TRANSDISCIPLINARY RESEARCH ON LEARNING AND TEACHING: CHANCES AND CHALLENGES

EDITED BY: Matthias Stadler, Frank Fischer and Arthur C. Graesser
PUBLISHED IN: Frontiers in Psychology and Frontiers in Education

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-277-9
DOI 10.3389/978-2-88971-277-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

TRANSDISCIPLINARY RESEARCH ON LEARNING AND TEACHING: CHANCES AND CHALLENGES

Topic Editors:
Matthias Stadler, Ludwig Maximilian University of Munich, Germany
Frank Fischer, Ludwig Maximilian University of Munich, Germany
Arthur C. Graesser, University of Memphis, United States

Citation: Stadler, M., Fischer, F., Graesser, A. C., eds. (2021). Transdisciplinary Research on Learning and Teaching: Chances and Challenges. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-277-9

Table of Contents

frontiers in Psychology

EDITORIAL
published: 06 July 2021
doi: 10.3389/fpsyg.2021.696219

Editorial: Transdisciplinary Research on Learning and Teaching: Chances and Challenges

Matthias Stadler[1], Arthur Graesser[2] and Frank Fischer[1]*

[1] Department of Psychology, Ludwig-Maximilians-Universität Munich, Munich, Germany, [2] Department of Psychology and the Institute for Intelligent Systems, University of Memphis, Memphis, TN, United States

Keywords: interdisciplinary, transdisciplinary, Education, teachers, methods

Editorial on the Research Topic

Transdisciplinary Research on Learning and Teaching: Chances and Challenges

OPEN ACCESS

Edited and reviewed by:
Jana Uher,
University of Greenwich,
United Kingdom

****Correspondence:***
Matthias Stadler
matthias.stadler@lmu.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *16 April 2021*
Accepted: *09 June 2021*
Published: *06 July 2021*

Citation:
Stadler M, Graesser A and Fischer F (2021) Editorial: Transdisciplinary Research on Learning and Teaching: Chances and Challenges. Front. Psychol. 12:696219. doi: 10.3389/fpsyg.2021.696219

INTRODUCTION

The goal of the present Research Topic is to provide a forum where research groups, investigating teaching and teachers from multiple perspectives involving multidisciplinary (i.e., different disciplines working on different aspects of a problem independently within their disciplinary boundaries), interdisciplinary (i.e., restructuring and integrating existing disciplinary approaches to address problems relevant for all participating disciplines) and ideally transdisciplinary (i.e., seeking to integrate different lines of work from contributing disciplines to create new approaches or even new scientific disciplines) approaches (Klein, 2017; Hall et al., 2018), can present and discuss the opportunities and challenges of such endeavors. The articles published in the Research Topic can be broadly classified into three categories: Conceptual reviews of transdisciplinary research on teaching and teachers, the results of transdisciplinary research projects, and methodological challenges and innovations related to transdisciplinary cooperation.

CONCEPTUAL REVIEWS OF TRANSDISCIPLINARY RESEARCH

The Research Topic is initiated by Pea and Linn, who provide their personal perspectives on the emergence of the learning sciences as a transdisciplinary research community from the early 1970's to today. In line with the Research Topic's aim, the paper illustrates how the specific approach of the learning sciences integrated approaches from disciplines as diverse as science education, psychology, and computer science to create a new and more holistic scientific discipline devoted to research on learning and instruction under a situated cognition perspective.

Their article is complemented by Lund et al., who discuss how research in education draws widely from the social sciences and humanities. The study uses bibliometric analyses to determine the place of educational research in the larger context of social science research. The authors argue that modern educational research cannot be considered to be a single discipline but rather a multidisciplinary field, thus implementing the initial goal described by Pea and Linn.

This interdisciplinarity of modern educational research is also mirrored in Hmelo-Silver and Jeong's review on the benefits and challenges of interdisciplinarity in computer-supported collaborative learning, a research field remarkably diverse regarding contributing disciplines. While the authors agree that diversity should be cultivated, they also caution to be mindful that research outcomes need to be exchanged and appropriated actively across participating disciplines in order for our understanding of CSCL rises above individual disciplines.

Two articles discussing student learning and development in university settings as a systematic way to integrate different academic disciplines complete the section. Budwig and Alexander take a firm stance for reorganizing universities and curricula on campus efforts to allow disciplinary integration that requires not only alignment and support from the learning and developmental sciences but also local, national, and transnational efforts with relevant learning communities.

Pammer-Schindler et al. present their experiences with interdisciplinary doctoral training on technology-enhanced learning (TEL) in Europe. Based on a survey of 35 doctoral education programs in Europe, the authors argue that cross-institutional doctoral training might be key to progress TEL as a field.

While these challenges are significant, all five articles in this section show how much progress was made in the last decades toward transdisciplinary research on teaching and teachers.

RESULTS OF INTERDISCIPLINARY RESEARCH PROJECTS

The second section of this Research Topic takes a step back from the theoretical challenges of transdisciplinary research to illustrate concrete lessons learned from current transdisciplinary research projects.

Schilcher et al. detail how the FALKE (Fachspezifische Lehrerkompetenzen im Erklären; Engl.: subject-specific teacher competency in explaining) research project integrates 14 heterogeneous disciplines in order to examine the pedagogical quality of teacher explanations empirically. The authors discuss how trans-, multi-, and interdisciplinary projects, in particular, are primarily shaped by the nature of the problem, the scientists and stakeholders involved, and the institutional setting. Moreover, they present an example on how to tackle some of these issues.

Closely related to this, Heitzmann et al. illustrate the potential but also the challenges of large transdisciplinary projects. The authors review why many promising projects fail beyond the actual research conducted. They argue that ideas from the field of collaborative problem solving have the potential to yield valuable insights when designing or conducting cross-disciplinary research in learning and instruction.

Bauer et al. present an innovative analytic approach based on epistemic network analysis to compare diagnostic activities in medical and teacher education. Based on their results, the authors recommend that educators think beyond individuals' knowledge and systematically teach and increase the awareness of disciplinary standards.

Finally, Fleckenstein et al. investigate whether text length is a construct-relevant aspect of writing competence, a transdisciplinary issue concerning the research areas of educational assessment, language technology, and classroom instruction.

All articles in this section provide examples of successful interdisciplinary research projects and highlight the challenges that come with such endeavors.

METHODOLOGICAL CHALLENGES AND INNOVATIONS

The final part of this Research Topic focuses on the methodological challenges posed by transdisciplinary research. Lindl et al. provide recommendations on tackling the often highly complex data resulting from investigating unique subject-specific aspects on the one hand and transdisciplinary, generalizable effects on the other. They compare meta-analysis, multilevel models, latent multilevel structural equation models, and machine learning methods discussing the advantages and disadvantages of all methods.

Levy et al. contrast classical and machine learning approaches in estimating value-added scores in large-scale educational data. Aside from statistical features, the authors discuss possible ethical concerns and practical implications regarding using machine learning methods for decision-making in education.

The Research Topic is concluded by Rienties et al. The authors review future research directions on teaching and teacher education, defining the boundaries between artificial intelligence in education, computer-supported collaborative learning, educational data mining, and learning analytics. The article encourages researchers to cross the boundaries of their respective fields and work together to address the complex challenges in education.

In this collection, we included meta-level and theoretical papers on collaborations between various disciplines in research on learning, the design of learning environments, and teaching. These approaches can serve as models for future collaborations to tackle complex phenomena and problems that are beyond what individual disciplines can tackle successfully.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was funded by grants from the Deutsche Forschungsgemeinschaft (for 2385; COSIMA; Teilprojekt M) and the National Science Foundation Learner Data Institute (grant # 1934745).

