

REVIEW

Open Access



Modern international large-scale assessment in education: an integrative review and mapping of the literature

Daniel Hernández-Torrano*  and Matthew G. R. Courtney 

*Correspondence:
daniel.torrano@nu.edu.kz;
d.hernandeztorrano@gmail.com
Graduate School
of Education, Nazarbayev
University, 53 Kabanbay
Baty, Nur-Sultan 010000,
Kazakhstan

Abstract

Research in international large-scale assessment (ILSA) has become an increasingly popular field of study in education. Consequently, interest and debate in the field by practitioners, researchers, policymakers, and the public has grown over the past decades. This study adopts a descriptive bibliometric approach to map modern research on ILSA in education and provide an up-to-date picture of the recent developments and structure of the field. The analysis of 2,233 journal articles indexed in the Web of Sciences database revealed that ILSA research in education is an emerging field in a stage of exponential growth that has become increasingly international with recent substantive contributions from China, Spain, and Turkey. Research in the field is currently produced by a tupid network of scholars with diverse geographical backgrounds that engage frequently in national and international research collaborations. Also, the field is relatively interdisciplinary and has developed grounded on nine differentiated historical paths. The PISA program has received the greatest attention in the field, and a wide variety of topics have been addressed in the literature in the last decades, including equity and quality education, globalization and education policy, measurement and statistics, student motivation and self-concept, and interpersonal relationships. The paper concludes by pointing to the potential of future ILSA research to make use of new more relevant instrumentation, data linkages, and trans-regional collaborations.

Keywords: Bibliometric analysis, Large scale assessment, Education, PISA, TIMSS, PIRLS, ICCS, ICILS

Introduction

Studies on international large-scale assessment (ILSA) in education, such as the Programme for International Student Assessment (PISA) or the Trends in Mathematics and Science Study (TIMSS), have received increasing attention from educational practitioners, researchers, policymakers, and the public at large. The goal of research in ILSA is to shed light on how educational infrastructures, policies, and contexts operate and how related factors might contribute to educational and social outcomes across different schools and countries. The outcomes of such studies feature prominently in discussions about the comparative quality of educational jurisdictions, and have had important policy implications across and within multiple countries (Johansson, 2016).

Although the theoretical conception, design, implementation, scale validation, and descriptive analysis for these large-scale projects are generally conducted by international institutions such as the Organisation for Economic Co-operation and Development (OECD), the International Association for the Evaluation of Educational Achievement (IAE), and the United Nations Educational, Scientific and Cultural Organization (UNESCO), some programs also invite public participation –for example, “countries and economies are invited to submit questions that are then added to items developed by the OECD’s experts and contractors” (PISA, 2020; PisaFAQ). The sampling designs for ILSAs shape possible analyses and are characterized by collections of data from large representative samples of students and schools across multiple jurisdictions. While large representative samples of students sit in standardized assessments and give responses to questions about their school experience, principals, for example, report on the implementation of specific school policies, teachers report on aspects of the classroom, and parents report on home support for learning.

ILSA research has come a long way from the early international studies in education (e.g., FIMS) (Rutkowski et al., 2014). Today, ILSA research in education is a flourishing field that has experienced significant growth in recent decades (Addey et al., 2017; Kamens & McNeely, 2010). The modern scholarship in this broad field includes publications about methods and analytic procedures, research on secondary analysis of ILSA datasets, and studies that apply a social science framework to the ILSA phenomenon. The field brings together “scholars and practitioners who are contributing to the analysis, critique, and development of large-scale assessment methods and results, as well as their implications for education policy” (CIES, 2021).

As the volume of research in ILSA grows and empirical evidence accumulates in the field, it is increasingly necessary to synthesize the knowledge generated to offer educational scholars, practitioners, and policymakers a coherent vision of the recent research trends in the literature as well as the structural cornerstones of the field. Such epistemological syntheses are important because they envision the past, present, and future developments of a field. While program-specific reviews have provided some insights into the contribution of some ILSA programs at a given point in time (e.g., see Hopfenbeck et al., 2018 for systematic review of PISA research to 2015), a more thorough review of research on ILSA in education more generally, one that provides clarity about the overall growth and trajectory of this broad field, has yet to be undertaken. The purpose of this study is to provide a quantitative synthesis of modern research available revolving around the five major current and ongoing ILSA programs in education: PISA, TIMSS, PIRLS, ICCS, and ICILS. We use a descriptive bibliometric approach to map the growing literature in research on ILSA in education and describe recent developments of the field, as well as its actual structure based on publication and citation data.

The study is guided by the following research questions:

- RQ1: What is the volume and growth trajectory of modern research on ILSA in education?
- RQ2: What journals, articles, authors, and countries have had the highest impact on the dissemination of modern research in the field?

- RQ3: As evidenced by the literature, what are the collaborative links among researchers in ILSA in education to produce knowledge in the field?
- RQ4: What disciplines and historical paths underlie the foundations of the field?
- RQ5: As reflected in the literature, what research themes have captured the greatest attention in modern ILSA research in research?

This study contributes to the literature by examining how recent research on ILSA in education is built on the basis of different social, intellectual, and conceptual frameworks and by identifying the key players in the development of the field. The findings have the potential to inform future studies by identifying strengths and gaps in ILSA research in education in terms of its growth patterns, relevance, and coverage.

In the following sections, a brief history of ILSA in education is presented prior to a description of common subject areas. This is followed by a discussion vis-a-vis the motivations of participating jurisdictions. Thereafter, details of the prominent ILSA programs, focal to this study, are provided prior to a review of the strengths and weaknesses of such research programs. Finally, a short summary of previous reviews of ILSA research in education is provided prior to providing a rationale for the current study.

A brief history and scope

Though the field of ILSA in education has become quite prominent over the past two decades, its history can be traced back to the late 1950s. At the UNESCO Institute for Education in Hamburg, a group of educational researchers met and discussed the possibilities of conducting research on student academic performance and its antecedents across multiple countries (Husén, 1979). One of the key initial intentions of this proposed enquiry was for countries to learn from the experiences of others and to avoid developments that produced less than satisfactory results (Husén, 1979). This initial meeting represents the origins of the IEA, which, from the 1960s, has conducted regular studies in international large-scale assessment. Today, the IEA's TIMSS with its focus on Mathematics and Science and Progress in International Reading Literacy Study (PIRLS) with its focus on Reading Literacy, alongside the OECD's PISA triennial surveys, represent prominent ILSA research programs with global reach.

Broadly, the shift in early ILSA studies in the 1960s to that in the current era can be characterized politically, temporarily, and disciplinarily: politically, by a shift in focus with countries not only interested in educational accountability, but now also interested in becoming embedded in world society (Pizmony-Levy, 2013); temporarily, by a more high-paced, synchronized administration, and more immediate publication of results (Landahl, 2020); and, disciplinarily, by an expansion of common academic subjects to include problem-solving, and, more recently, research in ICT skills and civic education.

The expansion to include ICT and civic education has enabled the examination of a more broad range of cognitive, psychological, and educational processes and outcomes. For example, Torney-Purta and Amedeo (2013) posit that while civic education is not commonly conceptualized as ILSA research, it has “the potential to contribute to understanding many aspects of the school's role cross countries” (p. 254). To the authors' first point about conceptualization, we point out that the first civics education study was authorized by the IEA in the early 1970s (Torney et al.,

1975). To the authors' second point about potential contribution to understand the roles of schools, we concur and point to recent studies that buttress this argument—for example, a study by Maurissen et al. (2020) contributed to an understanding of how *student*-level processes, such as student–teacher relationships, contribute to perceived civic outcomes within schools. In addition, studies in civic education have enabled researchers to better understand youths' civic identities (see Knowles et al., 2018 for full review). For these reasons, this study also includes the more recently established International Computer and Information Literacy Study (ICILS) and Civic and Citizenship Education Study (ICCS), both administered by the IEA.

Motivations of participating governments and jurisdictions

Research by DeBoer (2010) and Kijima (2010) identifies four forms of motivations for a jurisdiction's participation in ILSA study programs: the *rational choice model*, the *policy diffusion model*, the *macro-dissatisfaction perspective*, and, herein defined as, the *financial aid model* (Kijima, 2010). Jurisdictions that are driven by the *rational choice model* partake in ILSA studies to inform decision-making and uphold their international reputation, while states that function under the *policy diffusion model* are motivated by identifying effective educational processes and have an interest in promoting the transfer of specific educational practices. However, jurisdictions with a *macro-dissatisfaction perspective* see such international studies as “bringing attention to a perceived crisis and prompting a focus on education” (Torney-Purta & Amedeo, 2013, p. 249). Finally, empirical work undertaken by Kijima (2010) suggests that recently participating jurisdictions may be motivated by an associated increase in educational foreign aid. Kijima found that countries that participate in major cross-national assessments, such as PISA, receive an average of 37% more funding than non-participating country counterparts. While it might be argued that these four motivations exist, we note that they are not mutually exclusive.

Torney-Purta and Amedeo (2013) state that for jurisdictions “operating according to the rational choice and the policy diffusion models, secondary analysis has obvious advantages” (p. 249). Indeed such analysis is better aligned with initial goals. Though, we argue that secondary analysis may benefit all participating jurisdictions regardless of initial motivational orientations—while participating jurisdictions must manage their country's reported rank-ordering of student performance and potential political fallout, all jurisdictions can also make use of the research infrastructure that ILSAs provide (Johansson, 2016). Research has also suggested that jurisdictional motivation for participation may also be driven financially.

Nevertheless, countries and jurisdictions leverage off the fact that they (1) have access to large-scale representative data, rather than common smaller convenience samples, (2) can save on the time and costs associated with research design, data collection, and management, (3) are able to undertake research that identifies student-, school-, and country-level related conditions that may be effective for learning outcomes of interest, (4) test competing statistical models, and (5) replicate and compare findings across jurisdictions and regions (Donnellan et al., 2011; Johansson, 2016).

Description of ILSAs in education

The definition of ILSAs adopted in the present study resembles that given by Bos (2002): “Studies in which both achievement of certain age/grade in one or more subjects is compared across education systems and effects of contextual factors at the system, school, classroom and student level on achievement are studied” (p. 2). Specifically, the IEA (TIMSS, PIRLS, and ICCS) and OECD (PISA) study programs under review are very similar in that they both use comparable psychometric methods to analyze, validate, and scale student response data (von Davier et al., 2013; IEA, 2013) and make use of two-stage clustered sampling designs (OECD, 2009), first with random school samples weighted on school size, then either (a) a random sample of one or two intact classes (IEA studies), or (b) a random sample of the school’s 15-year-olds (OECD studies).

Table 1 provides a summary of the similarities and differences associated with the five ILSA programs of interest in the study. As illustrated, all five studies employ cyclical designs with a focus on measuring trends, though target populations are only comparable for TIMSS Mathematics and Science and ICCS assessment programs (Grade 8).

The programs differ in terms of general design doctrine. PISA can be conceived as being based on Human Capital Theory (see Cardoso, 2020; and, Sellar & Lingard, 2013 for a comprehensive review), whereas TIMSS uses a curricula model (IEA, 2019a) as the major organizing principle where participating countries’ intended, implemented, and attained curriculum inform the general design of the survey program. PIRLS uses the “attainment of societal goals” encompassing literacy experience and the processes of comprehension to attain societal goals (Mullis & Martin, 2015), whereas the IEA’s ICCS can be conceived as evolved from ecological systems theory (e.g., Neal & Neal, 2013) and is based on cognitive and affective-behavioral dispositions toward civics and citizenship (IEA, 2016). Finally, IEA’s ICILS is underpinned on modern conceptions of computer and information literacy and computational thinking (IEA, 2019b).

