, 2016). However, our current work is grounded in the intraresearch criteria of Fensham’s conceptual framework, arguing that an examination of intraresearch parameters such as research questions, study contexts, and methodologies over time will contribute to defining BER as a maturing field.

Rationale for Our Current Study

To facilitate an introspective and reflective discussion on how BER has matured over time, we analyzed research articles from *LSE* and abstracts selected for presentation at the SABER annual meeting from their inceptions (*LSE* in 2002 and SABER in 2011) until 2015. The aim and scope of *LSE* as stated in the inaugural editorial are to 1) “provide an opportunity for scientists and others to publish high-quality, peer-reviewed, educational scholarship of interest of ASCB members”; 2) “provide a forum for discussion of educational issues”; and 3) “promote recognition and reward for educational scholarship” (Ward, 2002). More information about the journal can be found at www.lifescied.org. We believe that *LSE* is an appropriate data source to capture the development of BER because of its rising influence in the field and because it is now the lead research journal in BER. According to Scimago journal rankings (Scimago, 2007), *LSE* has an h-index of 40 and has consistently

been in the first quartile of education journals over the past 5 years; since its inception in 2002, all reported parameters have steadily increased: total cites (from 13 to 337), citations per document (from 0.43 to 1.97), external cites (defined as citations from outside *LSE*) per document (from 0.03 to 1.49), and the percent of cited documents overall (from 35.48 to 52.82%).

We chose to analyze SABER abstracts for a number of reasons. First, SABER has been a major BER conference in the United States since its inception in 2011. At the time when we began this study, there were no other conferences solely dedicated to BER scholarship, although the Gordon Reference Conference has since added a biannual meeting on undergraduate biology education research in 2015. Second, conference presentations provide a unique data source compared with journal articles. The acceptance threshold for a conference presentation is likely to be lower than that for a publication, and we were therefore able to capture work that was in its earlier stages or that might have been considered not exciting or impactful enough for submission to a journal (e.g., negative results). More information about SABER can be found at: <https://saberbio.org>. By analyzing SABER abstracts in combination with *LSE* articles, we reason that our study may be able to present a more comprehensive picture of work being conducted in BER, complementing existing work in the literature.

Our study complements others in the existing literature in two important ways. First, we focus specifically on trends related to intraparameters, such as research questions, study contexts, and methodologies, adding to the historical perspectives and broad trends identified in other studies (DeHaan, 2011; Dirks, 2011). Second, our work expands on a similar study (Gul and Sozbilir, 2016) by including *LSE*, a major journal in the BER field. Our study is also the first in BER to include abstracts from a research conference in this type of analysis, providing additional triangulation from data sources that might hold different standards for intraresearch criteria.

In this paper, we identified existing research questions, study contexts, and methodologies in BER through an analysis of research articles from *LSE* and abstracts accepted for presentation at SABER before 2015. The comprehensive nature of our study sample also allowed us to examine changes in the field over time. Specifically, our study was guided by the following research question: What are the longitudinal trends in *LSE* articles and SABER abstracts in relation to the intraresearch parameters of research questions, study contexts, and methodologies?

METHODS

Sample Selection

A comprehensive sample of *LSE* publications since the founding of the journal in the year 2002 to the year 2015 (volumes 1–14, approximately four issues in each volume) were analyzed. Each issue contained multiple sections, including features, essays, research methods, and articles. All publications classified as “articles,” including special articles, were used as analytical units. Articles were specifically selected, because they are “for dissemination of biology education research that is designed to generate more generalizable, basic knowledge about biology education” (American Society for Cell Biology, 2017) and are

most likely to be representative of a scholarly approach to BER. There were 436 articles in all of the noted volumes. During review for content validity, an additional 97 articles (22%) were excluded, because they were nonempirical (i.e., did not collect original data). This resulted in a final sample of $n = 339$ articles (78%), with more than 90% of these articles reporting on studies conducted in the United States.

All abstracts presented at the SABER annual meeting since its inception in 2011 and up to 2015 ($n = 688$) were analyzed. These abstracts are publicly available and can be found on the SABER website. Most abstracts were empirical ($n = 652$, or 95%), including interventions-based research, investigations not based on interventions, and synthesis of existing empirical work. The remaining abstracts ($n = 36$, or 5%) were nonempirical and were excluded from subsequent analyses.

All *LSE* articles and SABER abstracts were de-identified, with information such as authors and institutions removed. The year of publication or presentation (for *LSE* articles and SABER abstracts) and the presentation format (for SABER abstracts) were retained, although the researchers were blinded to this information in the qualitative data analysis stage of the study to avoid potential biases.

We acknowledge that *LSE* articles and SABER abstracts addressed a variety of biology subdisciplines and also potentially included subsamples of non-biology students and/or faculty. However, as each of these articles and abstracts was deemed appropriate for the journal or conference by their respective reviewers, we view them as representative of the field of BER more broadly.

Coding Schemes

LSE articles and SABER abstracts were coded in relation to the three intraresearch parameters: research questions, study contexts, and methodologies. Data analysis took a combination of deductive and inductive approaches. For research questions and methodologies, external frameworks already existed (NRC, 2002; Earle *et al.*, 2013; Chiu *et al.*, 2016) and were used deductively as starting points for coding schemes. For study contexts (in *LSE* articles and SABER abstracts) and a more detailed analysis of methodologies (in *LSE* articles), coding schemes were generated inductively from the data using content analysis (Mayring, 2000). Briefly, preliminary codes were generated from an analysis of all the data. These codes were refined through independent analysis and consensus discussions in iterations of small subsets of the data. Disagreements were discussed, and consensus was reached.