REFERENCES

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., et al. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *Am. Psychol.* 73:532. doi: 10.1037/amp0000319

Klein, J. T. (2017). "Typologies of interdisciplinarity: the boundary work of definition," in *The Oxford Handbook of Interdisciplinarity,* eds R. Frodeman, J. T. Klein, and R. C. S. Pacheco (Oxford: Oxford University Press), 21–34.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

CONCEPTUAL ANALYSIS
published: 16 July 2020
doi: 10.3389/feduc.2020.00130

Personal Perspectives on the Emergence of the Learning Sciences: 1970s–2005

Roy Pea[1] and Marcia C. Linn[2]*

[1] H-STAR Institute, Graduate School of Education, Stanford University, Stanford, CA, United States, [2] WISE TELS Research, Graduate School of Education, University of California, Berkeley, Berkeley, CA, United States

We describe the emergence of the interdisciplinary learning sciences field and its consequential transformations, drawing on experiences that brought us together. Starting with our undergraduate years, the account culminates with the formation of the International Society of the Learning Sciences (ISLS). We identify six themes shaping the emergence of the learning sciences and our own trajectories: (a) broadening the community and incorporating new disciplinary perspectives; (b) appropriating and developing new methods; (c) reconceptualizing challenges; (d) creating artifacts; (e) developing abstractions; and (f) developing people. We intend this personal account to stimulate new initiatives and deepening insights as the journey of the learning sciences continues.

Keywords: learning science, design-based research, interdisciplinarity, education, teaching

OPEN ACCESS

Edited by:
Matthias Stadler,
Ludwig Maximilian University of Munich, Germany

Reviewed by:
Bart Rienties,
The Open University (United Kingdom), United Kingdom
Beth O'Brien,
Nanyang Technological University, Singapore

****Correspondence:***
Roy Pea
roypea@stanford.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *18 April 2020*
Accepted: *29 June 2020*
Published: *16 July 2020*

Citation:
Pea R and Linn MC (2020) Personal Perspectives on the Emergence of the Learning Sciences: 1970s–2005. Front. Educ. 5:130. doi: 10.3389/feduc.2020.00130

INTRODUCTION

We describe the emergence of the interdisciplinary learning sciences field and its consequential transformations drawing on our own experiences. We start with our undergraduate and graduate years, using our first names to describe our separate experiences. We refer to our joint perspectives using "we." We highlight a series of opportunities that brought us together shortly after graduate school and have arisen throughout our careers. We follow the development of the field up to 2005, including the formation of the International Society of the Learning Sciences (ISLS) in 2002. We conclude with themes that emerged in our own trajectories and which shaped the learning sciences.

While our experiences intersect with those of our international colleagues and research programs, they inevitably skew toward programs of scientific research funding and educational policies in the United States, where we have lived and worked for the past four decades. We have extensively learned from and deeply appreciated the profound contributions of the long-standing computer supported collaborative learning (CSCL) community, which is solidly international in its origins and leadership. We look forward to reading related personal historical accounts by all the learning sciences contributors.

To characterize learning science, we take a Wittgensteinian approach in which the meaning of terms is defined by their uses. We have often argued that "learning sciences" *is* simply what "learning scientists" *do*. We do not offer nor seek a set of necessary and sufficient conditions that define "learning science" or "learning scientist." Rather, we capture themes to characterize the emergence of the learning sciences and highlight some events and experiences to illustrate the trajectory.

INITIAL STEPS TOWARD THE LEARNING SCIENCES

The learning sciences interdisciplinary approach to deepening understanding involving *instruction, psychology, and computer science* initially attracted both Roy and Marcia. These interests converged when the National Institute of Education (NIE) sought research on the cognitive consequences of computer programming.

Beginnings

Marcia and Roy both developed interdisciplinary interests starting as undergraduates and benefitted from mentors who nurtured their nascent desires to bridge multiple fields. These mentors were generous in brokering opportunities to advance their professional learning pathways.

As a Stanford undergraduate, Marcia's initial interest in mathematics morphed into an interdisciplinary focus on statistics, computer science, and models of student learning in psychology. She joined Richard Atkinson's group as an undergraduate and learned about the nuances of mathematical learning theories. She served as an unpaid intern for Patrick Suppes who founded the Computer Curriculum Corporation in 1967 to catalyze the computerized learning movement. Marcia explored the mainframe program that tutored students in logic. The learning sciences were a perfect focus for her interests in multiple aspects of learning, instruction, and technology.

Marcia strengthened her knowledge of learning and computing in her first job working for a startup founded by classmate Larry Tesler. Having taken one computer science course where she learned Algol, Marcia developed her programming expertise as an apprentice to Tesler. Tesler eventually joined Xerox PARC and then Apple where he co-invented the Lisa machine and became Chief Scientist. She wonders how things might have unfolded if she had persisted in the technology industry rather than entering graduate school.

As a graduate student at Stanford, Marcia joined the School of Education research group led by Lee Cronbach where her interests bridged computing, psychometrics, and complex reasoning. When Cronbach went on sabbatical, he arranged for her to spend a year in Piaget's Genevan lab where she struggled to learn French and explored the reasoning elicited in clinical interviews. Cronbach encouraged her interest in studying instructional scaffolds for complex reasoning to measure Vygotsky's zone of proximal development. Piaget teased her about her interest in instruction, calling it "the American question." Marcia designed clinical interviews to explore her question and conducted them at the American School in Geneva. Returning to Stanford, Marcia wrote a computer simulation of a clinical interview for a course taught by Edward Feigenbaum. Encouraged by her committee, Cronbach, Richard Snow, and Ernest Hilgard, she studied the interactions between instructional supports and complex reasoning in her dissertation, motivating a long-term interest in how instruction can guide learners to integrate their ideas.

Roy's interdisciplinary interests in what became the learning sciences began with his serendipitous undergraduate opportunity to study epistemology with philosopher of science Stephen Toulmin, one of Wittgenstein's last Cambridge University Ph.D. students. Toulmin was Roy's largest influence as a lifelong mentor since his freshman year and later as a colleague at Northwestern. Stephen encouraged Roy's pursuit of an independently defined major in 'cognition' at Michigan State University from 1970 to 1974, enacted as a double major in psychology and philosophy with a minor in linguistics, a few years before the cognitive sciences would emerge in 1979 as an international society and a journal. Stephen focused Roy's interests on the need for empirical studies which would illuminate the philosophical issues embodied in the development of logic, language, and cognition in social context.

These interests were well met by the opportunity to study child language development in Jerome Bruner's new laboratory at Oxford University's Department of Experimental Psychology, where, in 1974, Roy joined Bruner's lab as his doctoral advisee. As Roy wrote his dissertation in 1977, Bruner encouraged him to join George Miller's and Michael Cole's research groups at Rockefeller University. In doing so, Roy was able to bridge the experimental psycholinguistic paradigms of Miller's Lab and the cultural psychological, anthropological, and video interaction analytic studies of Cole's Laboratory of Comparative Human Cognition, where he also learned from Sylvia Scribner and Ray McDermott.

Roy integrated these varied approaches and research questions he encountered in Oxford and New York when he began to study children's learning with computers with his New York City colleagues at the Bank Street College of Education's Center for Children in Technology from 1980 to 1986. These interdisciplinary studies began with a watershed project funded by the Spencer Foundation: "The Impact of a Classroom Computer Experience on Children's Problem-Solving, Planning, and Peer Collaboration" (1981–1984). Roy and CCT Director Karen Sheingold were Co-Principal Investigators, with Jan Hawkins a central collaborator, as socio-cognitive developmental psychologist.

Sister Grants for Studying Cognition and Computing, 1983

Marcia and Roy's careers converged in 1983 when their interdisciplinary paths prepared them to win the first two grants awarded by the US Department of Education's NIE to research the cognitive consequences of computer programming (see **Table 1**). Marcia in California was a leader of one in collaboration with Bill Rohwer and Ellen Mandinach, a recent Lee Cronbach Ph.D. Roy in New York City was a leader of the other, in collaboration with Midian Kurland, a recent Robbie Case Ph.D. and Ann Brown postdoc.

In this bi-coastal work, the projects explored distinct learning contexts and held regular networking discussions on research priorities and methodologies, culminating in a special issue they co-edited: (Mandinach et al., 1986). The two projects drew on expertise from software designers, computer

TABLE 1 | The National Institute of Education competition in 1982 resulted in two large grants to investigate the cognitive consequences of programming.