In terms of focal subject areas, it is sometimes argued that PISA Science, Reading, and Mathematics literacy focus more on applications to real-world situations compared to TIMSS and PIRLS assessments, which focus more on mastered factual and procedural knowledge taught in school (ACER, 2020), though some studies have found PISA Maths and TIMSS numeracy to be highly comparable forms of assessments (see, for example, Wu, 2009). One particular point of difference for the ICCS study is that it benchmarks student outcomes to the UNESCO sustainable development goals related to citizenship. In addition, the more recent number of participating countries also differ with 79 countries participating in the 2018 PISA, 61 in TIMSS and PIRLS, and 24 in the ICCS surveys. Important to note, studies under all programs cyclically provide international league tables that tend to draw a large amount of media attention and policy debate, yet, as Hopmann et al. (2007) note, there has been a tendency to treat the results of the examinations uncritically and make direct links to a nation’s relative school quality and economic future.

Criticisms of ILSAs

Compared to the level of attention paid to the results of international league tables in public and governmental domains, there exists a dearth of criticism of such programs.

Table 1 A comparison of PISA, TIMSS & PIRLS, ICCS, and CLIL ILSA Programs

Descriptor	PISA (OECD)	TIMSS & PIRLS (IEA)	ICCS (IEA)	ICILS (IEA)
Administration periods	2000, 2003, 2006, 2009, 2012, 2015, 2018...	1995, 1999, 2003, 2007, 2011, 2015, 2019... (TIMSS) 2001, 2006, 2011, 2016... (PIRLS)	2009, 2016, 2022...	2013, 2018...
Target Populations	15-year-olds (aged between 15 years and three months, and 16 years and two months at time of assessment)	Grade 4 (though next grade higher when average age of testing is less than 9.5; TIMSS & PIRLS); Grade 8 (though next grade higher when average age of testing is less than 13.5; TIMSS) ²	Grade 8 (though next grade higher when average age of testing is less than 13.5) ³	Grade 8 (though next grade higher when average age of testing is less than 13.5) ⁷
General Design Doctrine	Human Capital Theory (Sellar & Lingard, 2013)	TIMSS: National Curricula Focussed (IEA, 2019a) PIRLS: "Reading to do" (Stiggins, 1982) and the attainment of societal goals (Mullis & Martin, 2015)	Ecological Systems Theory (Neal & Neal, 2013)	Computer and information literacy and computational thinking (IEA, 2019b)
Focal subject areas	Scientific, Reading, and Mathematical Literacy every cycle as focal subjects (with collaborative problem solving, problem solving, and financial literacy included intermittently)	Mathematics and Science (TIMSS); Reading (PIRLS)	Trends in civic knowledge and engagement ³	Computer and Information Literacy ⁷
Focus of assessment	How well students can apply understanding and skills in science, reading, and mathematics to real-world situations ¹	How well students have mastered factual and procedural knowledge taught in school mathematics and science curricula (TIMSS) ¹ Reading comprehension at Grade 4 (PIRLS)	UNESCO Sustainable Development Goal (SDG), specifically, 4.7.4: "the percentage of students... showing adequate understanding of issues relating to citizenship and sustainability" ⁴	Knowledge of basic software and basic use of computers as information source Sufficient knowledge and skills of ICT for information gathering and use Critical thinking while searching for information online ⁷
Most recent sample design	Minimum 5250 students from 150 schools ⁶	Minimum of 4,000 students from 150 schools ²	Minimum of 3,000 students from 150 schools ⁵	3000 8th Grade students from each country
Most recent number of participating territories/economies	79	61 (50 countries; and 11 benchmarking entities including different regions, and age/language groups)	24	13

¹ Information adapted from Australian Council for Educational Research (ACER, 2020); ² adapted from IEA (2019a, 2019b); ³ adapted from IEA (2020); ⁴ adapted from UNESCO (2020); ⁵ IEA (2018); ⁶ OECD (2019); ⁷ most recent administrative period in bold; ⁷ IEA (2014)

Though, academic criticism have centered around the idea that (a) ILSAs promote regional and global isomorphism, (b) results from ILSAs are often reported in a poor uncriticized way, (c) causal language is often misused when presenting ILSA results, and (d) ILSAs are exclusionary and lack consequential and other components of validity.

A common criticism levelled against ILSAs is that they promote isomorphic ideologies (see Wiseman et al., 2014) across jurisdictions and cultures. Here it is argued that similarities in educational structures, policies, pedagogical approaches, and curricula content emerge as a consequence of shared insights from ILSAs (Pettersson, 2008). While theorists have argued for regional (Dale, 2000) and global (Meyer et al., 1992) educational homogenization, it is difficult to identify the degree to which ILSAs themselves drive isomorphism (Johansson, 2016) or whether commonalities across curricula happen to exist.

Torney-Purta and Amedeo (2013) point to the common lack of criticism of country rankings in league tables often reported in the media given the inherent inaccuracies associated with measurement. Specifically, the authors point to the imprecision associated with estimates of country-level means in ILSAs. A calculation of sampling errors alone¹ for the more recent 2018 PISA Science scores helps to illustrate this point: while results afford Australia, the U.S., and New Zealand mean scale scores of 502, 503, and 508, we can only be 95% confident that the countries' true means fall between 499.15 and 504.85, 501.31 and 504.69, and 505.40 and 510.60, respectively (see Additional file 1: Technical Appendix, Note 1). Therefore, it is only reasonable to assert that the Scientific literacy of 15-year-olds in New Zealand likely exceeds that of counterparts in both Australia and the U.S. In addition, ILSA programs commonly report shifts in country-level means (or growth trends) cycle-to-cycle. However, such estimates *also* involve item equating (or item link) methodologies that align the results across different assessment administrations (Wu, 2010). While this contributes to an even larger amount of error in estimating cyclical shifts in country-level means (Michaelides & Haertel, 2004), associated inaccuracies (equating errors) are commonly not reported (Wu, 2010). Other measurement critics have argued that PISA Reading data has reflected at least two underlying dimensions and that alternate scaling models are more appropriate (Goldstein et al., 2007).

Another criticism of ILSA is, that, due to the cross-sectional survey designs of such programs, it is not easy to draw causal inferences about the data (Rutkowski & Delandshere, 2016). To this end, Gustafsson (2008), argues that one key negative consequence of ILSAs is the causal language used to explain findings. Further, Allardt (1990) argues that there is little to no evidence that comparative surveys in any field of the social sciences have been able to create knowledge about causal phenomena. Relatedly, Goldstein (2004) points to ILSA's general lack of longitudinal data (on the same sample of students) that would enable researchers to attribute differences in groups to educational systems per se. Despite these criticisms, it should be noted that recent studies have made an attempt to use methodologies that enable causal inferences from ILSA data (see for example, Hogrebe & Strietholt, 2016).

¹ Not accounting for inaccuracies associated with measuring individual students, though these become negligible at the country mean level; see Wu (2010, Supplementary material, Technical Appendix, Note 1).

Some academics have argued that ILSA is exclusionary and lacks consequential validity. ILSA relies on sophisticated statistical techniques such as item-response theory (IRT), computer adaptive testing (multi-stage CAT), and matrix sampling designs (with different groups responding to different sets of items) to enable efficient and unbiased estimation of system-level performance. However, Gustafsson (2008) argues that this complexity also makes insights and analyses of these data exclusionary for some countries, only available to a small number of experts. Therefore, insofar as the interpretation and *use* of measurement devices form an important part of validity (see Messick's 1989 reconceptualization), it could be argued that ILSAs lack this specific component.

Despite criticisms, the ILSAs carry with them many advantages. As mentioned, ILSAs provide a data infrastructure for research into a variety of issues and avoid the ethical problems and costs associated with randomized experiments. The strong methodological foundations enable secondary analysts to make use of the openly-available data to provide stable and generalizable descriptions of knowledge trends within and between schools, and between-countries, and across administrative cycles.

Summary of reviews on ILSA research in education

As research on ILSAs continues to grow, researchers have begun to provide summaries and reviews to provide practitioners, researchers, and policymakers a general picture of this developing field. This has generally involved narrative and critical reviews of the literature. These kinds of reviews are particularly useful to offer relevant insights on the themes, methodologies, and theoretical underpinnings of the research on a particular field or topic. Examples of such studies include an examination of the scientific contribution and impact of TIMSS and the influence of school and classroom contexts on student academic achievement (Drent et al., 2013), a critical review on PISA effects on education governance (Pons, 2017), and methodological reviews of research strategies and statistical techniques used in ILSA research in education (e.g., Lin et al., 2014; Liou & Hung, 2015). Meta-analytic approaches have also been used to review research in the field. Such procedures are appropriate to systematically assess the results of previous research to derive conclusions about that body of research. An example is the meta-analysis by Else-Quest et al., (2010) which used PISA 2013 and TIMSS 2013 data to explore cross-national gender differences in mathematics achievement.

A third path to synthesize the research literature in a field is a bibliometric approach. Bibliometrics can be generally defined as “a set of quantitative methods used to measure, track, and analyze print-based scholarly literature” (Roemer & Borchardt, 2015, p. 234). Bibliometrics has been proposed as a valuable approach to map vast amounts of research available in particular disciplines and to describe their developmental trends and current status in a comprehensive, systematic, and replicable manner (Linnenluecke et al., 2020). Such syntheses are important because they envision the past, present, and future developments of a field. A few bibliometric reviews of the literature on ILSA research in education are available. In 2012, Domínguez et al. (2012) provided an early descriptive bibliometric study on the impact of the PISA studies on academic journals from 2002 to 2010. The study, analyzing 322 recognized papers, identified prominent researchers, journals, and topics of the publication period. The authors identified the prominence of Western countries and authors, with a focus on student performance, and an emergence of ICT use and equality topic areas.

Lindbald et al., (2015) conducted a systematic review of 337 journal articles on ILSA (PISA, TIMSS, and ICCS) research that focuses on comparative achievement between schools. The authors identified prominent journal articles, journals, author origins, and associated fields of education. The authors found a Western country bias, fields focused on educational comparative- and policy-related research, and complementary fields of economics, psychology, and sociology, and concluded that the field is conducted and managed by few and thus potentially immature and vulnerable.

Supplementary to the aforementioned study, Lenkeit et al., (2015) undertook a bibliometric specific to PIRLS-related journal publications spanning the 2001, 2006, and 2011 cycles. The authors identified 114 articles covering secondary analysis (63.2%), critique (28.1%), and policy categories (8.8%), noting that the prominently Western countries focused on multi-country comparisons.

More recently, Hopfenbeck et al. (2018) categorized and analyzed 654 journal articles that covered PISA-related topics. The authors noted a steady rise in the number of publications from just several in the early 2000s to 103 in 2014, and also noted the Western-centric outputs covering a similar suite of fields to that identified in the Lindbald et al., (2015) study. The authors also identified that the journals covered three main research categories of secondary analysis (61.8%; with focus on demographic inequalities), critique (16.2%, with focus on construct validity, research design, and various technical issues), and policy categories (22.0%; focusing on the effects of PISA on policy and governance).

The present study

The purpose of this study is to contribute to the growing literature on ILSA research in education by providing an overall picture of the recent development and structure of research in the field using a bibliometric approach. More specifically, we aim to map modern ILSA research in education by describing the developmental trajectory of the field based on publication and citation data related to the five major recent and ongoing PISA programs (i.e., PISA, TIMSS, PIRLS, ICCS, and ICILS), the core journals and most influential publications in the field, the leading scholars and countries and the patterns of scientific collaboration between them, disciplines and historical developmental paths underlying the foundations of ILSA research in education, and the major research topics addressed in the literature.