Research Questions

This intraresearch parameter deals with the scientific nature of educational research (NRC, 2002) and includes the codes descriptive, causal, and mechanistic. Descriptive studies asked “What is happening?” questions and included work that was exploratory in nature and not necessarily hypothesis driven. Causal studies asked “Does x lead to y ?” questions and included work that examined the outcomes of a particular intervention in an attempt to identify causal (or at least correlational) relationships. Mechanistic studies asked “How or why does x lead to y ?” questions and included work that further examined a causal relationship to identify particular components that contributed to the effects.

Study Contexts

This intraresearch parameter refers to the population studied in an article or abstract. Our codes were generated inductively from the data and included K–12 (students or teachers), community college (students), undergraduate (students at 4-year institutions), graduate student or postdoctoral scholar, and faculty (instructors at any type of higher education institutions). More than one study population could be reported in an article or abstract.

We further examined the study population in relation to the presence or absence of demographics data being reported. Demographic descriptions regarding gender, race/ethnicity, or socioeconomic status (SES) were coded if those factors were clearly described in an article or abstract. While our codes do not represent a comprehensive list of demographic descriptions, gender, race/ethnicity, and SES were reported with high enough frequencies and consistencies to be tracked over time. We also acknowledge that race and ethnicity are related but different constructs; we combined race and ethnicity into one code, as they are often reported together or not clearly distinguished in the articles and abstracts that we examined.

Methodologies

This intraresearch parameter deals with the types of methods used in research and their underlying theoretical frameworks. Our first level of analysis included the codes qualitative, quantitative, and both (qualitative and quantitative; Earle et al., 2013; Chiu et al., 2016). Qualitative methodologies involve data that are not numerical in nature, whereas quantitative methodologies involve data that have numerical information. The “both” code indicated the incorporation of qualitative and quantitative methodologies. SABER abstracts especially have limited lengths and often did not include enough information to determine whether the qualitative and quantitative methodologies triangulated the data enough to fall under the strict definitions of mixed methods (Johnson and Onwuegbuzie, 2004), so we opted instead for the more generic “both” code.

The full-length nature of *LSE* articles allowed us to delve further into the methodologies and examine the types of data collected and reported. We analyzed mechanisms by which data were collected, and the codes were generated inductively from the data and included concept inventory or survey instruments, interviews, observations, or artifacts. Instruments, interviews and observations were further broken down in protocols that were validated in the existing literature or generated by the authors. We also quantified the number of articles that reported information on validity, reliability, and effect sizes.

Reliability Measurements

For *LSE* articles, five members of the research team (G.E.G., J.R., V.N.-F., P.C., E.S.) analyzed the data. Interrater reliability (IRR) was confirmed by selecting a random subsample (10%), redistributing those articles to the research team at random, recoding them, and comparing these codes with original codes, achieving an IRR of 84% and a Cohen's κ of 0.70 (Cohen, 1960), indicating a high level of reliability (Landis and Koch, 1977). A subsample (20%) of all SABER abstracts was analyzed by two members of the research team (S.M.L., B.K.S.), with an IRR of 85% and a Cohen's κ of 0.78 (Cohen, 1960). Because of the high level of reliability (Landis and Koch, 1977), the remain-

ing abstracts were coded by a single member of the research team. Any coding disagreements that arose in the *LSE* or SABER subsamples were discussed, and consensus was reached.

Statistical Analysis

We performed statistical analyses to test the hypothesis that the distribution of each intraresearch parameter was independent of year of publication or presentation (*LSE* articles and SABER abstracts) and the presentation format (SABER abstracts). We used contingency analysis to explore the distribution of one categorical variable (each of the intraresearch parameters) across another categorical variable (year or presentation format) and used chi-square statistics to test their independence (Cochran, 1952). Correspondence analysis, conceptually similar to principal component analysis (Jolliffe and Ringrose, 2006), is a multivariate method that decomposes the chi-square statistic from contingency tables into orthogonal factors and displays the set of categorical data in two-dimensional descriptive, graphical form (Greenacre, 2010). Points on the two major dimensions identified in correspondence analyses were put into groups by cluster analysis with hierarchical clustering. All statistical analyses were performed in JMP Pro (versions 11.0–13.0).

In similar studies, spans of at least 5 years yielded results that allowed the authors to examine change and variation over time (Koh et al., 2014; O'Toole et al., 2018). Accordingly, we combined *LSE* articles into multiyear blocks of approximately 5 years (2002–2006, 2007–2011, and 2012–2015). Rather than aligning the 2011–2015 *LSE* data with the SABER 2011–2015 data, we decided to use the blocks 2002–2006, 2007–2011, and 2012–2015 because of the smaller number of articles in the earlier years of publication of *LSE*. For SABER abstracts, we analyzed the results separated by individual years, because the number of abstracts per year was large and the rubrics for abstract selection changed multiple times within the 5-year time frame.

Characterization of SABER Attendees

To characterize the individuals who attend and present at SABER, we surveyed the 2016 national meeting attendees ($n = 194/284$, 68.3% response rate) in one of the meeting's plenary sessions. Each attendee was asked to provide information on a handout regarding current role, institution type, year and field of most recent degree, years teaching in higher education, number of SABER meetings attended, and number of education publications. This data collection was performed with approval from the University of California Irvine Institutional Review Board (HS# 2016-2669).