1983–1985: Linn et al., Principal Investigators, NIE funded project: OE 0400-83-0017: Assessing the Cognitive Consequences of Computer Environments for Learning.
1983–1985: Pea R. and Kurland D.M., Principal Investigators, NIE funded project: OE 0400-83-0016: The Demands and Cognitive Consequences of Learning to Program.

scientists, cognitive and developmental psychologists, science and technology precollege teachers, science educators, and educational anthropologists. They helped each other refine methodologies for this emerging field, exploring case studies of expert child programmers; observational studies of computer science classroom instruction; and design studies of assessments of student progress in learning programming languages and of transfer of planning and problem solving from learning programming to other domains. The artifacts produced included introductory programming languages, curriculum materials, and assessments.

NATIONAL OPPORTUNITIES SHAPING THE LEARNING SCIENCES

Many national initiatives brought together and funded interdisciplinary collaborations that shaped the learning sciences. As the scope of the learning sciences expanded to incorporate new fields, our work both contributed and benefitted.

Board of Reviewers: National Science Foundation (NSF) Research on Teaching and Learning (RTL), 1983

Starting around 1983 and continuing until around 1990, Roy and Marcia had the opportunity to help shape the research agenda for the learning sciences by serving on the standing Board of Reviewers for the National Science Foundation (NSF) Research on Teaching and Learning (RTL) program led by Raymond Hannapel. This program emerged "*out of the dust*" after newly elected President Ronald Reagan's 23% slash in the Fiscal Year 1982 NSF budget. Although the NSF is an independent federal agency created by the US Congress in 1950 "to promote the progress of science; to advance the national health, prosperity, and welfare" and funds about 24% of all federally supported basic research conducted by America's colleges and universities, Reagan scuttled all programs in science education and behavioral sciences. Fortunately, incoming NSF Director Edward Knapp oversaw a FY 1984 increase of 35% in appropriations compared to FY82, enabling the re-establishment of key programs in science education including RTL.

As members of the Standing Board of Reviewers we met twice a year alongside mathematics and science educators and learning researchers including Audrey Champagne, Robert Davis, Jim Kaput, Judy Sowder, and a changing group of other STEM learning scholars. We collectively developed a trajectory for the emerging research funded by the RTL program. We reviewed grant proposals and recommended funding awards to build an interdisciplinary field. Reviewers negotiated the meaning of interdisciplinarity, valuing it in both the mix of funded projects and in the leadership of each grant. Funded projects, led by interdisciplinary teams from many fields, contributed to the emerging field of learning sciences. Teams typically included cognitive psychologists, science and mathematics educators, and experts in the relevant disciplines (science, mathematics, computer science). Funded projects conducted research in K-12 schools, science museums, or other complex settings, illustrating the growing purview of the learning sciences. The RTL program brought together individuals from a wide range of fields and produced methodologies, artifacts, and abstractions that bridged those fields. To illustrate, lead investigators included cognitive psychologists such as John Anderson (CMU), John Bransford's Cognition and Technology Group (Vanderbilt), Jim Greeno and Lauren Resnick (U. Pittsburgh), David Klahr (CMU), Carol Smith and Susan Carey (Harvard); cultural psychologists such as Geoffrey Saxe (UC Berkeley); computer scientists Andrea diSessa (Berkeley), Wallace Feurzeig, John Frederiksen, John Richards, and Barbara White (BBN); science educators such as John Clement (U Mass, Amherst), Fred Goldberg (San Diego State), Richard Hake (Indiana), David Hestenes (U. Arizona), Lillian McDermott (U. Washington), Joe Novak (Cornell), Fred Reif (UC Berkeley); and math educators such as Jere Confrey (Cornell), Elizabeth Fennema (U. Wisconsin), James Hiebert (U Delaware), Glenda Lappan (Michigan State), Alan Schoenfeld (UC Berkeley), Ed Silver (Pittsburgh), Leslie Steffe (U. Georgia).

Community Building: The Science of Science Education, 1986

Community-building initiatives funded by NSF introduced leaders from disparate fields to shape the learning sciences. "One meeting I recall vividly" Roy noted, was in January 1986 at the Lawrence Hall of Science at the University of California, Berkeley where vibrant discussions arose among leaders from disparate fields including Fred Reif (UC Berkeley physics educator), Jill Larkin (Herb Simon's CMU protégé), Jim Greeno (UC Berkeley cognitive scientist working on early mathematical cognition, recently arrived from University of Pittsburgh's Learning Research and Development Center), Andy DiSessa (UC Berkeley physics educator, new arrival from Papert's MIT lab), Lauren Resnick and Bob Glaser (U. Pittsburgh's LRDC co-directors), Alan Schoenfeld (mathematics educator, U. Rochester), and Glenn Seaborg (Nobel Laureate, UC Berkeley Chemistry Professor). The meeting was co-hosted by Marcia who reported that 45 mathematicians, scientists, cognitive scientists, mathematics and science educators, and curriculum and technology experts, often meeting each other for the first time, convened at Berkeley for a planning conference on research and science education. The conference concluded that leaders from these diverse fields "must combine their efforts to add to a systemic, comprehensive research base" (Linn, 1987, p. 192).

Cultural Foundations: The Institute for Research on Learning, 1987–2000

The centrality of the social and cultural foundations of learning in environments beyond schools developed at the Xerox-funded non-profit think tank The Institute for Research on Learning and at Stanford University's School of Education from 1987 to 2000 and have had far reaching impacts on the field. Roy's distributed intelligence studies of learning interactions (Pea, 1987, 1993b,c, 1994) contributed to understanding of the foundational nature of culture. This occurred alongside the development of the "situated learning" perspective in the learning sciences anchored in the works of IRL researchers (Lave, 1988; Brown et al., 1989; Lave and Wenger, 1991; Greeno and The Middle School Mathematics Through Applications Project Group [MMAP], 1998).

Design-Based Research: National Design Experiments Consortium (NDEC), 1990

Broadening the scope of learning to encompass cultural and social factor while also expanding the expertise relevant to the study of learning, motivated the 1990 formation of the *National Design Experiments Consortium* (NDEC). Annual meetings facilitated the development of design-based research methods now recognized as at the heartland of the learning sciences. This network of American learning and technology lab leaders organized by Jan Hawkins out of the Center for Children and Technology involved leaders of many groups funded by the NSF RTL program including Marcia and Roy, Ann Brown, John Bransford, Joseph Campione, Sharon Carver, Allan Collins, John Frederiksen, Shelley Goldman, Susan Goldman, Jim Greeno, Marlene Scardamalia, and Janet Schofield. These leaders grappled with the design of innovations to promote complex skills such as inquiry learning as well as the methods required to establish their validity. Design-based research methods were further refined across multiple scholarly networks including special interest groups in established organizations such as the American Educational Research Association (AERA).

Learning Sciences Organizations: AERA SIG-EST, 1990

Taking advantage of the growing community of learning scientists attending the annual meetings of the leading American forum for educational research, The AERA, Marcia and Roy as founding co-chairs launched AERA's SIG-EST: Education in Science and Technology. Over three decades, this organization morphed into the current AERA SIG-Learning Sciences and Advanced Technologies for Learning. These annual SIG symposia and paper sessions continue to attract emerging learning sciences initiatives, whether founded in design-based research, situative learning perspectives, or expanding on the advanced technologies represented in studies of the learning sciences.

Learning Sciences Ph.D. Program, 1992

Responding to growing interest, Northwestern in 1992 launched the first doctoral program called the learning sciences (Pea, 2016; Schank, 2016). The scope of "learning sciences" and the definition of "learning scientist" have both subsequently expanded. Roy oversaw the design of the program with his interdisciplinary colleagues in psychology, education, and computer science and directed it in its first years. Its three emphases and integrative focus (Pea, 1993a) built upon Roy's formative experiences in research on children's learning and classroom studies of children using computers. The initial program description foreshadows the field's eventual developments:

> The design and use of technologies play a special role in Learning Sciences inquiries. Multimedia computing and telecommunications are increasingly prevalent in society, in the world of work, and in schools, as new tools for enhancing workplace activities and educational practices. Computer tools have also served as new instruments for investigative research on cognition, learning, and social interaction. Integrations of computing and video provide tools for deeper analyses of learning and teaching situations, and designs for novel architectures of learning, teaching, and assessment tools. Research and theory in the Learning Sciences Program pays constructive and critical attention to these issues by integrating three areas of specialization in its core coursework and methodological foundations:
>
> > *Environments*: Deepening understanding of the social, contextual, and cultural dynamics of learning in situations ranging from classrooms to out-of-school settings.
> >
> > *Cognition*: Articulating scientific models of the structures and processes of learning and teaching of organized knowledge, skills, and understanding.
> >
> > *Architectures*: Theory-guided design, construction, and use of multimedia computing and telecommunications technologies for supporting learning and teaching processes (op cit., p. 27).