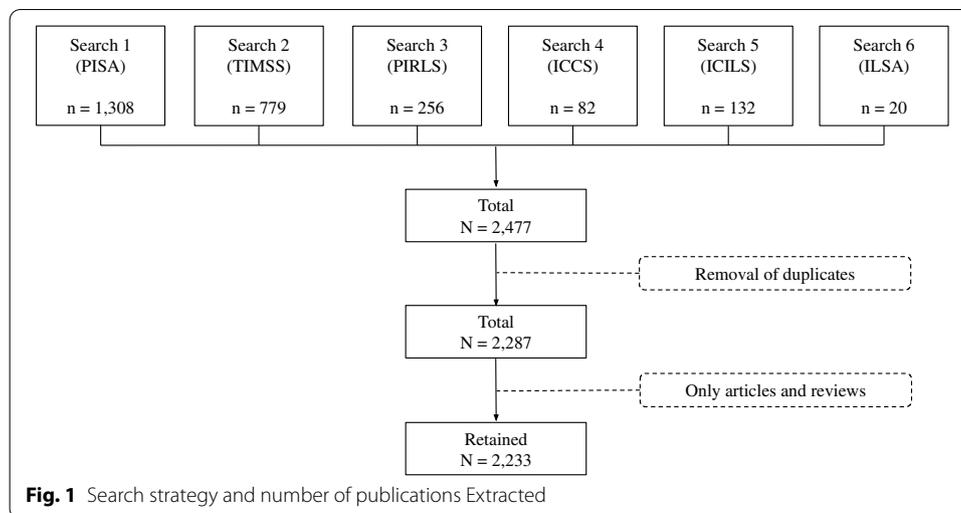
Our study complements and differs from previous reviews of the literature in several ways. First, our study provides a wider coverage of the extant literature by considering publications addressing five major ILSA programs in education, in contrast to previous studies that have focused on specific assessments (e.g., Drent et al., 2013; Hopfenbeck et al., 2018; Lenkeit et al., 2015; Pons, 2017). Early ILSA studies (e.g., First International Science Study, FISS; First International Mathematics Study, FIMS; Six Survey Study) were not considered in this study as they have been discontinued decades ago and do not provide relevant insights to current education systems and policies. Second, previous reviews have predominantly examined documents published in academic journals in which the language of publication is English. In our study, we do not exclude documents published in other languages to account for a broader geographical representation of ILSA research in education across the world. Third,

although other reviews have examined the development and the major journals, authors, and countries in ILSA research in education, no study to date has explored how researchers and countries collaborate in the production of scientific knowledge in the field (i.e., social structure). Similarly, this is the first study that explores in a systematic way what disciplines underlie the foundations of the field and the historical paths that contributed to its development (i.e., intellectual structure). Fourth, we provide a detailed analysis of the major research topics addressed in the literature using a data-driven approach that provides a unique and objective perspective to account for the research topics that have received the greatest attention in recent decades, instead of using content analysis of publication content as in previous studies (i.e., focussing on knowledge structure) (e.g., Domínguez et al., 2012; Hopfenbeck et al., 2018; Lenkeit et al., 2015).

Data and methods

A bibliometric approach was used to build a corpus of publications on modern ILSA research in education using metadata extracted from four indexes of the Web of Science (WoS) database: The Science Citation Index-Expanded (SCI-Expanded); the Social Sciences Citation Index (SSCI); the Arts & Humanities Citation Index (A&HCI); and the Emerging Sources Citation Index (ESCI). WoS database was used because it is the most widely used database for harvesting research metadata (Meho & Yang, 2007) and is considered the industry standard in most disciplines, including education (Ivanović & Ho, 2019). Also, WoS is multidisciplinary and thus allowed for the compilation of publications on ILSA research emerging from multiple disciplines (McVeigh, 2009). Moreover, WoS provides a wider coverage of citation information compared to other multidisciplinary databases such as Scopus and Google Scholar (Li et al., 2010) and better accuracy in journal classification (Wang & Waltman, 2016).

Six complementary searches were conducted in the four WoS indexes on April 22, 2021. The first five searches were aimed to retrieve publications related to the five ILSA programs under investigation in this study and included the following key terms: (1) ["PISA 20*" or "Programme for International Student Assessment"], (2) ["TIMSS" or "Trends In International Mathematics And Science Study"], (3) ["PIRLS" or "Progress In International Reading Literacy Study"], (4) ["ICCS 20*" or "International Civic and Citizenship Education Study"], and (5) ["ICILS" or "International Computer and Information Literacy Study"]. The sixth search was intended to extract more general publications in ILSA research in education and included the key terms ["ILSA" or "International Large Scale Assessment"] and ["Education"]. All search terms used were imputed in the *Topic* field. Figure 1 presents the number of documents retrieved from this search strategy, which overall yielded a total of 2,477 publications, which were subsequently reduced to 2,287 after the removal of the duplicates. The documents were then filtered by type of document and only articles and reviews were extracted because other document types do not contain complete metadata information. No filter was applied for the language of publication or year of publication. This resulted in a total of 2,233 publications. For each publication, the following metadata was extracted: publication title, abstract, and keywords; publication year, journal, and

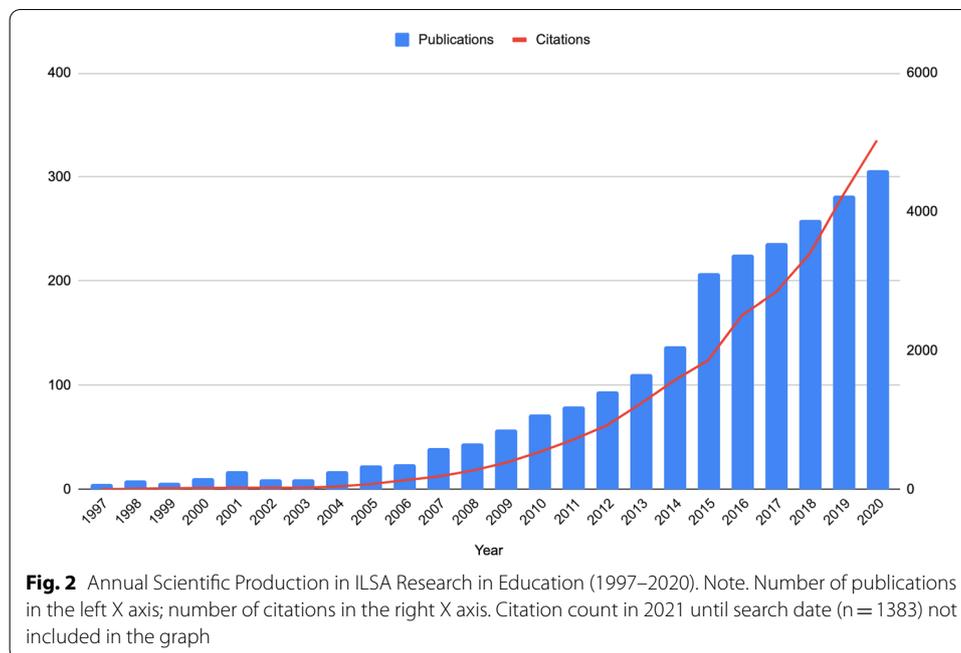


number of citations, and references; and author's names and country of affiliation (i.e., the territory where the institution with which the author is affiliated is located).

Data analysis procedures

The data analysis comprised three steps. First, descriptive bibliometric analyses in Bibliometrix version 3.0 were performed to provide an overview of the development of ILSA research in education and the key players in the field (RQ1, RQ2). Bibliometrix is an open-source tool developed in the open-source R programming language that calculates multiple bibliometric and scientometric indicators for science mapping (Aria & Cucurullo, 2017). More specifically, descriptive analyses using the *biblioAnalysis* and *plot* functions were performed to illustrate the dynamic growth of publication and citation data, the core journals and most influential publications, the leading authors and countries, and the most frequently used keywords in ILSA research. Frequency counts of the total number of publications for a given year and the total citations of the articles published each year provided an account of the evolution of the interest in ILSA research in education through time. The Standard Competition Ranking (SCR) was used to rank the productivity of journals, authors, and countries based on the total number of publications. The most influential articles were ranked based on the total number of citations.

Second, multiple social network analyses were conducted in VOSViewer to explore the structure of research on ILSA in education (RQ3, RQ4, RQ5). VosViewer is a freely available software for viewing and constructing bibliometric maps (Van Eck & Waltman, 2010; Waltman et al., 2010). In VOSViewer, the units of analysis are journals, publications, citations, authors, countries, or keywords, depending on the focus of the analysis. To create the bibliometric maps, VOSViewer normalizes the differences between the units of analysis and builds a two-dimensional map where these units are represented as circular nodes. The size of the nodes accounts for their volume (e.g., number of publications in the dataset by an author) and the distance between the nodes reflects the similarity between these nodes. Nodes are connected with lines or edges, which indicate a relationship between nodes. Edge thickness indicates the strength of that relationship. Finally, VOSViewer groups closely related nodes into clusters, where each



color represents a cluster (Van Eck & Waltman, 2014). Co-authorship analyses were used to explore the networks of scientific collaboration between authors and countries (RQ3) (Newman, 2001, 2004). A co-citation analysis of the most cited journals in the dataset was performed to explore the disciplines underlying the structure of the field (RQ4) (Ding et al., 2000). A co-occurrence analysis of the most frequent keywords in the dataset was conducted to explore the knowledge base of research on ILSA in education (RQ5) (Rijsbergen, 1977). Additional information is presented in the Results section to facilitate the interpretation of the co-authorship, co-citation, and co-occurrence bibliographic maps.

Third, additional social network analyses were conducted to further elaborate on the historical evolution of the field. A historiographic analysis in Bibliometrix using the *hisNetwork* and *histPlot* functions was performed to uncover the historical citation paths of the most cited documents in the database (Garfield, 2004). Finally, three keyword co-occurrence maps were generated in VOSViewer to describe the thematic evolution of ILSA research in education across three "time-slicing" periods based on the overall time distribution of keywords in the dataset. Additional details are provided in the appropriate Results section to facilitate the interpretation of the analyses about the historical evolution of ILSA research in education.

Results

Growth trajectory

The developmental patterns of a field can be illustrated by examining the trends in publication and citation data. The 2,233 publications in ILSA in education research have been cited a total of 27,406 times. Figure 2 shows the dynamic growth of publications and citations in the field. Overall, the trends suggest a steady increase in the scholarly interest of ILSA in education research from 1997 to 2020 that can be organized into three stages:

Table 2 Leading Authors in ILSA in Education Research

Rank	Author	Publications	Citations
1	Jerrim J	18	217
2	Marsh HW	15	587
3	Rindermann H	15	566
4	Luedtke O	14	250
5	Goldhammer F	13	137
5	Liou PY	13	105
7	Rutkowski L	12	266
7	Saelzer C	12	41
7	Scherer R	12	337
10	Borgonovi F	11	81
10	Han SW	11	59
10	Lee J	11	305

an emergence stage, in which publications rose slowly (1997–2006); a fermentation stage, with the number of publications growing notably (2007–2014); and a take-off stage, in which more than 200 records have been published every year (2015–2020). The number of citations follows a similar pattern, increasing exponentially as we approach the present.