RESULTS

Characterization of Research Questions

We first asked whether and how the types of research questions in *LSE* articles and SABER abstracts have changed over time (Figure 1). Most *LSE* articles are descriptive in nature (52%), followed by causal studies (32%). Interestingly, SABER abstracts have the reverse trend, with 65% of studies being causal and 32% descriptive in nature. In both data sources, few studies focused on mechanisms: 6% in *LSE* articles and 2% in SABER abstracts. The predominant type of research question changed in *LSE* articles ($p < 0.0001$; Supplemental Table S1) but remained the same in SABER abstracts over time ($p = 0.38$;

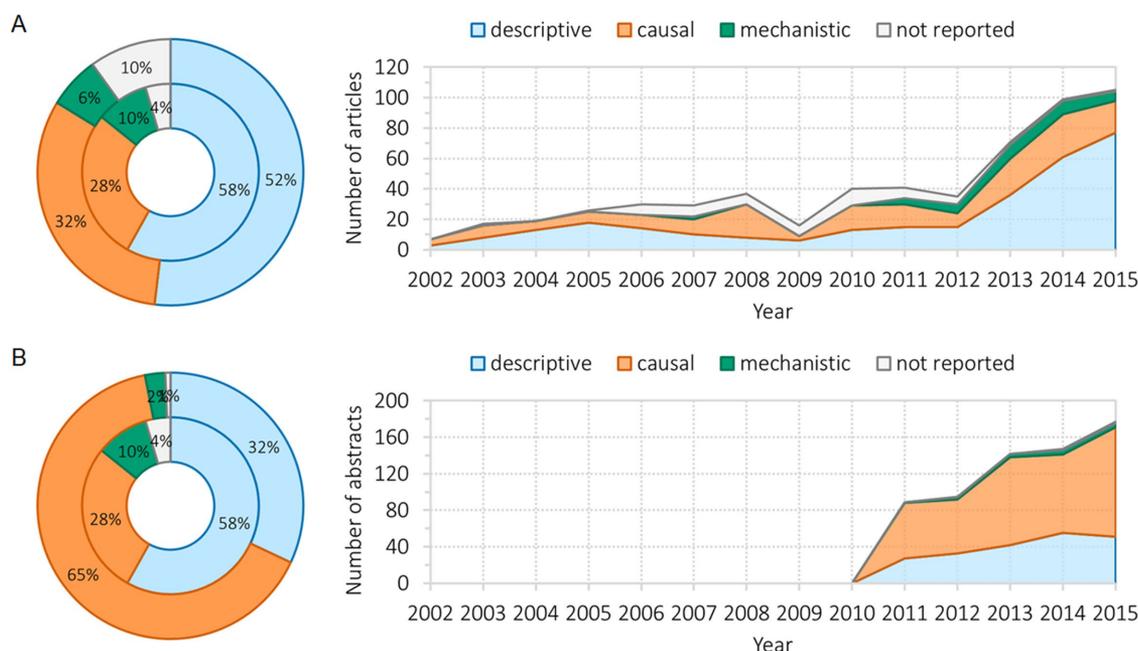


FIGURE 1. Types of research questions found in *LSE* articles and SABER abstracts. (A) Research questions found in *LSE* articles were coded as descriptive, causal, mechanistic, or not reported. The rings on the left represent the complete *LSE* data from 2002 to 2015 (outer ring) and *LSE* data from 2011 to 2015 (inner ring) for comparison with the SABER data in the corresponding years. The graph on the right illustrates the percentage of research questions found in *LSE* articles on an annual basis. (B) The same data are reported for SABER abstracts from 2011 to 2015. The rings on the left represent the 2011–2015 data for SABER abstracts (outer ring) and *LSE* articles (inner ring). The graph on the right illustrates the percentage of research questions found in SABER abstracts on an annual basis.

Supplemental Table S2). In *LSE* articles, causal studies were the majority in 2002–2006 (42%) and 2007–2011 (51%); this percentage dropped to 31% for causal research questions in 2012–2015, with descriptive research questions rising to 56%. Additionally, mechanistic studies were not found until 2007–2011. In SABER abstracts, the predominant type of research question was consistently causal: 69% in 2011 and 68% in 2015.

Of the presented studies, a large proportion focus on interventions: 48% in *LSE* articles and 69% in SABER abstracts. *LSE* articles are richer in information than SABER abstracts, allowing for more in-depth analysis. In *LSE* articles, most of the interventions are classroom based, either in a particular content area (44% of all interventions or 21% of all articles) or a course-wide intervention (40% of all interventions or 19% of all articles; Supplemental Figure S1).

Characterization of Study Contexts

For study population, a majority of both *LSE* articles and SABER abstracts collected data on undergraduates (79 and 69%, respectively; Figure 2). These percentages are similar after taking into account that ~8% of SABER abstracts did not report a study population at all. Much smaller numbers of studies focus on K–12 settings, faculty, and graduate students or postdoctoral scholars. Particularly of note, community college studies were almost nonexistent, at just 1% for both *LSE* articles and SABER abstracts. In *LSE* studies, there were significant differences over time in regard to the study population ($p < 0.0001$; Supplemental Table S3). While undergraduates represented the majority of study populations over all time periods, this dropped from 86% in 2007–2011 to 68% in 2012–2015. Correspondingly small

increases were observed in all other study populations. For SABER abstracts, there was no difference over time for study population ($p = 0.45$; Supplemental Table S4).

For demographic descriptions, in both data sets, SES is the least frequently reported among the three metrics in our codes (gender, race/ethnicity, and SES), although all three metrics have been collected consistently over time (Figure 3). In *LSE*, gender, race/ethnicity, and SES were reported only in 26, 26, and 8% of articles in 2002–2006; these values were 33, 19, and 4%, respectively, in 2007–2011 and 44, 25, and 4%, respectively in 2012–2015. These differences were statistically significant for gender ($p = 0.02$) but not race/ethnicity ($p = 0.38$) or SES ($p = 0.39$; Supplemental Table S5). In SABER abstracts, 7% reported gender, 3% race/ethnicity, and 0% SES in 2011, compared with 3, 5, and 1%, respectively, in 2013, and 9, 7, and 1%, respectively, in 2015 (Supplemental Table S6). The lower frequency of these data found in SABER abstracts is likely due to the space restrictions of an abstract submission. The difference by year is significant for gender ($p = 0.03$), likely because of the lower percentages in the middle years, but not for race/ethnicity ($p = 0.21$) and SES ($p = 0.87$).