The pursuit of a Ph.D. in the Learning Sciences will provide students with a deep and action-oriented understanding of the dynamics of learning environments; the nature of the cognitive processes involved in learning and teaching; and how to design, construct, and use technology to support the learning and teaching processes (op. cit., p. 38).

Broadband Networking and Technology-Enhanced Learning, 1992: CoVis, CLP, KIE, and WISE

The emerging broadband network supporting opportunities for technology-enhanced learning enticed both Roy and Marcia to initiate research funded by NSF starting in 1992. Marcia and Roy continued to build on each other's work by advising each other's projects. The multidisciplinary projects funded during this time broadened the fields involved in the learning sciences and prepared a new generation of leaders.

Co-Vis. Beginning in 1992, in collaboration with Northwestern University colleagues Louis Gomez and Daniel Edelson, Roy served as PI of the Collaborative Visualization (CoVis) Project funded by NSF and industry partners Bellcore and Ameritech: "The CoVis Collaboratory: High school science learning supported by a broadband educational network with scientific visualization, videoconferencing, and collaborative computing."

The CoVis project abstract reveals how at the edge of possibility the networked learning environments we sought to develop were:

> "The next decade brings widespread, networked, multimedia interpersonal collaborative computing. Data collection, exploration, analysis, and collaborative work is being transformed throughout science by new flexible data visualization and communications tools. A question-centered and collaboration-focused pedagogy is supplanting more traditional didactic K-12 instruction. The Learning Through CoVis Project will install a high-bandwidth testbed network using public-switched ISDN services to support synchronous and asynchronous collaboration with rich data sharing (e.g., complex images, large datasets) and desktop videoconferencing among high school students across schools, who also use the network to communicate with university researchers and other scientific experts. We describe students' uses of new CoVis tools for supporting collaborative project-enhanced science learning: a multimedia 'collaboratory notebook,' and specially-tailored visualization tools for atmospheric science allowing students to record their work and thinking during project-based inquiry using the same data as leading scientists."

Since one of Roy's CoVis collaborators was the UIUC's National Center for Supercomputing Applications (NCSA), and its atmospheric sciences faculty, when undergraduate Marc Andressen created Mosaic, the first publicly released World Wide Web browser, our CoVis Project was one of the first to establish a "distributed multimedia learning environment" (Pea and Gomez, 1992) employing the Internet for CSCL (Pea, 1993d; Ramamurthy et al., 1996); for more details, see Edelson et al., 1996; Pea et al., 1997; Gomez et al., 1998). Among our first learning sciences doctoral students at Northwestern working on this project for their dissertation research studies were Barry Fishman and Joseph Polman (ISLS Past-President), now both ISLS Fellows.

CLP, KIE, and WISE 1984–1999. In 1984 Marcia began a lifelong collaboration with the late inventor, visionary, and physicist Robert Tinker (Chief Science Officer at TERC, later President and founder of the Concord Consortium) who was developing probeware technologies for classroom computers, and with middle school science teacher Douglas Kirkpatrick (fondly known as Mr. K). The group started developing and testing ways to leverage probeware for real-time data collection, to teach thermodynamics in a series of design studies initially supported by an Apple Wheels for the Mind Grant of 16 Apple][computers. In 1988 the NSF funded a project called Computer as Laboratory Partner (CLP). In 1992, with increasingly powerful computers becoming available, the group took advantage of networked communication within the classroom as well as interactive scientific models and simulations while also broadening the focus to thermodynamics plus light, and sound. Funded by a new NSF grant called CLP, the new technologies made it possible to study ways to design productive online classroom discussions and investigate the impact of connecting real-time experiments to simulations of everyday scientific phenomena. Students were able to test ways to keep a drink cold for lunch or investigate ways to propagate light for room illumination. Marcia collaborated with Sherry Hsi, a former graduate student (now Vice President of the Concord Consortium) to synthesize this work in a constructivist instructional framework called knowledge integration and in a set of design principles to help guide instructional decision making (Linn and Hsi, 2000).

Our collaboration, like CoVis, immediately began to explore the advantages of Mosaic, a user friendly, graphical web browser. In 1994 in collaboration with graduate students Philip Bell (now Professor, University of Washington) and Betsy Davis (now Professor, University of Michigan) we proposed the Knowledge Integration Environment (KIE) and developed the first web-enabled learning environment for K-12 science informed by the knowledge integration framework (Bell et al., 1995). KIE leveraged existing web resources such as the UC Berkeley repository of images of frog deformities and sought to instill a healthy skepticism of uncurated resources. KIE researched instructional patterns that could provide designers a starting point when wishing to use online resources to promote critique, argument construction, collaborative investigations, and hands-on learning. KIE refined the design-based research paradigm with the goal of developing detailed design knowledge while strengthening theoretical knowledge of learning and cognition. Marcia collaborated with Davis and Bell to edit the 2004 book "Internet Environments for Science Education," capturing the contributions of the KIE team (Linn et al., 2004). The stunning collaborators have gone on to become leaders in the learning sciences. Graduate students included Douglas Clark (Professor, University of Calgary), Brian Foley (Professor, California State University), Chris Hoadley (Professor, NYU), Sherry Hsi (Vice President, Concord Consortium), Eileen Lewis (NSF), Linda Shear (SRI) and Nancy Songer (Emeritus Professor, University of Michigan and Dean, University of Utah), and Judith Stern (Education Technology Services, UC Berkeley).

A major contributor to this work starting in 1996 was James Slotta (a student of LRDC's Micki Chi), who is now a University of Toronto Professor and ISLS Board Member. Slotta joined as a postdoctoral scholar and designed the next generation of KIE, the Web-based Inquiry Science Environment (WISE) project, funded in 1998.

NSF Centers for Learning Technologies 1995–2005

Leaders in education and computer science at NSF began to envision a transformation of understanding of learning and instruction made possible by a combination of advanced technologies and understanding of cognition. In October 1995, these leaders convened a multidisciplinary workshop to set a Computer Science research agenda in educational technology. The goal was "to conduct, in a collaborative fashion, interdisciplinary research and systems development that can lead to significant breakthroughs in our understanding of learning and cognitive functioning—from empirical research to theory development to classroom practices—as well as in the application of advanced technologies and new understanding of cognition and the learning process to intelligent systems to use in all facets of education, including informal and self-directed learning"

(Sabelli and Pea, 2004; Pea, 2004, pp. 2–3). In response to this agenda, NSF solicited proposals to establish one or more centers for collaborative research on learning technologies, with the expectation that these centers would have the ability to undertake large, cross-disciplinary projects; to act as technology transfer mechanisms by training new researchers; to support prototype or model projects; and to be impartial and comparative evaluation centers for learning technologies.

Three 4-year centers were established with differing philosophies for how to leverage their activities and achieve broad impact:

(1) *The Center for Innovative Learning Technologies*. CILT was formed to stimulate the development and implementation of important, technology-enabled solutions to critical problems in K–14 STEM learning. CILT was an open and inclusive national effort led by PIs at five institutions: Barbara Means (SRI), Roy Pea (SRI, Stanford), Marcia Linn (UC-Berkeley), John Bransford (Vanderbilt), and Robert Tinker (Concord Consortium). It focused on empowering research advances in learning using technology, specifically, in visualization, assessment, community tools, and ubiquitous computing (e.g., Pea et al., 1999). CILT was especially effective in broadening the community involved in the learning sciences. Funded from 1997 to 2003 for a total of $7.5M, CILT brought together researchers from a broad range of institutions along with technology industry leaders, precollege administrators and teachers, disciplinary specialists, software designers, and graduate students to develop research agendas and stimulate new initiatives. To synthesize the contributions of these individuals CILT developed a model of Synergy Research (see **Figure 1**).