Authors

All the 2,233 publications in the dataset have been published by a total of 3,508 researchers. The majority of authors in the dataset produce multi-authored documents (88%) and more than three-fourths of the publications are co-authored papers (77%). Table 2 shows the leading researchers in ILSA research, ranked by the number of publications. J. Jerrim (University of London, UK), H. W. Marsh (Australian Catholic University, Australia), and H. Rindermann (Technische Universität Chemnitz, Germany) appear as the most productive authors in the field, all with 15+ publications each in the dataset. H. W. Marsh and H. Rindermann obtained the highest number of citations.

The results of the co-authorship analysis of authors with five or more publications in the dataset ($n=122$) are presented in Fig. 3. In this analysis, each node represents an author and its size reflects the number of publications in the dataset for each author. The edges connecting the nodes account for co-authorship relationships (i.e., co-authored publications), and the clusters can be interpreted as networks of scientific collaboration between authors (i.e., research groups). The map suggests the existence of numerous networks of scientific collaboration, with eight large research groups at the core and all smaller groups appearing in the periphery. These eight groups present some collaborative ties between themselves and are led by productive researchers in the field. The group led by Goldhammer-Luedtke (red cluster) seems to be situated at the center of the collaborative network and maintains links with most of the other connected clusters. A considerably large cluster not connected to these central groups is the one composed by Borgonovi, Santin, and other authors (yellow).

Countries/territories

A ranking of the 50 most productive corresponding countries/territories publishing research in the dataset is presented in Table 3. The USA appears as the leading country

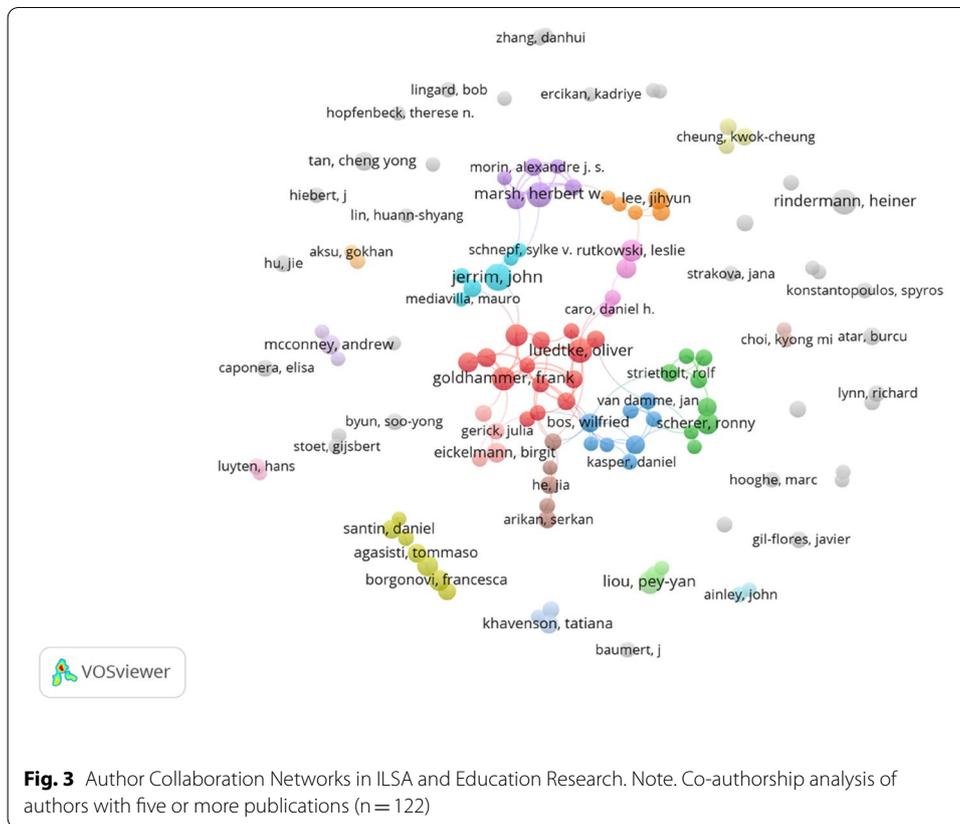
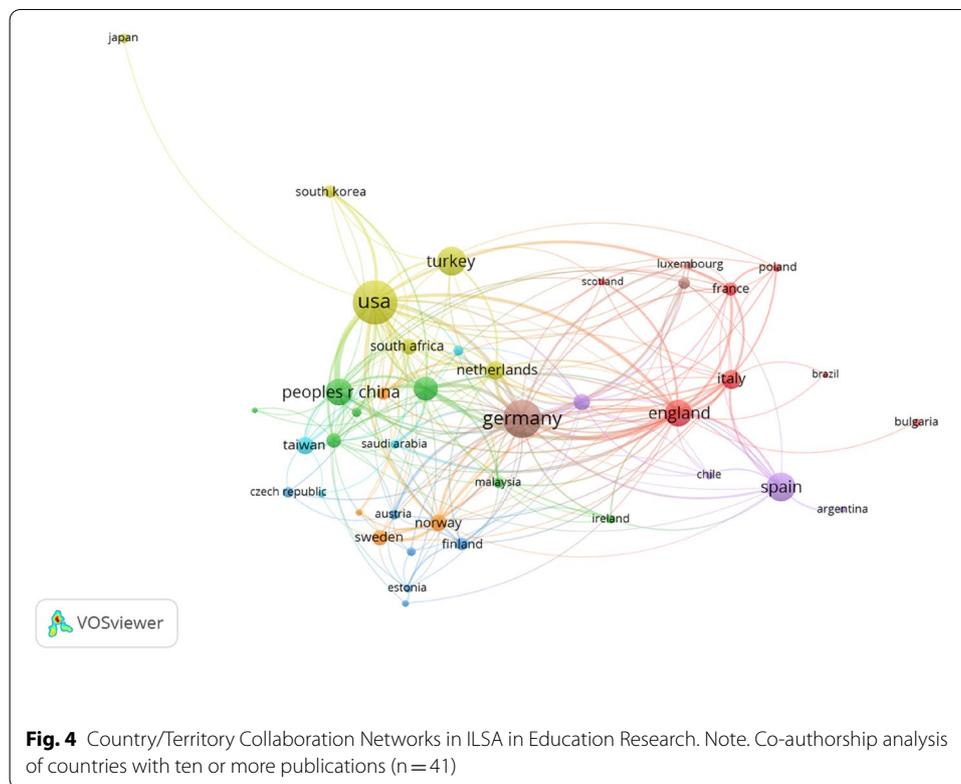


Table 3 Leading Corresponding Countries/Territories in ILSA and Education Research

Rank	Country/Territory	Articles	Citations	SCP	MCP	MCP_Ratio
1	USA	343	6,951	269	74	0.22
2	Germany	273	4,126	207	66	0.24
3	China	199	1,469	142	57	0.29
4	Turkey	184	853	175	9	0.05
5	Spain	169	1,126	154	15	0.09
6	United Kingdom	116	2,221	70	46	0.40
7	Australia	102	2,859	70	32	0.31
8	Italy	73	569	55	18	0.25
9	Netherlands	56	878	39	17	0.30
10	Sweden	46	423	34	12	0.26
11	Belgium	43	454	29	14	0.33
12	Norway	42	567	29	13	0.31
13	South Africa	41	176	35	6	0.15
14	Canada	37	450	27	10	0.27
15	Finland	31	536	25	6	0.19

TC Total citations; SCP Single country publication, MCP Multiple country publication

both in terms of number of publications and total citations, with Germany, China, Turkey, Spain, the United Kingdom and Australia, all with more than 100 publications each, following in the rankings. It should be noted that while China has been relatively



productive, Chinese students themselves are not proportionally surveyed. For example, in PISA, only students in Shanghai are sampled, while in TIMSS and ICCS, only students from regionally-associated jurisdictions, Taiwan and Hong Kong, are sampled. Nevertheless, the USA, Germany, Australia, and the United Kingdom are the countries with the highest number of citations. The United Kingdom emerges also as the country with the highest collaboration intensity in the ranking (i.e., rate of publications with at least one co-author from a different country), while Turkey and Spain demonstrate substantively lower levels of international collaboration.

Figure 4 displays the networks of collaboration at the county/territory level. Only countries/territories with ten or more publications were included in the analysis ($n = 41$). Overall, modern research on ILSA in education seems to be generated by the collaboration of geographically proximal countries/territories. Germany appears at the center of the map and demonstrates co-authorship ties with other productive countries in the dataset, especially in Western (red cluster) and Northern Europe (blue, orange). The USA, the most productive country in ILSA education research, collaborates closely with Turkey, South Africa, and East Asia (Japan, South Korea). Australasian countries tend to cluster together (green). Finally, an international collaborative network between predominantly Spanish speaking countries is represented by the purple cluster.

Journals

The 2,233 publications in the dataset have been published in 616 journals. Table 4 shows the top core journals in ILSA research ranked by number of publications. Most of the

Table 4 Core Journals in ILSA in Education Research

Rank	Journals	Articles	Citations
1	International Journal of Science Education	52	1124
1	Large-Scale Assessments in Education	52	275
3	International Journal of Science and Mathematics Education	41	604
4	Egitim Ve Bilim-Education and Science	37	245
5	Zeitschrift Fur Erziehungswissenschaft	36	155
6	Revista de Educacion	33	197
7	Learning and Individual Differences	30	656
8	Frontiers in Psychology	29	129
8	School Effectiveness and School Improvement	29	439
10	Computers & Education	27	743
11	Journal of Measurement and Evaluation in Education and Psychology	26	19
12	Compare-A Journal of Comparative and International Education	25	140
13	Educational Psychology	24	196
13	International Journal of Educational Development	24	232
13	Comparative Education	23	902
13	Economics of Education Review	23	702
13	International Journal of Educational Research	23	93

core journals in the field seem to be indexed in the Education and Educational Sciences WoS category. Only one journal in the Social Sciences category (*Economics of Education Review*), and two journals in the Psychology category (*Educational Psychology*, *Frontiers in Psychology*) appear in the ranking of core journals. The journals accumulating the highest number of publications in the dataset are the *International Journal of Science Education* and the journal *Large-scale Assessment in Education* (a journal rising to prominence largely in the “take off” stage, see subsection on growth trajectories). With regard to the total number of citations, the *International Journal of Science Education* and *Comparative Education* were the most cited sources. Figure 5 presents the yearly growth in the number of publications in the five core journals in the field and reveals that the journal *Large-scale Assessment in Education* has become the core journal in the field.