Characterization of Methodologies

For the final intraresearch parameter, we asked what types of research methodologies are used in these studies. Both *LSE* articles and SABER abstracts appear to be heavily dependent on quantitative methodologies. For example, 37% of *LSE* studies use quantitative methodologies, while 48% use both quantitative and qualitative, whereas only 11% of studies collected only qualitative data (Figure 4A). For SABER, 34% of abstracts

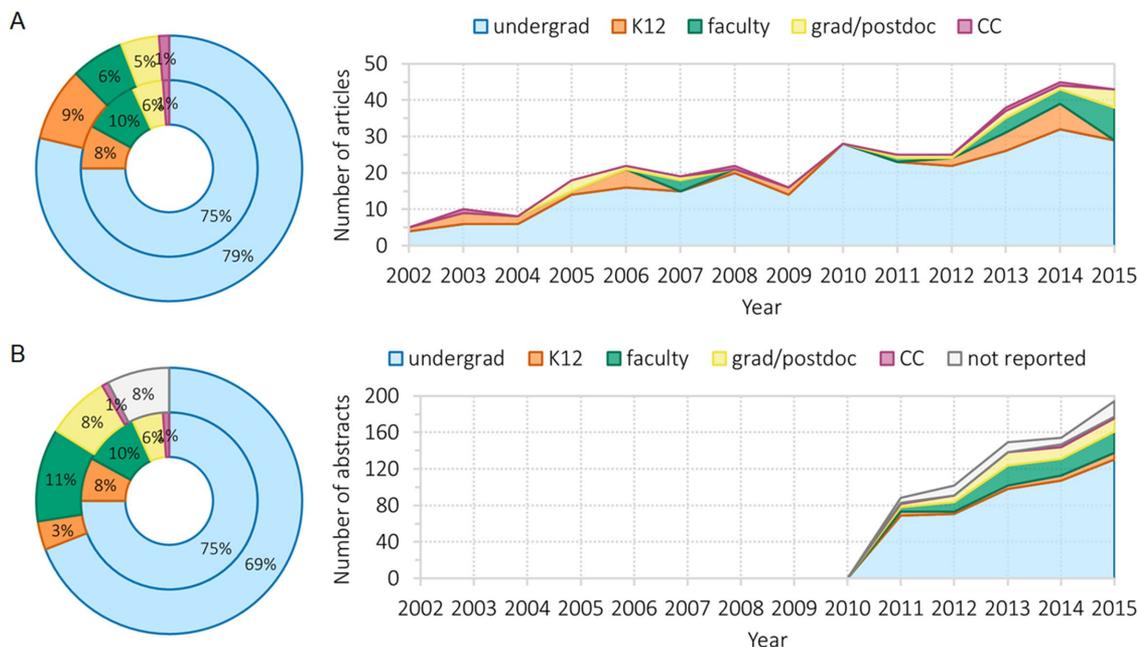


FIGURE 2. Study context for *LSE* articles and SABER abstracts. (A) Study contexts found in *LSE* articles were coded as undergraduate, K–12, faculty, graduate student or postdoctoral scholars (grad/postdoc), community college (CC), or not reported. The rings on the left represent the complete *LSE* data from 2002 to 2015 (outer ring) and *LSE* data from 2011 to 2015 (inner ring) for comparison with the SABER data in the corresponding years. The graph on the right illustrates the percentage of studies with the particular study context found in *LSE* articles on an annual basis. (B) The same data are reported for SABER abstracts from 2011 to 2015. The rings on the left represents the 2011–2015 data for SABER abstracts (outer ring) and *LSE* articles (inner ring). The graph on the right illustrates the percentage of studies with the particular study context found in SABER abstracts on an annual basis.

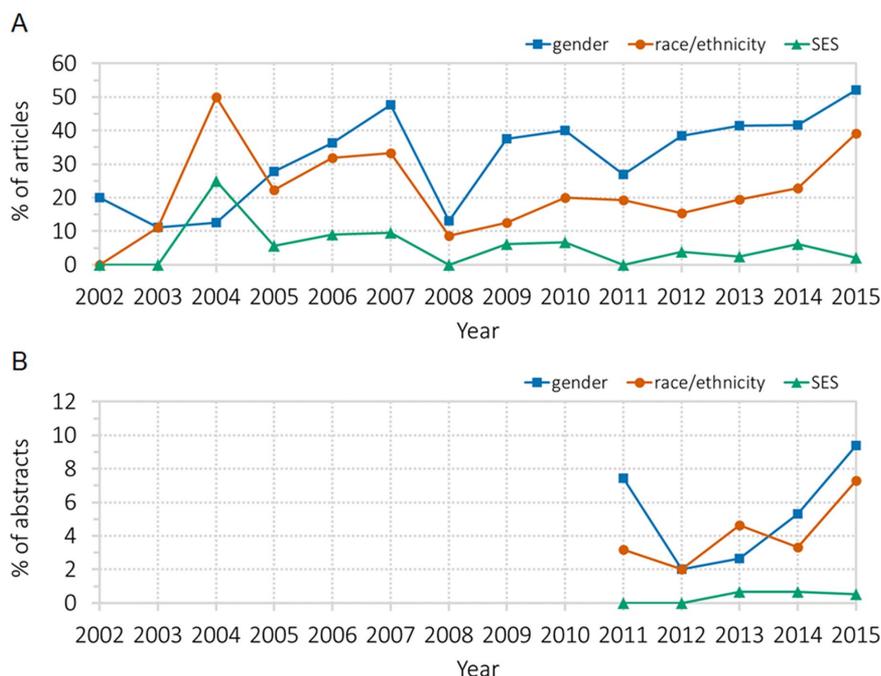


FIGURE 3. Reporting of gender, race, and SES in *LSE* articles and SABER abstracts. (A) The percentage of *LSE* articles that include gender, race, and SES is presented on an annual basis from 2002 to 2015. (B) The percentage of SABER abstracts that include gender, race, and SES is presented on an annual basis from 2011 to 2015.

contain solely quantitative data, and 38% use a combination of quantitative and qualitative methodologies; fewer abstracts employ only qualitative methodologies (14%), while some do not report methods (14%; Figure 4B). Interestingly, we found that SABER presentation formats correlate with methodologies (Supplemental Figure S2), as well as topics of study (Supplemental Figure S3). We identified three predominant combinations of presentation formats and methodologies at SABER: talks with quantitative or both methods, posters with qualitative methods, and roundtables with no methods reported (Supplemental Figure S2). Methodologies have not changed over time to a statistically significant degree for either *LSE* articles ($p = 0.29$; Supplemental Table S7) or SABER abstracts ($p = 0.08$; Supplemental Table S8).