(2) *The Center for Learning Technologies in Urban Schools*. LeTUS was formed to better serve urban science education needs through innovative, hands-on, project-based curricula. The center's premise was that urban schools represent a challenging and important setting for shaping and assessing new organizational and teaching practices supported by technology. LeTUS was a partnership among the Chicago Public Schools, the Detroit Public Schools, Northwestern University, and the University of Michigan[1]. LeTUS sought to imbue educational systems with technology supports for their own reform efforts, specifically, in science education and inquiry.

(3) *The Center for Interdisciplinary Research on Constructive Learning Environments*. CIRCLE had three main goals: first, to understand an extremely effective pedagogy, human tutoring; second, to build and test a new generation of computer tutoring systems that encourage students to construct the target knowledge; and third, to help integrate this new technology into existing educational practices. CIRCLE was a partnership between the University of Pittsburgh and Carnegie Mellon University[2]. CIRCLE sought to advance a learning technology (artificial intelligence tutoring systems) and disseminate its findings to the AI R&D communities.

Emerging Socio-Cognitive Scaffolding Systems

Systems developed to catalyze the variation and cohesion among the socio-cognitive scaffolding systems emerging in learning sciences projects such as CSILE, CoVis, KIE, and WISE, and Kids as Global Scientists that were informed by pedagogical principles from the cognitive sciences. These scaffolding systems structure classroom network-based or

[1]http://www.LeTUS.org

[2]http://www.pitt.edu/~circle/

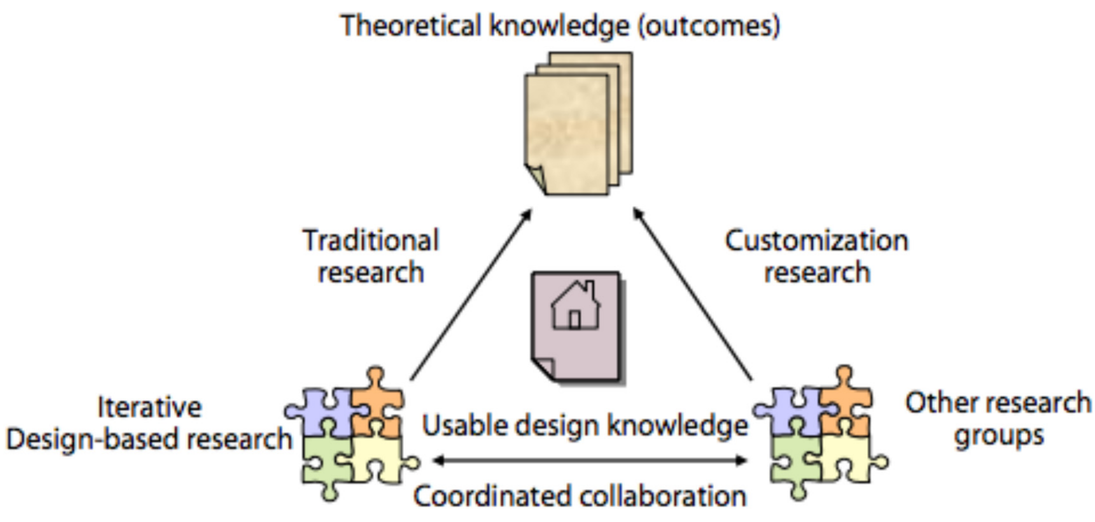

FIGURE 1 | A model of synergy research. The Center for Innovative Learning Technologies (CILT) leveraged close links among postdocs and participants in each CILT group with the partner schools, to explore issues of replication and accumulation of knowledge, which are so crucial to dissemination and implementation projects. The project developed methods for synthesizing like innovations where developers of an innovation attempt to create a similar educational activity using a novel learning environment and test it in the same classrooms where the initial research took place. Customization research explores the value of usable design knowledge in new contexts. Results of this research, along with coordinated collaboration across research groups, refine the outcomes of traditional education research. Source: https://web.stanford.edu/~roypea/RoyPDF%20folder/A121_CILT_fmal.pdf.

distributed learning models that facilitate the conduct of complex thinking, inquiry, and knowledge building. For example, CILT researchers developed a learning technologies vision paper for the 1999 National Governors' Association meeting (Means et al., 1999). The NSF LeTUS center (whose researchers including Louis Gomez, Joseph Krajcik, and Barry Fishman were frequent CILT workshop contributors), organized the 2004 special issue on "Scaffolding in Science Learning" of the *Journal of the Learning Sciences* (JLS). The CILT design principles database informed by the knowledge integration framework synthesized emerging insights from computer-based learning environments to guide designers of curriculum and instruction (Kali, 2006). The effort continues today[3].

INTERNATIONAL COLLABORATIONS AND THE INTERNATIONAL SOCIETY OF THE LEARNING SCIENCES, 2002

International collaboration was spurred by NATO conferences involving both North American and European participants, the founding of the influential *JLS*, started in 1991, along with special interest groups focused on learning sciences that emerged at international conferences including the European Association for Research on Learning and Instruction (EARLI) and the European Science Education Research Association (ESERA). These and related activities contributed to the 2002 founding of the ISLS. With leadership from ISLS, the international field of the learning sciences has grown and thrived.

NATO Conferences 1988–1993

NATO Advanced Research Workshops, part of the NATO Special Program on Advanced Educational Technology under the auspices of the NATO Science Committee occurred from 1988 to 1993. Designed to bring together researchers from the United States and Europe, each conference featured published proceedings. Marcia was thrilled to be a co-organizer of a NATO workshop led by Erik De Corte that Roy also attended (De Corte et al., 1992). The conference entitled *Computer-based learning environments and problem solving* featured amazing three course lunches with wine along with opportunities to build enduring relationships between United States and European leaders. The NATO workshop organized by Tinker (1996) entitled *Microcomputer Based Labs: Educational Research and Standards* was especially exciting because it brought together physicists from Europe and the United States who had independently designed education-oriented probeware and were genuinely interested in the relationship of their work to research on learning (Linn, 1996). Another very influential NATO workshop was hosted in 1989 by Claire O'Malley in Maratea, Italy, on *CSCL* (O'Malley, 1994). This workshop was the precursor to the CSCL conference that became integral to ISLS.

[3] http://wise.berkeley.edu/design/

United States–German Collaboration 2002–2003

One exciting conference in 2001 attended by Roy and Marcia convened a cross-Atlantic collaboration of US NSF-funded researchers and German researchers funded by the German Science Foundation (DFG). It was entitled "Research Methods for International Collaboration" and held in Freiburg, Germany. The focus illustrated the advantage of merging fields and integrating research methods. Initially the differences between European and US methods led to discussions about which approach was more valid. Eventually, the discussion turned productively to explorations of the tradeoffs between laboratory investigations and research in classrooms or out-of-school settings. This led to reflections on aspects of validity, utility, and generalizability. A follow-on NSF-funded conference entitled, "Implementation of an American-German research network in the field of technology-supported education," led by Roy and Ken Koedinger (CMU) brought German and US learning scientists together in Washington DC in 2004 to formulate a collaborative research agenda for technology supported education. This network has grown and flourished, benefitting from the leadership of Frank Fischer from Ludwig-Maximilians-Universität (LMU) in Munich.

The International Society of the Learning Sciences, 2002

The International Society of the Learning Sciences was founded in 2002 by Chris Hoadley, Janet Kolodner, and Tim Koschmann. Both Roy and Marcia have served as President of ISLS and were Inaugural Fellows. ISLS set out to unite the traditions started by the JLS, the International Conferences of the Learning Sciences (ICLS), and the CSCL Conferences. This marked the coalescence of the field as reflected in the ISLS vision statement:

> The educational challenges of our world are increasingly global, requiring interdisciplinary problem solving, knowledge building, and collaboration involving multiple forms of expertise for better understanding the complex phenomena of learning and for guiding the design and improvement of learning environments for valued outcomes. The ISLS is the leading professional society for academics, professionals, and students who seek to advance the sciences and practices of learning, broadly speaking, with special attention to how they may be augmented by technology. ISLS brings together those interested in learning experiences across schools, homes, workplaces, and communities, and who seek to understand how learning and collaboration is enabled by knowledge, tools and networks, and multiple contexts of experience and layers of social structures.