Figure 6 presents the co-citation analysis of journals in the dataset with more than 50 citations ($n=283$). In this analysis, the nodes denote journals in the dataset and their size is a reflection of the number of co-citations with other journals in the dataset. Two journals are co-cited if a third journal in the dataset cites a publication in these journals. Frequently co-cited journals are assumed to share theoretical and semantic grounds, and therefore clusters can be interpreted as disciplines from which modern ILSA research in education emerges. The analysis suggests that there are five distinctive but related disciplines contributing to the development of modern ILSA research in education. The green cluster groups journals in the fields of *Psychology and Behavioral Sciences* (e.g., Educational, Social and Personality, Counseling, Developmental psychology). The blue cluster pulls together several journals publishing research on multiple topics in the *Educational Sciences* (e.g., science education, school leadership/management, teacher education, school effectiveness, policy education). Journals in *STEM Education* (science

Table 5 Most Influential Publications on ILSA Research in Education

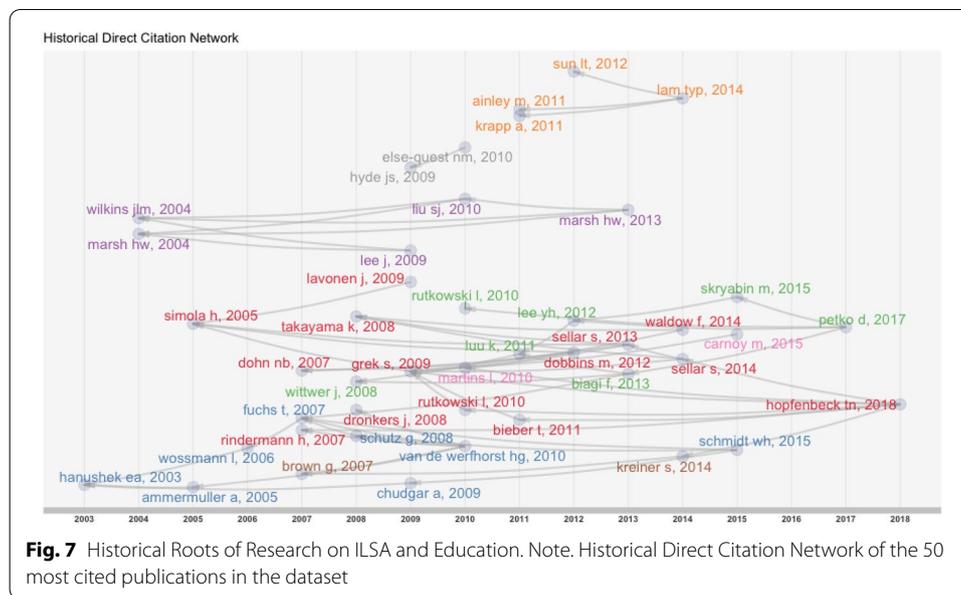
Rank	Publication	Title	TC	TC/Year
1	Else-Quest et al., (2010), <i>Psych Bullet</i>	Cross-national patterns of gender differences in mathematics: A meta-analysis	592	49.3
2	Grek, (2009), <i>J. Educ Policy</i>	Governing by numbers: The PISA 'effect' in Europe	470	36.1
3	Van De Werfhorst et al., (2010), <i>Annu Rev Sociol</i>	Achievement inequality and the institutional structure of educational systems: A comparative perspective	255	21.2
4	Krapp et al., (2011), <i>Int J Sci Educ</i>	Research on interest in science: Theories, methods, and findings	253	23.0
5	Simola, (2005), <i>Comp Educ</i>	The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education	214	12.6
6	Ainley et al., (2011), <i>Contemp Educ Psychol</i>	Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science	197	17.9
7	Marsh et al., (2004), <i>J Educ Psychol</i>	Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries	195	10.8
8	Rindermann, 2007, <i>Eur J Personality</i>	The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations	190	12.7
9	Hyde et al., (2009), <i>P Natl Acad Scien</i>	Gender, culture, and mathematics performance	188	14.5
10	Lee, (2009), <i>P Natl Acad Sci USA</i>	Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participant countries	184	14.1

TC Total citations

Publications

Table 5 presents the 10 most influential publications in ILSA research, ranked by the number of total citations. The most cited publication in the dataset is the article by Else-Quest et al. (2010), a meta-analysis exploring gender gaps in mathematics achievement using TIMSS 2003 and PISA 2003 data. Although the most influential publications address a variety of research topics, an overarching issue across the publications is the study of cross-country and cross-national patterns in variables measured by ILSA assessments. Other recurrent topics in these publications are equity in education, academic achievement, student interest, engagement and self-concept, and the globalization of education. The proportion of theoretical-empirical publications in the list is approximately 50–50, but there is a clear predominance of documents examining or discussing issues around the PISA test.

To elucidate the historical roots of the modern research on ILSA in education, a historical direct citation network analysis of the 50 most cited publications was performed (see Fig. 7). In this analysis, each node represents a highly cited paper in the field and the direction of the arrows account for the chronological change of research trends in the past. Each path represents the historical evolution of a research theme based on the chronological relations of the most relevant citations in the field, which



permits the understanding of the genealogical antecedents and descendants of ILSA in education research. The historiography suggests the existence of nine historical developmental paths. Based on the findings, the earliest developments correspond to the blue and purple paths at the beginning of the twentieth century. The publication by Hanushek (2003), which explores human capital and quality education around the world using TIMSS data, serves as a seed for the development of broad research on (in)equity in education through ILSA data (blue path). The studies by Wilkins (2004) and Marsh (2004) on academic self-concept initiate research on academic self-concept and motivation, especially in Math and Sciences (purple). Another early development is marked by the red path, which represents the chronological evolution of research on education policy and globalization. This is an intricate path that starts with Simola's (2005) publication examining the social, cultural, and historical factors explaining the success of the Finish comprehensive school system. The next historical paths in the timeline correspond to the development of research on ICT literacy and engagement (green); model fit, robustness, and scaling of ILSAs in education (brown); gender gaps in ILSA research (grey); family socio-economic inequalities (pink); and most recently, science literacy (orange). All publications in the historical citation analysis are provided in Additional file 2 to guide the reader in understanding these developmental trends.

Keywords

Our bibliometric analysis revealed the presence of 4,774 author keywords (AKW) and 2,316 Keyword Plus (KWP) in the dataset. Table 6 shows the 10 most frequently occurring AKW and KWP on ILSA in education research and Fig. 8 presents the yearly growth of occurrence for the period 1997–2020. PISA and TIMSS clearly stand out as the most common AKW, while achievement, performance, and education appear as the most frequently used KWP.

Table 6 Most Frequently Occurring AKW and KWP on ILSA in Education Research

Rank	AKW	Occurrences	KWP	Occurrences
1	Pisa	448	Achievement	352
2	Timss	231	Performance	258
3	Education	112	Education	243
4	Achievement	83	Students	173
5	Mathematics	70	School	165
6	Mathematics achievement	67	Mathematics	155
7	Pirls	66	Motivation	131
8	Assessment	62	Academic-achievement	121
9	Science	49	Science	97
10	Science achievement	47	Impact	96

AKW Author Keywords, KWP Key Word Plus

Figure 9 presents the results of the co-occurrence analysis of all keywords in the dataset occurring five or more times ($n = 649$). In this analysis, nodes account for keywords in the publications and edges connect keywords that frequently appear together (i.e., co-occur) in the publications. Clusters accumulate frequently co-occurring words across publications and can be interpreted as research topics or themes addressed in the literature. The map suggests that the interest of researchers in the field has revolved around six major research topics: (1) Measurement and testing, with frequently occurring keywords such as model, validity, item response theory, tests, and measurement equivalence (red); (2) Educational policy and reform related to policy, governance, impact, accountability, standards, and OECD (green); (3) Education quality/effectiveness around issues of efficiency, competition, choice, and class size (light blue); (4) Equity in education pertaining to segregation, inequality, gender, attainment, and migration (dark blue); (5) Interpersonal relationships connected to school climate, family, socialization, victimization, and bullying (yellow); and (6) Motivation and beliefs (e.g., self-concept, self-efficacy beliefs, attitudes, and interests) (purple). There are three secondary additional topics that seem to have captured the attention of ILSA in education researchers. These are related to students characteristics and include student knowledge and comprehension (brown), attitudes and interests (orange), and engagement and values (pink).

The thematic evolution of the topics addressed in the literature of ILSA in education was explored through three additional co-occurrence analysis of the keywords included in the dataset. A co-occurrence map of frequently occurring keywords was generated for each of the three of the development of the field: emergence (1997–2006), fermentation (2007–2014), and take-off (2015–2020). The three maps are displayed in Fig. 10 using density visualization. These maps are similar to the one presented in Fig. 9 and display the most frequently occurring keywords. Labels represent highly frequently occurring keywords in the dataset; the larger the label, the higher its occurrence. Keywords are placed in the map based on their co-occurrence relationship. However, instead of displaying clusters of frequently co-occurring words as in Fig. 9, the density visualization colors each point of the map based on the density of items in that area. These colors can go from red to blue. High density areas are those with a large number of items in the neighborhood and are represented by reddish colors. Low density areas include fewer items and are colored in blueish. The density view is particularly useful to determine

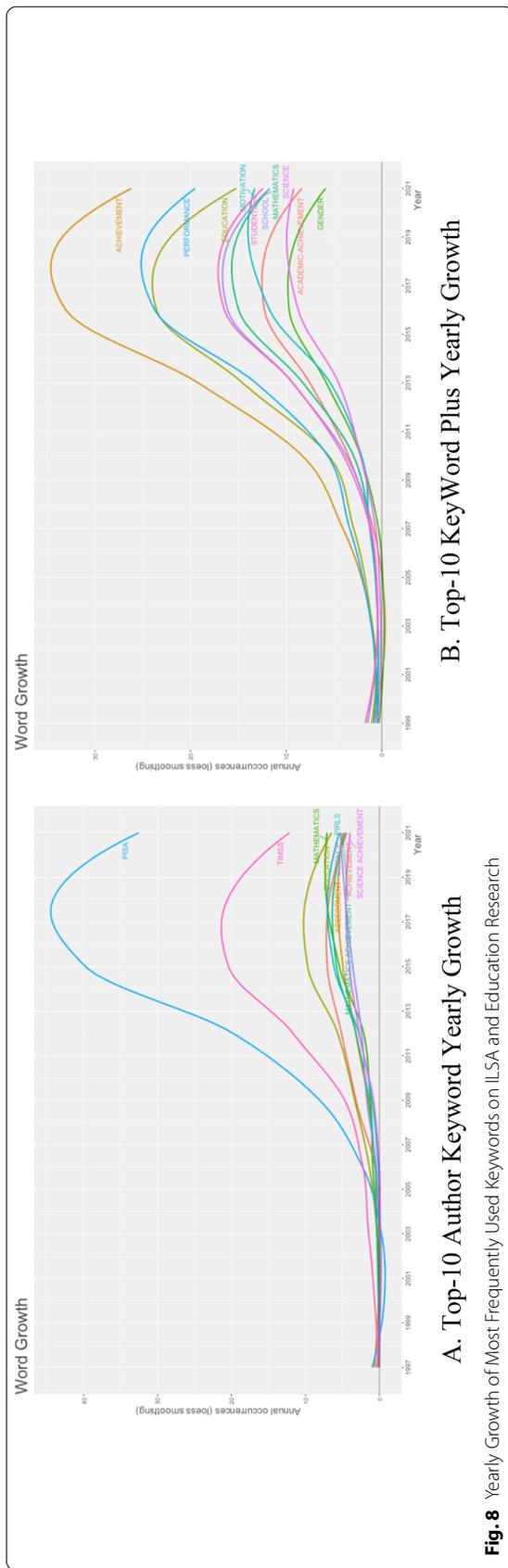
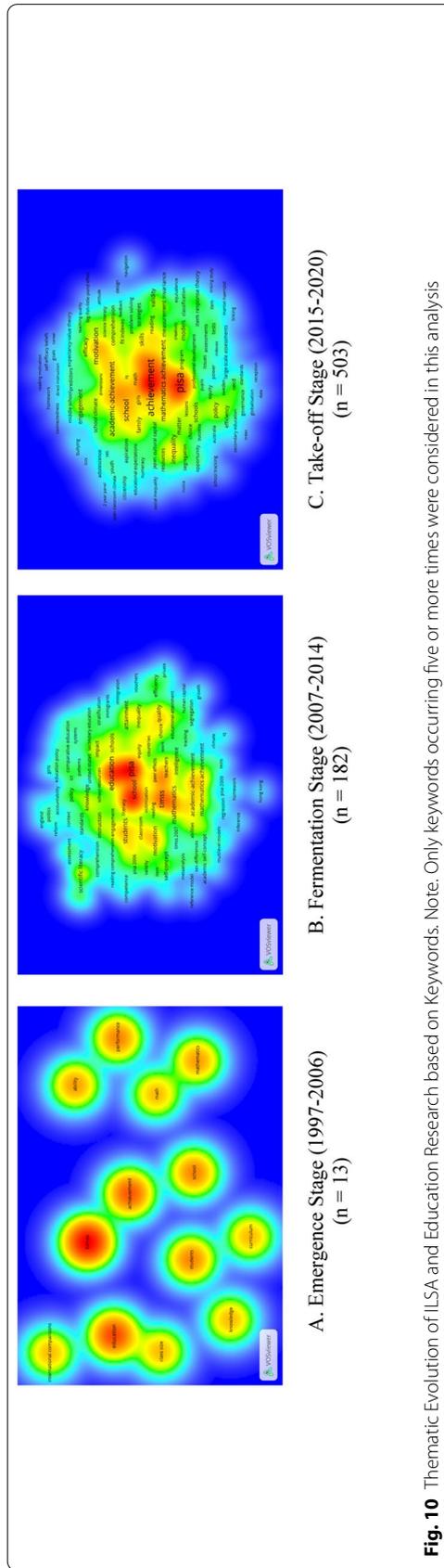


Fig. 8 Yearly Growth of Most Frequently Used Keywords on ILSA and Education Research



policy (e.g., efficiency, governance, globalization, standards) and measurement (e.g., models, validity, measurement invariance, item response theory, tests) also emerge more clearly as important topics during this period.