LSE articles contain much more information than SABER abstracts, allowing us to delve further into their methodologies. We examined the data-collection instruments used by the authors (Figure 5A and Supplemental Table S9). Statistical analysis indicated a significant relationship

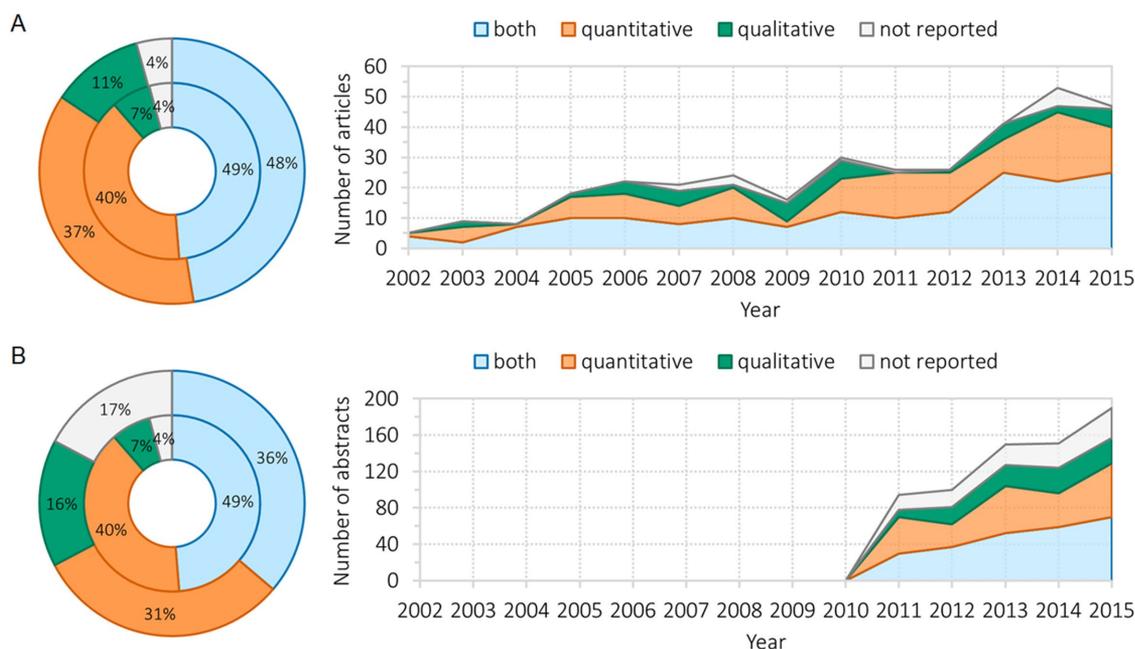


FIGURE 4. Research methodologies for LSE articles and SABER abstracts. (A) Research methodologies found in LSE articles were coded as quantitative, qualitative, both, or not reported. The rings on the left represent the complete LSE data from 2002 to 2015 (outer ring) and LSE data from 2011 to 2015 (inner ring) for comparison with the SABER data in the corresponding years. The graph on the right illustrates the percentage of studies with the particular research methodologies found in LSE articles on an annual basis. (B) The same data are reported for SABER abstracts from 2011 to 2015. The rings on the left represent the 2011–2015 data for SABER abstracts (outer ring) and LSE articles (inner ring). The graph on the right illustrates the percentage of studies with the particular research methodologies found in SABER abstracts on an annual basis.

between data-collection instrument and the year of publication ($p = 0.02$). In 2002–2006, the data-collection instruments used include researcher-designed (“new”) written instruments such as course exams (53%), standardized (“existing”) written instruments such as concept inventories (5%), and artifacts such as lesson plans (17%). In 2007–2011, there was an increase in the use of existing written instruments (14%) and artifacts (26%). In 2012–2015, there was an increase in the use of new interview and observation protocols (10 and 3%, respectively), which coincided with the increase in the number of descriptive research questions (Figure 1 and Supplemental Table S2).

While the variety of data-collection instruments has increased over time, a sizable portion of these instruments were created by the researchers. We were curious to see whether the instruments are being checked for validity and reliability. We found this to be the case, as reporting of both validity and reliability values have risen over time (Figure 5B). In 2002–2006, only 26% of articles reported validity, compared with 37% in 2007–2011 and 65% in 2012–2015 ($p < 0.0001$). For articles reporting reliability, the percentage increased from 24 to 28 and 60% at the same time points ($p < 0.0001$). The emphasis on reporting effect size in BER studies in recent years (Maher *et al.*, 2013) is also evident from the number of articles including the statistic over time: from 3% in 2002–2006 and 2007–2011 to 19% in 2012–2015 (Figure 5B and Supplemental Table S10), which should provide important statistical information for future meta-analyses (Ruiz-Primo *et al.*, 2011).