Today ISLS brings members together to advance the themes we identified as characterizing the learning sciences:

Broadening the Community and Incorporating New Disciplinary Perspectives

The International Society of the Learning Sciences actively recruits members from every continent and country and welcomes new disciplinary perspectives in pre-conference workshops and presentations. For example, a preconference workshop offered for the 2020 conference is entitled, "Expanding the field: How the learning sciences might further computing education research." In addition, as President of ISLS Frank Fischer initiated the Network of Academic Programs in the Learning Sciences (NAPLeS) to foster high-quality Learning Sciences programs by developing online materials for instructors, supporting student exchanges, developing a repository of course syllabi, and forming a community of programs that now includes over 60 universities.

Appropriating and Developing New Methods

The International Society of the Learning Sciences members regularly offer preconference workshops on new methods, up-to-date ways to use existing methods (such as design-research), and ways to use methods formerly developed in other fields such as data science for learning analytics. Both the JLS and the *International Journal of Computer Supported Collaborative Learning* (iJCSCL) regularly publish articles featuring new methods.

Creating Artifacts

The ISLS conferences are ideal ways to encourage creation of artifacts and to introduce new artifacts to a receptive audience. Both JLS and iJCSCL publish articles reporting research on learning sciences artifacts.

Developing Abstractions

The ISLS conferences are ideal ways to develop and refine abstractions, to encourage creation of artifacts whose uses for learning will be researched to further refine abstractions, and to introduce new artifacts and abstractions to a receptive audience keen to build on the latest works.

Developing People

The International Society of the Learning Sciences has a wide range of activities designed to develop members and to attract newcomers to join the organization. The Doctoral Consortium and Early Career Workshop are sought-after opportunities for organization members. Marcia, as President of ISLS and chair of the Education Committee, facilitated the development of a newcomers' event to welcome new members, and a mid-career workshop to support members as they navigate new challenges post-tenure such as managing leadership responsibilities, taking up new research foci, or mentoring younger faculty.

THEMES SHAPING THE LEARNING SCIENCES

Reflecting on the trajectory of the learning sciences, we identify six themes that emerged as the learning sciences grew and expanded over the past four decades. These themes are ongoing areas of intellectual work.

Broadening the Community and Incorporating New Disciplinary Perspectives

The learning sciences community welcomes new disciplinary perspectives and incorporates them systemically. These perspectives strengthen understanding of all the fields involved including learning, development, instruction, technology, computer science, linguistics, anthropology, neuroscience, cultural studies, and others. Many participants embraced the learning sciences because they were already bridging several fields and valued others who shared their interdisciplinarity. Roy's interests in philosophy and psychology and Marcia's interests in computer technology and learning led them to the emerging field of the learning sciences.

Each new perspective spurs a reconceptualization of the learning sciences and expands the challenges the learning sciences embrace. A major factor in the evolution of the learning sciences was a focus on the practices that develop to support cultural communities. Anthropologists who initiated in-depth studies of learning in cultural contexts including midwifery, tailoring, candy selling among youthful entrepreneurs, and cooking spurred learning scientists to seek overlooked complexities in more typical foci for studies of learning such as reading and mathematics in schools (Carraher et al., 1985; Saxe, 1985). Furthermore, combining perspectives revealed new dimensions previously ignored. For example, studies of contextually rich learning involving realistic problems delineated the limited ecological validity of laboratory studies in decontextualized settings and highlighted fundamental roles for learning played by the cultural backgrounds of participants (Cole et al., 1982; Cole and Griffin, 1987).

Expanding perspectives can improve the conceptualizations of the problems being addressed while also adding scientific understandings that facilitate progress in related areas. For example, adding computer science to science learning motivated designs for computer tutors that, in turn, captured very nuanced data about student learning trajectories. Efforts to design tutors revealed stark differences between curricular subjects as researchers focused on topics featuring closed systems such as geometry proofs or mechanics where it was relatively possible to analyze student progress and offer guidance. Research teams struggled to create guidance for more open-ended problems.

Appropriating and Developing New Methods

By welcoming new perspectives, the learning sciences also adopted, adapted, created, and refined methods for the learning sciences. Methods as diverse as Piagetian clinical interviews and microgenetic studies of learners reasoning about phenomena in the material world were combined with controlled experiments, frameworks from biology, mathematics, and physics, and computational models of children and adults thinking and reasoning during problem solving. Researchers investigated networked knowledge-building communities where the unit of analysis is a group (Stahl, 2004), or classroom (Scardamalia and Bereiter, 1994), rather than an individual child;

others studied classroom discourse and interactional analysis, leveraging sociolinguistics, educational anthropology, and identity theory, among many other important developments in the field. Recently, researchers have reconsidered computational linguistics/natural language processing, collaborative eye-gaze tracking, motion and emotion sensing, and multimodal learning analytics.

Researchers expanded the nature of experimental studies to include methods for comparing designs for learning environments conducted in complex settings. To investigate the impact of designed environments as they were tested and refined, researchers described what were called design-based research methods (diSessa, 1991; Brown, 1992; Collins, 1992). These early papers were followed by more insights across the emerging community including Cobb et al. (2003) and the network of young scholars represented in the Design-based Research Collective (2003). Design-based research methods were especially well-aligned with the advances in learning technologies that supported a rich assortment of information about student and teacher interactions such as logs of student work and portfolios.

Reconceptualizing Challenges

Broadening the community has expanded and sharpened challenges for the field of the learning sciences (for sampling breadth, consider: Lave and Wenger, 1991; Pea and Gomez, 1992; Bruer, 1993; Anderson, 1996; Koschmann, 1996; Nasir et al., 2006; Linn, 2012; Penuel and Spillane, 2013; Esmonde and Booker, 2016; Niemi et al., 2018). For example, strengthening ties to engineering in the learning sciences has challenged designers to create valid engineering activities for middle school students (Kolodner et al., 2003; Chiu and Linn, 2011).

Creating Artifacts

New perspectives and methods have motivated design or reformulation of artifacts for advancing the learning sciences. Learning artifacts are technologies that augment, transform, and strengthen opportunities to teach, learn, and investigate. These have included *programming languages*: Logo (Papert, 1980, 1991), BOXER (diSessa, 1985; diSessa and Abelson, 1986), AgentSheets (Repenning and Sumner, 1995), NetLogo (Wilensky, 1999), Scratch (Resnick et al., 2009); *learning environments*: CSILE (Scardamalia and Bereiter, 1994), KIE/WISE (Linn et al., 1998, 2003; Linn and Hsi, 2000), BGuILE (Reiser et al., 2001), CoVis/Learning through CoVis (Edelson et al., 1996; Pea et al., 1997; Gomez et al., 1998), SimCalc/MathWorlds (Kaput, 1992; Roschelle and Kaput, 1996), Carnegie Learning's Tutors (Anderson et al., 1995; Koedinger et al., 1997); and *tools*: Geometer's Sketchpad (Jackiw, 1991), Thinkertools (White, 1993).

Developing Abstractions

The learning sciences have created principles, frameworks, theories, and other abstractions to synthesize trends and findings for advancing understanding. These include "Bayesian knowledge tracing" (Corbett and Anderson, 1995), "cognitive apprenticeship" (Collins et al., 1989); "collaborative inquiry learning" (White et al., 1999; Roschelle and Pea, 2002; Kollar et al., 2007; Linn, 2012), "distributed intelligence" (Pea, 1993c), "knowledge-building communities" (Scardamalia and Bereiter, 1994), "scaffolding" (Wood et al., 1976; Pea, 2004); "knowledge integration" (Linn and Eylon, 2011), and many others.

Developing People

The learning sciences have embraced the goal of preparing newcomers in programs for undergraduates, graduate students, early career scholars, and established professionals. The methods and instructional designs emerging in the field have been customized to prepare those interested in the learning sciences. Thus, programs have incorporated apprenticeship models to communicate cultural practices and technologies for collaborative learning. Instructional programs reflect the field's interdisciplinarity and emerging patterns of reasoning, theories, and methods for conducting research studies accountable to the growing learning sciences community.