Discussion

The purpose of this study was to map the growing global literature on ILSA research in education and to provide an overall picture of the recent development and structure of the field in terms of volume and growth trajectory, leading publications, journals, authors and countries, networks of scientific collaboration in the field, disciplines and historical developmental paths underlying the foundations in the field, and the research topics that have received the greatest attention in the literature, as well as their evolution. We used a novel approach relying on bibliometric techniques and a data-driven approach to contribute to the scholarly debate in the field and provide directions for future research.

Volume and growth trajectory of ILSA research in education

The findings of the study suggest that the scholarly interest in modern ILSA research in education has consistently increased from the end of the twentieth century to date, especially after 2014, which is consistent with previous reviews conducted for earlier periods (e.g., Hopfenbeck et al., 2018). Overall, the developmental paths of publication data reflect the patterns of an emerging field. The typical development of a field is organized in several stages. At its birth, a few scholars explore new ideas and theoretical frameworks and produce the first publications (preliminary stage). Later, these ideas permeate to the research community and a larger group of researchers generate a notable growth in the number of publications (exponential growth stage). This is followed by a period in which the number of publications remains stable and ultimately declines when interest in the area is lost or replaced by new ideas (saturation phase) (Dabi et al., 2016; Keathley-Herring et al., 2016). Based on our findings, ILSA research in education seems to be situated at the beginning of the exponential growth stage, or what we have denominated as the take-off stage. Based on the growing citation patterns over time (Fig. 1), we would expect that a new suite of papers from the 2010 to 2015 era that holds more specific relevance (e.g., to student ICT-related behaviors, well-being) to begin to establish prominence in the field over the 2020 to 2025 period.

Several reasons can potentially explain the increasing trends in publication and citation data. First, the knowledge derived from ILSA programs accumulates over time due to their cyclical nature, offering researchers a periodically larger pool of updated data. Second, the number of jurisdictions participating in ILSA programs also increases with each cycle, which makes ILSA data relevant to a continuously wider audience. Third, new assessment domains have been incorporated to ILSA programs in recent cycles (e.g., creative problem solving, student well-being), further expanding their interest to the research community. Fourth, international organizations (e.g., IEA, OECD, World Bank, Soros Foundation, UNESCO, IDB) have been committed to fostering research capacity for analyzing ILSA data over the last two decades, especially in low- and middle-income countries (Ababneh et al., 2016; Wagemaker, 2013). Fifth, the more pronounced growth in publications can be partially explained by taking into account that specialized journals

in the field have been launched in the last 10 years. Finally, concerns about the implications of ILSA programs in education policy-making have also been on the rise, which has likely contributed to the increased productivity in the field.

Leading journals, articles, authors, and countries in the field

The analysis of the core journals in the field suggests that ILSA research in education is, not surprisingly, an educational topic. Most publications in the field are disseminated in educational journals, although research emerging from other disciplines also exists, with some of the core journals specializing in psychology, economics, and measurement and psychometrics. Also to note, most of the leading journals in the field publish research predominantly in English-language, although there are journals disseminating research in other languages, such as German, Turkish, and Spanish, which illustrates the current global interest in the field. Our analysis of journal yearly growth points to the sharp rise in prominence of a subject-specific journal, *Large-Scale Assessments in Education*, and the relative decline of three journals that burgeoned in the 2005 to 2011 period (*Journal of Science and Mathematics Education*, *International Journal of Science Education*, and *Education and Science*). This suggests that the launch of *Large-Scale Assessments in Education* in 2013, as part of a commitment to promote quality ILSA research in education and enhance consequential validity of the field, was timely with the journal well-positioned to play a key role in collating and disseminating ILSA research over the next five to ten years. Specialized journals are indeed capital for the development of a field because they provide an avenue for the dissemination of specialized knowledge, the exchange of ideas, and the formation of a scholarly community of experts in the field (Vanderstraeten et al., 2016). However, the fact that there is only one specialized journal in the field may be limiting the development of the field. Publishers and professional organizations may consider launching competing ILSA-specific journals in light of the growing volume and global interest in modern ILSA research in education revealed in this study.

The analysis of the most influential publications revealed that research in the field tends to focus on cross-national differences in student outcomes measured in the ILSA programs. This is to be expected, as ILSA provides a valuable tool for improving the quality of education systems by learning from the experiences of initiatives and practices implemented in diverse contexts. An interesting insight derived from our analysis, however, is the particular prominence of the gender- and policy-centric publications from Else-Quest et al. (2010) and Grek (2009), respectively, establishing that these research themes are quite central to ILSA research in education. In addition, most of the influential publications seemed to revolve around the analysis and discussion of the results from the OECD PISA program, which indicates that the PISA test captures the greatest attention in the field at the moment.

Research in the field is generated by scholars affiliated to different countries participating in ILSA programs, especially in the US and Germany. This is despite the apparently low interest of US academics on ILSA research (Green Saraisky, 2015). In terms of leading authors, our analysis suggests that the scholarly community of ILSA research in education is composed by a relatively large group of experts that produce a similar number of articles. This suggests that a depth of expertise, a constituent generally necessary

for a knowledge society (Grundmann, 2016), exists in ILSA research in education helping to ensure the ongoing sustainability of the field. Our wide coverage of the international literature demonstrated that while predominantly English-speaking are core to ILSA research in education, non-English speaking jurisdictions appear to be emerging as contributors to the field. Here, China, Spain, and Turkey stand out as major contributors of research in ILSA in education, now outperforming the United Kingdom and Australia. Interestingly, the productivity of these three countries seems to rely on different publication patterns. China seems to maintain high productivity by collaborating with researchers from overseas, whereas Spain and Turkey seem to produce research in the field predominantly within national borders and are likely to disseminate it in local languages via publications in national journals (e.g., *Egitim Ve Bilim-education and Science*, *Revista de Educación*). Still, our study shows that there is a relative scarcity of research produced in lower-middle-income countries (LMIC) that participate in these ILSA programs. Thus, the result of this study suggest that there might not be a Western bias in modern ILSA research in education, as suggested in previous studies (e.g., Domínguez et al., 2012; Hopfenbeck et al., 2018; Lenkeit et al., 2015; Lindbald et al., 2015), but a predominance of research produced in higher-income countries. Some of the reasons for this may pertain to LMIC's less developed research expertise and relative scholarly isolation, and language-based access issues given the disproportionate number of English-language journals representing the field (Wang et al., 2006; Ynalvez & Shrum, 2011).

Collaboration patterns in the production of ILSA research in education

There is evidence that research collaboration is growing around the world (Henriksen, 2016; Kliegl & Bates, 2010). Research collaboration is generally considered as a key contributor of the development of a discipline and as a metric of excellence and quality (Coccia & Wuang, 2016, Freshwater et al., 2006; Kim, 2006; Rolfe et al., 2004), specially in some global regions (e.g., Europe) (Kwiek, in press). Also, research collaboration has a positive influence on productivity and academic impact (Abramo et al., 2017; Kato & Ando, 2013) and the development of research capacity in developing research environments (Barrett et al., 2011).

Based on the findings of this study, collaboration in ILSA research in education seems to be the norm. This is a positive sign, considering that researchers in the humanities and social sciences tend to collaborate less often compared to those in the physical and life sciences (Kwiek, 2018, Yemeni, 2019). More specifically, the present study identified the existence of multiple networks of scientific collaboration at the author and country levels. Eight interrelated research groups are situated at the core and can be considered as influential collectives moving the field forward. These groups are led by some of the most productive researchers in the field and represent international and national collaborations between universities in different regions. International scientific networks tend to be characterized by co-authored publications by authors affiliated to geographically proximal countries and territories belonging to the same global regions (e.g., Europe, East Asia, Australasia) or authors who share cultural and linguistic ties (e.g., Spanish-speaking countries). This is a typical pattern in the social sciences (Mosbah-Natanson & Gingras, 2014), and such patterns of co-authorship could be monitored in future bibliometric studies so as to track shifts in the circulation and exchange of diverse ideas/

phenomena identified as potentially important to the development of the field (Barrett et al., 2011; Kato & Ando, 2013).

Intellectual roots of ILSA research in education

One of the aims of our study was to identify the intellectual roots of ILSA research in education. In this sense, we revealed two interesting insights into the development and structure of the field. First, we observed that the knowledge in modern ILSA research in education has emerged from the interdisciplinary research conducted in five related disciplines: educational sciences, STEM education, educational psychology, social sciences, and measurement and psychometrics (Fig. 6). This is another hopeful feature of the field, considering that interdisciplinary research is capital to integrating diverse approaches and providing holistic perspectives needed to solve complex problems and inform better decision-making (e.g., Aboelela et al., 2007).

Second, our historical direct citation network suggests that ILSA research in education has developed grounded on nine distinctive historical paths broadly accounting for research on equity, quality education, self-efficacy, STEM education, and education policy. These nine historic ILSA paths point to the breadth of the field and contribution to educational research in general. Earlier historical paths (i.e., most prominent at the early stages of the development of the field) seem to include research in educational quality, policy, and globalization, and (in)equality. The other developmental paths have been consolidated more recently, and account for research on science literacy, family socio-economic inequalities, gender gaps, computer skills and ICT engagement, and ILSA measurement. The densest and most intricate are the ones for research in policy and globalization, quality, and equity in education, suggesting that these focal areas will continue being an important part of future ILSA research.

Research themes/topics and their evolution through time

A major contribution of this study is the identification of the research topics that have captured the greatest attention of researchers in the field and their evolution since its origins until the present. Unlike previous studies that used content analysis techniques, we explored this issue using a data-driven approach. The major findings derived from our analysis are discussed below.

First, results from the analysis of the most frequently occurring keywords used in the literature suggest that work on the OECD “PISA” program supersedes “TIMSS”, “PIRLS”, “ICCS”, and “ICILS”. This is understandable given its more diverse set of subject areas, broader global reach, and more regular triennial administrative cycle.