Study Limitations

Our study has several limitations based on the chosen methodology. First, we used one journal and one conference as proxies for the BER community. We have argued that LSE and SABER are representative of the community, and these contexts are the main scholastic avenues for dissemination of BER in the United States. However, this geographically limits our data set, and accordingly, our analysis revealed that more than 90% of LSE articles described studies that were conducted in the United States. Nonetheless, our work complements existing studies that have examined BER trends in other geographic regions (Topsakal *et al.*, 2012; Gul and Sozbilir, 2015; Chiu *et al.*, 2016) and a recent study that largely focused on European journals (Gul and Sozbilir, 2016).

Second, we chose to focus on LSE articles and SABER abstracts with empirical research. While we acknowledge the value and place of theoretical research and perspectives, we ultimately limited our study to the articles and abstracts with enough details to examine the three related intraresearch parameters: research questions, study contexts, and methodologies. Nonetheless, the focus on empirical research did not significantly alter or bias the nature of our data set. Of the non-empirical studies excluded from our data set, a vast majority were curricular and product development projects that did not report empirical data and thus should not be counted as theoretical research. As such, the lack of theoretical research in our study is more reflective of the nature of our data set than the methodological decisions that we have made.

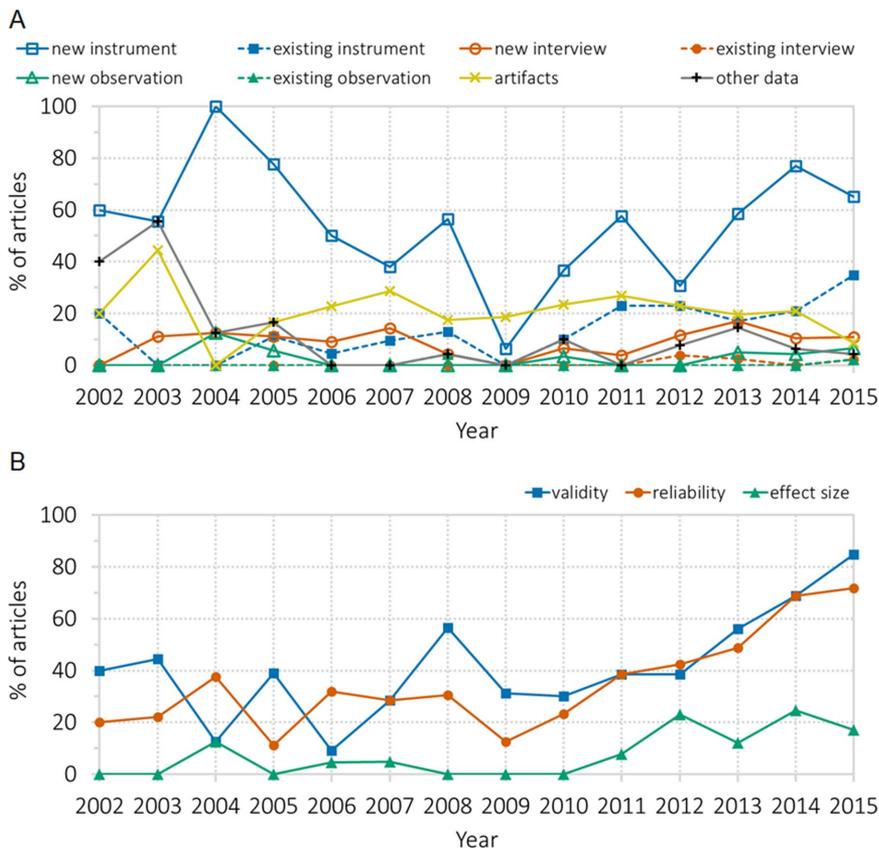


FIGURE 5. Data-collection instruments and data reported for *LSE* articles. (A) Data-collection instruments found in *LSE* articles were coded as new or existing instrument, new or existing observation protocol, new or existing interview protocol, artifact, or other data. The percentage of each is presented on an annual basis from 2002 to 2015. (B) The percentage of *LSE* articles that include validity, reliability, and effect size is presented on an annual basis from 2002 to 2015.

DISCUSSION

As the field of BER matures, self-reflection on its trajectory will assist stakeholders in defining cohesive results and agendas. Reflection allows the community not only to define itself as a unique field of academic inquiry (setting borders), but also to remain cognizant of the ways in which it is enhanced by other fields, while recognizing its own limitations that may be reinscribed by normalized practices (blurring borders). Our current study examines and summarizes the necessary information for such self-reflections to proceed by examining trends in peer-reviewed research at two of the foremost venues specific to dissemination of BER scholarship: *LSE* and SABER.

By focusing our reflection on two sources of data, we triangulate current and longitudinal trends in the field. For example, much of the research elucidated by these venues highlights a strong preference for causal research questions, a disproportionate focus on undergraduate student as the population of interest, and a reliance on quantitative data to make claims. These trends are perhaps unsurprising based on the historical development of *LSE* and SABER by biologists interested in pursuing education research (DeHaan, 2011). In our survey of participants at the SABER national meeting in 2016, 92% had a master's degree or PhD in a life sciences discipline, and 77%

had taught in higher education for more than 3 years. Such “crossover” biology education researchers, who are crossing disciplinary boundaries from the life sciences, may be predicted to rely on the data sources and research methodologies that are most familiar to them from their own (post)positivist research paradigms in the biological sciences (i.e., bench or field research with quantitative, generalizable results). In addition, as individual classroom research might prompt many such scholars to move into this type of BER work, it seems reasonable that they would study their own classrooms with populations of undergraduate students.