REFLECTIONS AND NEXT STEPS

The learning sciences will continue to grow and develop as new students join, members of related fields contribute, and individuals recognize new opportunities, create new artifacts, and formulate, test, and refine new abstractions. We look forward to the next new initiatives and deepening insights as the learning journey of the learning sciences continues.

AUTHOR CONTRIBUTIONS

The authors contributed equally, engaging in iterative refinement of the MS across 10 drafts. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by National Science Foundation grants: RP et al.: #DRL-8855582, #RED-9253462, #RED-9454729, # AAT-9453715, #CDA-9616584, #ESI-9720687, #REC-9720423, #CDA-9720384, #ITR-0326497, #SBE-0354453, #SMA/SBE-083585; ML et al.: NSF #1813713, #1451604, #1418423, #1119670, and Hewlett Foundation, POWERED projects.

ACKNOWLEDGMENTS

In addition to the National Science Foundation and Hewlett, ML thanks the University of California, Berkeley, Graduate School of Education. RP also thanks Spencer Foundation, and Apple's Advanced Technology Group and Apple Classroom of Tomorrow Project.

REFERENCES

Anderson, J. R. (1996). *The Architecture of Cognition*. Mahwah, NJ: Erlbaum.

Anderson, J. R., Corbett, A. T., Koedinger, K., and Pelletier, R. (1995). Cognitive tutors: lessons learned. *J. Learn. Sci.* 4, 167–207. doi: 10.1207/s15327809jls0402_2

Bell, P., Davis, E. A., and Linn, M. C. (1995). "The Knowledge Integration Environment: theory and design," in *Proceedings of the Computer Supported Collaborative Learning Conference (CSCL '95: Bloomington, IN)*, (Mahwah, NJ: Erlbaum), 14–21.

Brown, A. L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions. *J. Learn. Sci.* 2, 141–178. doi: 10.1207/s15327809jls0202_2

Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educ. Res.* 18, 32–42. doi: 10.3102/0013189x018001032

Bruer, J. (1993). *Schools for Thought*. Cambridge, MA: MIT Press.

Carraher, T. N., Carraher, D., and Schliemann, A. D. (1985). Mathematics in the streets and in schools. *Br. J. Dev. Psychol.* 3, 21–29. doi: 10.1111/j.2044-835x.1985.tb00951.x

Chiu, J., and Linn, M. C. (2011). Knowledge Integration and WISE Engineering. *J. Pre Coll. Eng. Educ.* 1, 1–14.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., and Schauble, L. (2003). Design experiments in educational research. *Educ. Res.* 32, 9–13. doi: 10.3102/0013189x032001009

Cole, M., and Griffin, P. (1987). *Contextual Factors in Education: Improving Science and Mathematics Education for Minorities and Women. Prepared for the National Research Council*. Madison, WI: University of Wisconsin.

Cole, M., Hood, L., and McDermott, R. (1982). "Ecological niche picking," in *Memory Observed: Remembering in Natural Context*, ed. U. Neisser (San Francisco: Freeman).

Collins, A. (1992). "Toward a design science of education," in *New Directions in Educational Technology*, eds E. Scanlon, and T. O'Shea (Berlin: Springer-Verlag).

Collins, A., Brown, J. S., and Newman, S. E. (1989). "Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics," in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, ed. L. B. Resnick (Hillsdale, NJ: Erlbaum), 453–494. doi: 10.4324/9781315044408-14

Corbett, A. T., and Anderson, J. R. (1995). Knowledge tracing: modeling the acquisition of procedural knowledge. *User Model. User Adapt. Interact.* 4, 253–278. doi: 10.1007/bf01099821

De Corte, E., Linn, M. C., Mandl, H., and Verschaffel, L. (eds) (1992). *Computer-Based Learning Environments and Problem Solving (NATO ASI Series F: Computer and System Series)*, Vol. 84. Berlin: Springer-Verlag.

Design-based Research Collective (2003). Design-based research: an emerging paradigm for educational inquiry. *Educ. Res.* 32, 35–37.

diSessa, A. (1991). "Local sciences: viewing the design of human-computer systems as cognitive science," in *Designing Interaction: Psychology at the Human-Computer Interface*, ed. J. M. Carroll (Cambridge, MA: Cambridge University Press), 162–202.

diSessa, A. A. (1985). A principled design for an integrated computational environment. *Hum. Comput. Interact.* 1, 1–47. doi: 10.1207/s15327051hci0101_1

diSessa, A. A., and Abelson, H. (1986). Boxer: a reconstructible computational medium. *Commun. ACM* 29, 859–868. doi: 10.1145/6592.6595

Edelson, D. C., Pea, R. D., and Gomez, L. (1996). "Constructivism in the collaboratory," in *Constructivist Learning Environments: Case Studies in Instructional Design*, ed. B. G. Wilson (Englewood Cliffs, NJ: Educational Technology Publications), 151–164.

Esmonde, I., and Booker, A. N. (2016). *Power and Privilege in the Learning Sciences*. New York, NY: Routledge.

Gomez, L. M., Fishman, B. J., and Pea, R. D. (1998). The CoVis project: building a large-scale science education testbed. *Interact. Learn. Environ.* 6, 59–92. doi: 10.1076/ilee.6.1.59.3608

Greeno, J. G., and The Middle School Mathematics Through Applications Project Group [MMAP] (1998). The situativity of knowing, learning, and research. *Am. Psychol.* 53:5. doi: 10.1037/0003-066x.53.1.5

Jackiw, N. (1991). *The Geometer's Sketchpad [Computer Software]*. Berkeley, CA: Key Curriculum.

Kali, Y. (2006). Collaborative knowledge-building using the Design Principles Database. *Int. J. Comput. Support Collab. Learn.* 1, 187–201. doi: 10.1007/s11412-006-8993-x

Kaput, J. (1992). "Technology and mathematics education," in *A Handbook of Research on Mathematics Teaching and Learning*, ed. D. Grouws (New York, NY: MacMillan), 515–556.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *Int. J. Artif. Intellig. Educ.* 8, 30–43.

Kollar, I., Fischer, F., and Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learn. Instr.* 17, 708–721. doi: 10.1016/j.learninstruc.2007.09.021

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: putting Learning by Design into practice. *J. Learn. Sci.* 12, 495–547. doi: 10.1207/s15327809jls1204_2

Koschmann, T. (1996). *CSCL: Theory and Practice of an Emerging Paradigm*. Mahwah, NJ: Erlbaum.

Lave, J. (1988). *Cognition in Practice*. Cambridge: Cambridge University Press.

Lave, J., and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Linn, M. C. (1987). Establishing a research base for science education: challenges, trends, and recommendations [Reprinted from National Science Foundation report of a conference held in 1986, January]. *J. Res. Sci. Teach.* 24, 191–216. doi: 10.1002/tea.3660240302

Linn, M. C. (1996). "From separation to partnership in science education: Students, laboratories, and the curriculum," in *Microcomputer Based Labs: Educational Research and Standards*, Vol. 156, ed. R. F. Tinker (Berlin: Springer-Verlag), 13–46.

Linn, M. C. (2012). "Insights for teaching and learning science," in *Digital Teaching Platforms: Customizing Classroom Learning for Each Student*, eds C. Dede and J. Richards (New York, NY: Teachers College Press), 55–70.

Linn, M. C., Bell, P., and Hsi, S. (1998). Using the internet to enhance student understanding of science: the knowledge integration environment. *Int. Learn. Environ.* 1, 4–38. doi: 10.1076/ilee.6.1.4.3606

Linn, M. C., Clark, D., and Slotta, J. D. (2003). WISE design for knowledge integration. *Sci. Educ.* 87, 517–538. doi: 10.1002/sce.10086

Linn, M. C., Davis, E. A., and Bell, P. (eds) (2004). *Internet Environments for Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., and Eylon, B.-S. (2011). *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration*. New York: Routledge.