Second, the results of our co-occurrence analysis for the total period analyzed (1997–2020) elucidated that ILSA research in education has revolved around several major research topics. Some of these topics, such as measurement and testing, equity and inequality, and cognition, have been identified in earlier studies (Domínguez et al., 2012; Hopfenbeck et al., 2018; Lenkeit et al., 2015). However, our search strategy and novel data analysis approach allowed us to identify a few additional topics not previously reported in the literature. This included issues related to educational policy and reform, education quality and effectiveness, interpersonal relationships, student motivation and engagement, and student knowledge, attitudes, and values. Overall, most of

these topics are connected with prominent research themes currently explored by education researchers (see Huang et al., 2020) and indicate that ILSA studies appear to be well-placed to make an ongoing major contribution to educational research in general.

Third, the co-occurrence analysis for each developmental stage in ILSA research in education revealed some interesting trends. For example, research on TIMSS was more popular during the early stages of modern ILSA research in education (when PISA data and reports were not available), but a focus on PISA research has become clear in recent years. Also, some topics seem to have continued to remain popular in the field. These are issues connected to student achievement, performance, and education. The attention paid to other themes has evolved through time, with some progressively increasing in popularity as we approach the present. For instance, research on ability, class size, student knowledge, and curriculum were the most popular during the emergence stage (1997–2005). In the fermentation stage (2007–2014), research on motivation and beliefs, education quality and effectiveness, and equity and equality in education appear as frequently addressed topics. In the most recent period, the take off-stage (2005–2020), these research themes consolidated and expanded to include issues on measurement and testing, education policy and globalization. This evolution partially mirrors the evolution of the research interests in education research (Huang et al., 2020), providing further evidence of the potential of ILSA research to inform international research in education and related disciplines.

Limitations

There are several limitations that should be considered. First, we only retrieved documents indexed in the Web of Science database, so we could have omitted relevant publications disseminated in journals not included there. Second, Web of Sciences, as other interdisciplinary databases (e.g., Scopus), is biased against research in the Humanities and the Social Sciences and published in other languages other than English, which might have excluded some other relevant publications (Mongeon & Paul-Hus, 2016). Third, only journal articles and reviews were examined in this study, omitting other types of publications that have also contributed to the development of the field. The consideration of alternative types of publications not included in this study could alter the ranking of leading authors and institutions, as well as the networks of scientific collaborations. Still, we believe that our search strategy, coupled with novel data analysis procedures, provide a meaningful and insightful contribution of the literature and the broadest account of the development and structure of the field provided to date. Future studies can replicate the findings of this study by using alternative databases (e.g., Scopus, ERIC) and including other types of documents beyond the ones included in this study (e.g., book chapters, conference procedures, grey literature).

Conclusion

The present study provides a broad overview of the development and structure of ILSA research in education based on the five major recent and ongoing ILSA programs in the field. There are several positive features that point to the progressive maturation of the field and its good standing to contribute to educational research in general. ILSA research in education is an emerging field currently situated at its take-off stage and

the number of publications in the field has grown considerably in the last twenty-three years. In addition, ILSA research in education is produced by a solid network of scholars with diverse geographical backgrounds that suggest the sustainability of the field. There is no one country clearly dominating the scholarly discourse on ILSA in education and, while most publications are disseminated in the English language, research is available in other languages as well. Also, the field is relatively interdisciplinary and has developed grounded on several differentiated historical paths. Moreover, scholars in the field have addressed a wide variety of topics ranging from equity and quality education, globalization, education policy, and measurement through to student self-efficacy, motivation, and interpersonal relationships. Still, there are some aspects that, ideally, should continue to evolve in the years to come. For example, the existence of a single specialized journal in the field and the scarcity of research emerging from LMIC might be limiting the development of the field. Future studies should examine if the progressive incorporation of territories in South America, Southeastern Asia, and the MENA region into ISLA programs alter these trends in the years to come. There is indeed much potential for future ILSA work. For example, nine countries (Australia, Argentina, Columbia, Czech Republic, Denmark, Georgia, Malta, Turkey, and Vietnam) engaged in 2018 OECD PISA and TALIS (Teaching and Learning International Survey) cycles. The merging of the PISA data with the TALIS data, rich with additional teacher- and school-level information, enables for a more broad and interesting set of research questions to be explored (Gil-Izquierdo & Cordero, 2018). These links along with new instrumentation in student well-being, financial literacy, and social engagement also allow for more relevant and diverse research hypotheses to be explored.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40536-021-00109-1>.

Additional file 1. Technical appendix.

Additional file 2. Developmental trends of modern research on ILSA in education.

Acknowledgements

Not applicable.

Authors' contributions

DHT contributed to the conception and design of the study, data collection, and performed the bibliometric analysis. DHT and MC wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Funding

The authors have no funding to disclose.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have approved the manuscript for submission. The content of the manuscript has not been published, or submitted for publication elsewhere.

Competing interests

The authors declare that they have no competing interests.

Received: 17 February 2021 Accepted: 12 July 2021

References

- Ababneh, E., Al-Tweissi, A., & Abulibdeh, K. (2016). TIMSS and PISA impact—the case of Jordan. *Research Papers in Education*, 31(5), 542–555. <https://doi.org/10.1080/02671522.2016.1225350>
- Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., Haas, J., & Gebbie, K. M. (2007). Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health Services Research*, 42, 329–346. <https://doi.org/10.1111/j.1475-6773.2006.00621.x>
- Abramo, G., D'Angelo, A. C., & Murgia, G. (2017). The relationship among research productivity, research collaboration, and their determinants. *Journal of Informetrics*, 11(4), 1016–1030. <https://doi.org/10.1016/j.joi.2017.09.007>
- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare*, 47(3), 434–452. <https://doi.org/10.1080/03057925.2017.1301399>
- Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4–12. <https://doi.org/10.1016/j.cedpsych.2010.08.001>
- Allardt, E. (1990). Challenges for comparative social research. *Acta Sociologica*, 33(3), 183–193. <https://doi.org/10.1177/000169939003300302>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Australian Council for Educational Research (ACER). (2020). *PISA and TIMSS: What's the Story?* Retrieved from <https://www.acer.org/files/TIMSSandPISA-background.pdf>
- Barrett, A. M., Crossley, M., & Dachi, H. A. (2011). International collaboration and research capacity building: Learning from the EdQual experience. *Comparative Education*, 47, 25–43. <https://doi.org/10.1080/03050068.2011.541674>
- Bos, K. T. (2002). *Benefits and limitations of large-scale international comparative achievement studies: The case of IAE's TIMSS study*. University of Twente. ISBN: 9036517397
- Cardoso, M. E. (2020). Policy Evidence by Design: International Large-Scale Assessments and Grade Repetition. *Comparative Education Review*, 64, 4. <https://doi.org/10.1086/710777>
- CIES. (2021). *Topic based SIGs*. Comparative & International Education Society. <https://www.cies.us/page/TopicSigs>
- Coccia, M., & Wang, L. (2016). Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences*, 113, 2057–2061. <https://doi.org/10.1073/pnas.1510820113>
- Dabi, Y., Darrigues, L., Katsahian, S., Azoulay, D., De Antonio, M., & Lazzati, A. (2016). Publication trends in bariatric surgery: A bibliometric study. *Obesity Surgery*, 26(11), 2691–2699. <https://doi.org/10.1007/s11695-016-2160-x>
- Dale, R. (2000). Globalization and education: Demonstrating a “common world educational culture” or locating a “globally structured educational agenda”? *Educational Theory*, 50(4), 427–448. <https://doi.org/10.1111/j.1741-5446.2000.00427.x>
- Ding, Y., Chowdhury, G., & Foo, S. (2000). Journal as markers of intellectual space: Journal co-citation analysis of information retrieval area, 1987–1997. *Scientometrics*, 47(1), 55–73. <https://doi.org/10.1023/A:1005665709109>
- Domínguez, M., Vieira, M. J., & Vidal, J. (2012). The impact of the Programme for International Student Assessment on academic journals. *Assessment in Education: Principles, Policy & Practice*, 19(4), 393–409. <https://doi.org/10.1080/0969594X.2012.659175>
- Donnellan, M., Trzesniewski, K., & Lucas, R. (2011). Introduction. In K. Trzesniewski, M. Donnellan, & R. Lucas (Eds.), *Secondary Data Analysis: An Introduction for Psychologists* (pp. 3–10). American Psychological Association.
- Drent, M., Meelissen, M. R., & van der Kleij, F. M. (2013). The contribution of TIMSS to the link between school and classroom factors and student achievement. *Journal of Curriculum Studies*, 45(2), 198–224. <https://doi.org/10.1080/00220272.2012.727872>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103–127. <https://doi.org/10.1037/a0018053>
- Freshwater, D., Sherwood, G., & Drury, V. (2006). International research collaboration: Issues, benefits and challenges of the global network. *Journal of Research in Nursing*, 11(4), 295–303. <https://doi.org/10.1177/1744987106066304>
- Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, 30(2), 119–145. <https://doi.org/10.1177/0165551504042802>
- Gil-Izquierdo, M., & Cordero, J. M. (2018). Guidelines for data fusion with international large scale assessments: Insights from the TALIS-PISA link database. *Studies in Educational Evaluation*, 59, 10–18. <https://doi.org/10.1016/j.stueduc.2018.02.002>
- Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education*, 11, 319–330. <https://doi.org/10.1080/0969594042000304618>
- Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel Structural Equation Models for the Analysis of Comparative Data on Educational Performance. *Journal of Educational and Behavioral Statistics*, 32(3), 252–286. <https://doi.org/10.3102/1076998606298042>
- Green Saraisky, N. (2015). The politics of international large-scale assessment: The Programme for International Student Assessment (PISA) and American education discourse, 2000–2012 (Doctoral dissertation, Columbia University).
- Grek, S. (2009). Governing by numbers: The PISA “effect” in Europe. *Journal of Education Policy*, 24(1), 23–37. <https://doi.org/10.1080/02680930802412669>
- Grundmann, R. (2016). The problem of expertise in knowledge societies. *Minerva*, 55, 25–48. <https://doi.org/10.1007/s11024-016-9308-7>
- Gustafsson, J.-E. (2018). International large scale assessments: Current status and ways forward. *Scandinavian Journal of Educational Research*, 62(3), 328–332. <https://doi.org/10.1080/00313831.2018.1443573>

- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455–476. <https://doi.org/10.1007/s11192-016-1849-x>
- Hogrebe, N., & Strietholt, R. (2016). Does non-participation in preschool affect children's reading achievement? international evidence from propensity score analyses. *Large-Scale Assessments in Education*, 4(1), 1–22. <https://doi.org/10.1186/s40536-016-0017-3>
- Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353. <https://doi.org/10.1080/00313831.2016.1258726>
- Hopmann, S. T., Brinek, G. & Retzl, M. (Eds) (2007). *PISA According to PISA: does PISA keep what it promises?* Lit Verlag
- Huang, C., Yang, C., Wang, S., Wu, W., Su, J., & Liang, C. (2020). Evolution of topics in education research: A systematic review using bibliometric analysis. *Educational Review*, 72(3), 281–297. <https://doi.org/10.1080/00131911.2019.1566212>
- Husén, T. (1979). An international research venture in retrospect: The IEA Surveys. *Comparative Education Review*, 23(3), 371–385. <https://doi.org/10.1086/446067>
- Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, 106(22), 8801–8807.
- IEA. (2013). TIMSS 2015 Assessment Framework. Retrieved from https://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf
- IEA. (2014). *ICILS – Past Cycles*. Retrieved from <https://www.iea.nl/icils-past-cycles>
- IEA. (2016). *IEA International Civic and Citizenship Education Study 2016 Assessment Framework*. Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-3-319-39357-5_2.pdf
- IEA. (2018). *ICCS 2016 Technical Report*. Retrieved from http://archivos.agenciaeducacion.cl/5_Reporte_tecnico_ICCS_2016.pdf
- IEA. (2019a). *Chapter 4: TIMSS 2019 Assessment Design*. Retrieved from <http://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks-Chapter-4.pdf>
- IEA. (2019b). IEA International Computer and Information Literacy Study 2018 Assessment Framework. Retrieved from <https://www.springer.com/gp/book/9783030193881>
- IEA. (2020). *ICCS 2016*. Retrieved from <https://www.iea.nl/studies/iea/iccs/2016>
- Ivanović, L., & Ho, Y. S. (2019). Highly cited articles in the education and educational research category in the social science citation index: A bibliometric analysis. *Educational Review*, 71(3), 277–286. <https://doi.org/10.1080/00131911.2017.1415297>
- Johansson, S. (2016). International large-scale assessments: What uses, what consequences? *Educational Research*, 58(2), 139–148. <https://doi.org/10.1080/00131881.2016.1165559>
- Kamens, D. H., & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, 54(1), 5–25. <https://doi.org/10.1086/64847>
- Kato, M., & Ando, A. (2013). The relationship between research performance and international collaboration in chemistry. *Scientometrics*, 97(3), 535553. <https://doi.org/10.1007/s11192-013-1011-y>
- Keathley-Herring, H., Van Aken, E., Gonzalez-Aleu, F., Deschamps, F., Letens, G., & Orlandini, P. C. (2016). Assessing the maturity of a research area: Bibliometric review and proposed framework. *Scientometrics*, 109(2), 927–951. <https://doi.org/10.1007/s11192-016-2096-x>
- Kijima, R. (2010). Why Participate? Cross-National Assessments and Foreign Aid to Education. In A. Wiseman (Ed.), *The Impact of International Achievement Studies on National Education Policymaking* (pp. 35–61). Emerald Group Publishing.
- Kim, K. W. (2006). Measuring international research collaboration of peripheral countries: Taking the context into consideration. *Scientometrics*, 66(2), 231–240. <https://doi.org/10.1007/s11192-006-0017>
- Kliegl, R., & Bates, D. (2010). International collaboration in psychology is on the rise. *Scientometrics*, 87(1), 149158. <https://doi.org/10.1007/s11192-010-0299-0>
- Knowles, R. T., Torney-Purta, J., & Barber, C. (2018). Enhancing citizenship learning with international comparative research: Analyses of IEA civic education datasets. *Citizenship Teaching & Learning*, 13(1), 7–30. https://doi.org/10.1386/ctl.13.1.7_1
- Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, 33(1), 27–50. <https://doi.org/10.1080/09500693.2010.518645>
- Kwiek, M. (in press). What large-scale publication and citation data tell us about international research collaboration in Europe: changing national patterns in global contexts. *Studies in Higher Education*. <https://doi.org/10.1080/03075079.2020.1749254>
- Kwiek, M. (2018). International research collaboration and international research orientation: Comparative findings about European academics. *Journal of Studies in International Education*, 22(2), 136–160. <https://doi.org/10.1177/1028315317747084>
- Landahl, J. (2020). The PISA calendar: Temporal governance and international large-scale assessments. *Educational Philosophy and Theory*, 52(6), 625–639. <https://doi.org/10.1080/00131857.2020.1731686>
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19(3), 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- Lenkeit, J., Chan, J., Hopfenbeck, T. N., & Baird, J. A. (2015). A review of the representation of PIRLS related research in scientific journals. *Educational Research Review*, 16, 102–115. <https://doi.org/10.1016/j.edurev.2015.10.002>
- Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation analysis: Comparison of Web of Science[®], Scopus[™], SciFinder[®], and Google Scholar. *Journal of Electronic Resources*, 7, 196–217. <https://doi.org/10.1080/15424065.2010.505518>
- Lin, M., Bumgarner, E., & Chatterji, M. (2014). Understanding validity issues in international large scale assessments. *Quality Assurance in Education*, 22(1), 31–41. <https://doi.org/10.1108/QAE-12-2013-0050>
- Lindblad, S., Pettersson, D., & Popkewitz, T. S. (2015). *International comparisons of school results: A systematic review of research on large scale assessments in education*. Stockholm: Swedish Research Council.

- Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, 45(2), 175–194. <https://doi.org/10.1177/0312896219877678>
- Liou, P. Y., & Hung, Y. C. (2015). Statistical techniques utilized in analyzing PISA and TIMSS data in science education from 1996 to 2013: A methodological review. *International Journal of Science and Mathematics Education*, 13(6), 1449–1468. <https://doi.org/10.1007/s10763-014-9558-5>
- Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96(1), 56. <https://doi.org/10.1037/0022-0663.96.1.56>
- Maurissen, L., Barber, C., & Claes, E. (2020). Classroom discussions and political tolerance towards immigrants: The importance of mutual respect and responsiveness. *Acta Politica*, 55, 242–266. <https://doi.org/10.1057/s41269-018-0114-0>
- McVeigh, M. (2009). Citation indexes and the Web of Science. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of Library and Information Sciences* (3rd ed., pp. 1027–1037). Chandos Publishing.
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google scholar. *Journal of the American Society of Information Sciences and Technology*, 58, 2105–2125. <https://doi.org/10.1002/asi.20677>
- Meyer, J., Kamens, D., Benavot, A., Cha, Y.-K., & Wong, S.-Y. (1992). *School Knowledge for the Masses: World Models and National Primary Curricular Categories in the Twentieth Century*. Falmer Press.
- Michaelides, M. P., & Haertel, E. H. (2004). *Sampling of common items: An unrecognized source of error in test linking*. Center for the Study of Evaluation (CSE), University of California
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Mosbah-Natanson, S., & Gingras, Y. (2014). The globalization of social sciences? Evidence from a quantitative analysis of 30 years of production, collaboration and citations in the social sciences (1980–2009). *Current Sociology*, 62(5), 626–646. <https://doi.org/10.1177/0011392113498866>
- Mullis, I. V. S., & Martin, M. O. (2015). *PIRLS 2016 Assessment Framework* (2nd ed.). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/pirls2016/framework.html>
- Neal, J. W., & Neal, Z. P. (2013). Nested or Networked? Future Directions for Ecological Systems Theory. *Social Development*, 22(4), 722–737. <https://doi.org/10.1111/sode.12018>
- Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409. <https://doi.org/10.1073/pnas.98.2.404>
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5200–5205. <https://doi.org/10.1073/pnas.0307545100>
- OECD. (2009). *PISA 2006 technical report*. Author.
- OECD. (2019). *Sampling in PISA*. Retrieved from <https://www.oecd.org/pisa/pisaproducts/SAMPLING-IN-PISA.pdf>
- Pettersson, D. (2008). "Internationell kunskapsbedömning som inslag i nationell styrning av skolan [International Knowledge Assessments: An Element of National Educational Steering]." PhD dissertation, Uppsala University.
- PISA (2020). *PISA FAQ*. Retrieved from <https://www.oecd.org/pisa/pisafaq/>
- Pizmony-Levy, O. 2013. *Testing for All: The Emergence and Development of International Assessments of Student Achievement, 1958–2012*. PhD. diss: Indiana University.
- Pons, X. (2017). Fifteen years of research on PISA effects on education governance: A critical review. *European Journal of Education*, 52(2), 131–144. <https://doi.org/10.1111/ejed.12213>
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21(5), 667–706. <https://doi.org/10.1002/per.634>
- Roemer, R. C., & Borchardt, R. (2015). *Meaningful metrics: A 21st-century librarian's guide to bibliometrics, altmetrics, and research impact*. Association of College and Research Libraries.
- Rolfe, M. K., Bryar, R. M., Hjelm, K., Apelquist, J., Fletcher, M., & Anderson, B. L. (2004). International collaborations to address common problems in healthcare: Processes practicalities and power. *International Nursing Review*, 51, 140–148. <https://doi.org/10.1111/j.1466-7657.2004.00237.x>
- Rutkowski, D., & Delandshere, G. (2016). Causal inferences with large scale assessment data: Using a validity framework. *Large-Scale Assessments in Education*, 4(1), 1–18. <https://doi.org/10.1186/s40536-016-0019-1>
- Rutkowski, D., Rutkowski, L., & von Davier, M. (2014). A brief introduction to modern international large scale assessment. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large Scale Assessment* (pp. 3–10). CRC Press.
- Sellar, S. & Lingard, B. (2013). The OECD and global governance in education. *Journal of Education Policy*, 28(5), 710–725. <https://doi.org/10.1080/02680939.2013.779791>
- Sellar, S., & Lingard, B. (2014). The OECD and the Expansion of PISA: New Global Modes of Governance in Education. *British Educational Research Journal*, 40(6), 917–936. <https://doi.org/10.1086/710777>
- Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41(4), 455–470. <https://doi.org/10.1080/03050060500317810>
- Stiggins, R. J. (1982). An analysis of the dimensions of job-related reading. *Reading World*, 82, 237–247
- Torney, J. V., Oppenheim, A. N., & Farnen, R. F. (1975). *Civic Education in Ten Countries: An Empirical Study*. New York: Almqvist & Wiksell International.
- Torney-Purta, J., & Amadeo, J.-A. (2013). International large-scale assessments: challenges in reporting and potentials for secondary analysis. *Research in Comparative and International Education*, 8(3), 248–258. <https://doi.org/10.2304/rcie.2013.8.3.248>
- Van de Werfhorst, H. G., & Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36, 407–428. <https://doi.org/10.1146/annurev.soc.012809.102538>
- Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>

- Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer.
- Van Rijsbergen, C. J. (1977). A theoretical basis for the use of co-occurrence data in information retrieval. *Journal of Documentation*, 33(2), 106–119. <https://doi.org/10.1108/eb026637>
- Vanderstraeten, R., Vandermoere, F., & Hermans, M. (2016). Scholarly communication in AERA journals, 1931 to 2014. *Review of Research in Education*, 40(1), 38–61. <https://doi.org/10.3102/0091732X16678836>
- von Davier, M., & Sinharay, S. (2013). Analytics in International Large-Scale Assessments: Item Response Theory and Population Models, chapter 7. In L. von Rutkowski, M. Davier, & D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. New York: Chapman and Hall/CRC.
- Wagemaker, H. (2013). International large-scale assessments: From research to policy. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 11–36). CRC Press.
- Waltman, L., Van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. <https://doi.org/10.1016/j.joi.2010.07.002>
- Wang, J., Peters, H. P., & Guan, J. (2006). Factors influencing knowledge productivity in German research groups: Lessons for developing countries. *Journal of Knowledge Management*, 10, 113–126. <https://doi.org/10.1108/13673270610679408>
- Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, 10(2), 347–364. <https://doi.org/10.1016/j.joi.2016.02.003>
- Wiseman, A. W., Astiz, M. F., & Baker, D. P. (2014). Comparative education research framed by neo-institutional theory: a review of diverse approaches and conflicting assumptions. *Compare*, 44(5), 688–709. <https://doi.org/10.1080/03057925.2013.800783>
- Wu, M. L. (2009). "A critical comparison of the contents of PISA and TIMSS mathematics assessments." Paper presented at the NCES "What we can learn from PISA" research conference, Washington, DC, June 2, 2009.
- Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27. <https://doi.org/10.1111/j.1745-3992.2010.00190.x>
- Yemini, M. (2019). International research collaborations as perceived by top-performing scholars. *Journal of Studies in International Education*. <https://doi.org/10.1177/1028315319887392>
- Ynalvez, M. A., & Shrum, W. M. (2011). Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country. *Research Policy*, 40, 204–216. <https://doi.org/10.1016/j.respol.2010.10.004>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