One of the observed longitudinal trends is the reduction in causal research questions and the corresponding increase in descriptive research questions in *LSE* over the history of the journal. This observation suggests that work published in *LSE* is changing as the field of BER is developing or maturing over time. In contrast, SABER abstracts have seen relatively little change in the prevalence of causal and descriptive research questions over time. This lack of change may be because of the short history of SABER as an association, as other studies have used spans of at least 5 years to examine change and variations over time (Koh et al., 2014; O'Toole et al., 2018). Furthermore, a substantial subset of SABER attendees were participating in the meeting for the first time (39.7%) based on data from our 2016 national meeting survey. These first-timers are likely to be crossover biology education researchers who have disciplinary training in the life sciences and are more accustomed to the corresponding (post)positivist research paradigms, thus sustaining the prevalence of causal research questions.

Triangulation also helps reveal gaps in the trajectory of research in BER. For example, mechanistic research questions that get at the how and why for inclusion, learning, and teaching in biology education have only begun to emerge in the past few years. For study contexts, studies on community college students were only represented in 1% of *LSE* articles and SABER abstracts; a similar percentage was reported in a recent analogous study of physics education research (Kanim and Cid, 2017), despite the fact that more than 40% of undergraduates begin their postsecondary education at community colleges, especially underrepresented minorities (51% of Hispanic students and 41% of African-American students; Crisp and Delgado, 2014). Consistent reporting of student demographics in many sample populations was also lacking, although we observed an increasing trend of reporting of gender in *LSE* articles over time. This lack of consistent reporting has important implications for research and practice on the opportunities and challenges related to diversity, equity, and inclusion in STEM. When interventions are only tested on a subset of a population, or when

theories are built using a limited demographic, the generalizability of these interventions and theories may not extend to underrepresented populations. To promote more inclusive study contexts in the field, journals can dedicate special issues focused on, for example, studies on community colleges; likewise, conferences can create themed sessions to encourage participation from more diverse demographics of study populations.

There is a lack of focus on qualitative or theoretical inquiry as a valid means of contribution to the field. This may be most apparent in the SABER abstracts, in which qualitative inquiries have been historically and disproportionately relegated to the poster format, which is often perceived as lower prestige. This difference in prestige can be evidenced by the call for SABER abstracts, in which posters are requested as “new or developing project[s]” and talks as “results that are being prepared for publication” (SABER, 2018). Furthermore, as much of the examined BER work was conducted by researchers trained in biology, it is not surprising that there is little to no theoretical work. As the field grows, it naturally can involve biology education researchers collaborating with other more mature DBER fields, including chemistry, engineering, mathematics, or physics, or with education researchers, cognitive scientists, psychologists, sociologists, and philosophers, who are much more grounded in the theories that underlie the intervention-based studies being conducted in BER. As Dolan (2015) notes in her commentary, “Biology Education Research 2.0,” the field of BER will largely need to move into asking more how and why research questions, which might be appropriately answered with more qualitative data sources. To facilitate such development, our field will likely need to continue to include more scholars who are trained in and comfortable with conducting and consuming rigorous qualitative research.

These results and the discussion here should serve as a renewed call for thoughtful introspection for our community. What are the next steps in using these data to reflect on and move the field of BER forward? Our data reveal not only the historical trends in BER but also the related gaps poised for new and exciting research questions, contexts, and methodologies. For example, what would be an appropriate changing distribution of scholarship with descriptive, causal, or mechanistic research questions over time? Could our field benefit from branching out to different study contexts such as informal education, clinical laboratory settings similar to cognitive science and psychology, or empirical literature research such as this paper? How should we train the next generation of emerging BER scholars in such diverse research methodologies that will allow our field to continue to grow and mature? It is our hope that this work would provide the necessary information for the community to reflect on our own identities as biology education researchers: Is this who we want to be as a field, and is this how we want our inquiry to be defined?

ACKNOWLEDGMENTS

We are grateful to all past authors at LSE and past participants at SABER who contributed articles and abstracts to this study and the 2016 participants at SABER for engaging with the survey at the meeting. We thank S. Hoskins, A.-M. Hoskinson, C. Trujillo, and M. P. Wenderoth for discussions originating at SABER that ultimately developed into this project.

REFERENCES

- Aikens, M. L., Corwin, L. A., Andres, T. A., Couch, B. A., Eddy, S. L., McDonnell, L., & Trujillo, G. (2016). A guide for graduate students interested in post-doctoral positions in biology education research. *CBE—Life Sciences Education, 15*, es10.
- American Association for the Advancement of the Science. (2011). *Vision and change in undergraduate biology education*. Washington, DC.
- American Society for Cell Biology. (2017). *Information for authors*. Retrieved August 13, 2018, from www.lifescied.org/info-for-authors
- Amin, T. G. (2015). Conceptual metaphor and the study of conceptual change: Research synthesis and future directions. *International Journal of Science Education, 37*, 966–991. doi: 10.1080/09500693.2015.1025313
- Bowen, C. W. (2000). A quantitative literature review of cooperative learning effects on high school and college chemistry achievement. *Journal of Chemical Education, 77*(1)116. doi: 10.1021/ed077p116
- Boyer, E. (1998). *The Boyer Commission on Educating Undergraduates in the Research University, Reinventing Undergraduate Education: A blueprint for America's research universities* (p. 46). Stony Brook, NY.
- Chang, Y., Chang, C., & Tseng, Y. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology, 19*, 315–331. doi: 10.1007/s10956-009-9202-2
- Chiu, M., Tam, H., & Yen, M. (2016). Trends in science education research in Taiwan: A content analysis of the *Chinese Journal of Science Education* from 1993 to 2012. In Chiu, M. (Ed.), *Science education research and practices in Taiwan* (pp. 43–78). Singapore: Springer.
- Cochran, W. G. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics, 23*, 315–345. doi: 10.1214/aoms/1177729380
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.
- Crisp, G., & Delgado, C. (2014). The impact of developmental education on community college persistence and vertical transfer. *Community College Review, 42*(2), 99–117. <https://doi.org/10.1177/0091552113516488>
- Darragh, L. (2016). Identity research in mathematics education. *Educational Studies in Mathematics, 93*, 19–33.
- DeHaan, R. L. (2011). Education Research in the Biological Sciences: A Nine-Decade Review. Paper presented at the Second Committee Meeting on the Status, Contributions, and Future Directions of Discipline-Based Education Research held October 18–19, 2010, in Washington, DC. Retrieved February 23, 2019, from http://sites.nationalacademies.org/dbasse/bose/dbasse_080124
- Dirks, C. (2011). The Current Status and Future Direction of Biology Education Research. Paper presented at the Second Committee Meeting on the Status, Contributions, and Future Directions of Discipline-Based Education Research held October 18–19, 2010, in Washington, DC. Retrieved February 23, 2019, from http://sites.nationalacademies.org/dbasse/bose/dbasse_080124
- Dolan, E. (2015). Biology education research 2.0. *CBE—Life Sciences Education, 14*(4), ed1.
- Earle, J., Maynard, R., Neild, R. C., Easton, J. Q., Ferrini-Mundy, J., Albro, E., & Winter, S. (2013). *Common guidelines for education research and development*. Washington, DC: Institute of Education Sciences and National Science Foundation.
- Fensham, P. J. (2004). *Defining an identity: The evolution of science education as a field of research*. Dordrecht, Netherlands: Kluwer Academic.
- 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 USA, 111*(23), 8410–8415. doi: 10.1073/pnas.1319030111
- Gijbels, D., Dochy, F., Bossche, P. V. d., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27–61. doi: 10.3102/00346543075001027
- Greenacre, M. (2010). Correspondence analysis of raw data. *Ecology, 91*, 958–963.
- Gul, S., & Sozibilir, M. (2015). Biology education research trends in Turkey. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(1), 93–109.

- Gul, S., & Sozibilir, M. (2016). International trends in biology education research from 1997 to 2014: A content analysis of papers in selected journals. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(6), 1631–1651.
- House of Lords, Science Technology Committee. (2012). *Higher education in science, technology, engineering and mathematics (STEM) subjects: 2nd Report of Session 2012–13*. London: Stationery Office.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33, 14–26. doi: 10.3102/0013189X033007014
- Jolliffe, I. T., & Ringrose, T. J. (2006). Canonical correspondence analysis. In Kotz, S., Read, C. B., Balakrishnan, N., Vidakovic, B., & Johnson, N. L. (Eds.), *Encyclopedia of statistical sciences*. Retrieved February 23, 2019, from <https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471667196.ess0609.pub2>
- Kanim, S., & Cid, X. C. (2017). *The demographics of physics education research*. Retrieved February 23, 2019, from arXiv:1710.02598.
- Koh, E., Cho, Y. H., Caleon, I., & Wei, Y. (2014). Where are we now? Research trends in the learning sciences. In Polman, J. L., Kyza, E. A., O'Neill, D. K., Tabak, I., Penuel, W. R., Jurrow, A. S., O'Connor, K., Lee, T., and D'Amico, L. (Eds.), *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS)*, 1, 535–542. Retrieved February 23, 2019, from www.isls.org/icls/2014/Proceedings.html
- Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lee, M., Wu, Y., & Tsai, C. (2009). Research trends in science education from 2003–2007: A content analysis of publications in selected journals. *International Journal of Science Education*, 31, 1999–2020. doi: 10.1080/0950069082314876
- Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE—Life Sciences Education*, 12, 345–351. doi: 10.1187/cbe.13-04-0082
- Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM: Country comparisons (Report for the Australian Council of Learned Academies)*. Retrieved February 23, 2019, from www.acola.org.au
- Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 1(2). doi: <http://dx.doi.org/10.17169/fqs-1.2.1089>
- National Academies of Sciences, Engineering, and Medicine (NAS). (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington, DC: National Academies Press.
- NAS. (2016). *Barriers and opportunities for 2-year and 4-year STEM degrees: Systemic change to support students' diverse pathways*. Washington, DC: National Academies Press.
- NAS. (2018). *Indicators for monitoring undergraduate STEM education*. Washington, DC: National Academies Press.
- National Research Council (NRC). (2002). *Scientific research in education*. Washington, DC: National Academies Press.
- NRC. (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.
- NRC. (2009). *A new biology for the 21st century*. Washington, DC: National Academies Press.
- NRC. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.
- NRC. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Washington, DC: National Academies Press.
- O'Toole, J. M., Freestone, M., McKoy, K. S., & Duckworth, B. (2018). Types, topics and trends: A ten-year review of research journals in science education. *Education Sciences*, 8, 73. doi: 10.3390/educsci8020073
- President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: U.S. Government Office of Science and Technology.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223–231. doi: 10.1002/j.2168-9830.2004.tb00809.x
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, 331(6022), 1269–1270. doi: 10.1126/science.1198976
- Scimago. (2007). *SJR: SCImago journal & country rank*. Retrieved July 21, 2015, from www.scimagojr.com
- Society for the Advancement of Biology Education Research. (2018). Poster abstract submission rubric. Retrieved February 23, 2019, from <https://saberbio.wildapricot.org/Call-for-abstracts>
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51. doi:10.3102/00346543069001021
- Taylor, J., Furtak, E., Kowalski, S., Martinez, A., Slavin, R., Stuhlsatz, M., & Wilson, C. (2016). Emergent themes from recent research syntheses in science education and their implications for research design, replication, and reporting practices. *Journal of Research in Science Teaching*, 53(8), 1216–1231. doi: 10.1002/tea.21327
- Topsakal, U. U., Calik, M., & Cavus, R. (2012). What trends do Turkish biology education studies indicate? *International Journal of Environmental & Science Education*, 7(4), 639–649.
- Vernon, D. T., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine*, 68(7), 550–563.
- Ward, S. (2002). Cell biology education. *Cell Biology Education*, 1, 1–2. doi: 10.1187/cbe.02-04-0012