Linn, M. C., and Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mandinach, E. B., Linn, M. C., Pea, R., and Kurland, M. (1986). Cognitive effects of computer learning environments [Special Issue]. *J. Educ. Comput. Res.* 3, 409–410.

Means, B., Pea, R. D., Gordin, D., and Roschelle, J. (1999). "The emerging fabric of distributed learning systems," in *Proceedings of the Vision paper for the National Governors' Assocation (NGA) Meetings*, Virginia.

Nasir, N. I. S., Rosebery, A. S., Warren, B., and Lee, C. D. (2006). "Learning as a cultural process: achieving equity through diversity," in *The Cambridge Handbook of the Learning Sciences*, ed. K. Sawyer (Cambridge: Cambridge University Press), 489–504. doi: 10.1017/cbo9780511816833.030

Niemi, D., Pea, R., Saxberg, B., and Clark, R. E. (2018). *Learning Analytics in Education*. Charlotte, NC: Information Age Publishing.

O'Malley, C. (1994). *Computer Supported Collaborative Learning. NATO ASI Subseries F, Computer and Systems Sciences, 128*. Heidelberg: Springer-Verlag.

Papert, S. (1980). *Mindstorms*. New York, NY: Basic Books.

Papert, S. (1991). "Situating constructionism," in *Constructionism*, eds I. Harel and S. Papert (Norwood, NJ: Ablex), 1–11.

Pea, R. (2016). "The prehistory of the learning sciences," in *The Learning Sciences: Past, Present, and Future*, eds M. A. Evans, M. J. Packer, and R. K. Sawyers (New York, NY: Cambridge University Press), 32–58. doi: 10.1017/cbo9781107707221.003

Pea, R. D. (1987). Socializing the knowledge transfer problem. *Int. J. Educ. Res.* 11, 639–663. doi: 10.1016/0883-0355(87)90007-3

Pea, R. D. (1993a). "Current developments in educational technology programs: doctoral training in the learning sciences at Northwestern University," in *Educational Media and Technology Handbook*, eds D. P. Ely and B. M. Minor (Englewood, CO: Libraries Unlimited).

Pea, R. D. (1993b). Learning scientific concepts through material and social activities: conversational analysis meets conceptual change. *Educ. Psychol.* 28, 265–277. doi: 10.1207/s15326985ep2803_6

Pea, R. D. (1993c). "Practices of distributed intelligence and designs for education," in *Distributed Cognitions*, ed. G. Salomon (New York, NY: Cambridge University Press), 47–87.

Pea, R. D. (1993d). The collaborative visualization project. *Commun. ACM* 36, 60–63. doi: 10.1145/155049.155063

Pea, R. D. (1994). Seeing what we build together: distributed multimedia learning environments for transformative communications. *J. Learn. Sci.* 3, 285–299. doi: 10.1207/s15327809jls0303_4

Pea, R. D. (2004). The social and technological dimensions of "scaffolding" and related theoretical concepts for learning, education and human activity. *J. Learn. Sci.* 13, 423–451. doi: 10.1207/s15327809jls1303_6

Pea, R. D., and Gomez, L. (1992). Distributed multimedia learning environments: why and how? *Interact. Learn. Environ.* 2, 73–109. doi: 10.1080/10494829 20020201

Pea, R. D., Gomez, L. M., Edelson, D. C., Fishman, B. J., Gordin, D. N., and O'Neill, D. K. (1997). "Science education as a driver of cyberspace technology development," in *Internet Links for Science Education*, ed. K. C. Cohen (New York, NY: Plenum Press), 189–220. doi: 10.1007/978-1-4615-5909-2_12

Pea, R. D., Tinker, R., Linn, M. C., Means, B., Bransford, J., Roschelle, J., et al. (1999). Toward a learning technologies knowledge network. *Educ. Technol. Res. Dev.* 47, 19–38.

Penuel, W., and Spillane, J. P. (2013). "Learning sciences and policy design and implementation: key concepts and tools for collaborative engagement," in *Cambridge Handbook of the Learning Sciences*, 2nd Edn, ed. K. Sawyer (New York, NY: Cambridge University Press).

Ramamurthy, M., Wilhelmson, R., Hall, S., Plutchak, J., Sridhar, M., Fishman, B., et al. (1996). "CoVis geosciences web server: an internet-based resource for the K-12 community," in *Proceedings of the 76th American Meteorological Society Meetings (Proceedings of the Fifth Symposium on Education)*, (Atlanta, GA: American Meteorological Society).

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., and Leone, T. J. (2001). "BGuILE: strategic and conceptual scaffolds for scientific inquiry in biology classrooms," in *Cognition and Instruction: Twenty-five years of progress*, eds S. M. Carver and D. Klahr (Mahwah, NJ: Erlbaum).

Repenning, A., and Sumner, T. (1995). Agentsheets: a medium for creating domain-oriented visual languages. *IEEE Comput.* 28, 17–25. doi: 10.1109/2. 366152

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., et al. (2009). Scratch: programming for all. *Commun. ACM* 52, 60–67.

Roschelle, J., and Kaput, J. (1996). SimCalc mathworlds for the mathematics of change. *Commun. ACM* 39, 97–99. doi: 10.1145/232014. 232041

Roschelle, J., and Pea, R. (2002). A walk on the WILD side: how wireless handhelds may change computer-supported collaborative learning. *Int. J. Cogn. Technol.* 1, 145–168. doi: 10.1075/ijct.1.1.09ros

Sabelli, N., and Pea, R. (eds) (2004). *Center for Innovative Learning Technologies (CILT): Six Years of Knowledge Networking in Learning Sciences and Technologies*. Menlo Park, CA: SRI International.

Saxe, G. B. (1985). The effects of schooling on arithmetical understandings: studies with Oksapmin children in Papua New Guinea. *J. Educ. Psychol.* 77, 503–513. doi: 10.1037/0022-0663.77.5.503

Scardamalia, M., and Bereiter, C. (1994). Computer support for knowledge-building communities. *J. Learn. Sci.* 3, 265–283.

Schank, R. C. (2016). "Why learning sciences?," in *The Learning Sciences: Past, Present, and Future*, eds M. A. Evans, M. J. Packer, and R. K. Sawyers (New York: Cambridge University Press), 19–31. doi: 10.1017/cbo978110770 7221.002

Stahl, G. (2004). "Building collaborative knowing," in *What We Know About CSCL and Implementing It In higher Education*, eds G. Stahl, J. W. Strijbos, P. A. Kirschner, and R. L. Martens (Dordrecht: Springer), 53–85. doi: 10.1007/1-4020-7921-4_3

Tinker, R. F. (1996). *Microcomputer Based Labs: Educational Research and Standards (NATO*, Vol. 156. Berlin: Springer-Verlag.

White, B. (1993). Thinkertools: casual models, conceptual change, and science education. *Cognition Instr.* 10, 1–100. doi: 10.1207/s1532690xci1001_1

White, B. Y., Shimoda, T. A., and Frederiksen, J. R. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: computer support for metacognitive development. *Int. J. Artif. Intellig. Educ.* 10, 151–182.

Wilensky, U. (1999). *NetLogo (and NetLogo User Manual), Center for Connected Learning and Computer-Based Modeling*. Evanston, IL: Northwestern University.

Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *J. Child Psychol. Psychiatry* 17, 89–100. doi: 10.1111/j.1469-7610.1976. tb00381.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

MINI REVIEW
published: 17 July 2020
doi: 10.3389/feduc.2020.00128

Defining the Boundaries Between Artificial Intelligence in Education, Computer-Supported Collaborative Learning, Educational Data Mining, and Learning Analytics: A Need for Coherence

Bart Rienties[1], Henrik Køhler Simonsen[2] and Christothea Herodotou[1]*

[1] Institute of Educational Technology, The Open University, Milton Keynes, United Kingdom, [2] Smartlearning, Copenhagen, Denmark

Edited by:
Matthias Stadler,
Ludwig Maximilian University of Munich, Germany

Reviewed by:
Judit García-Martín,
University of Salamanca, Spain
Jessica Levy,
University of Luxembourg, Luxembourg

****Correspondence:***
Bart Rienties
bart.rienties@open.ac.uk

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *17 April 2020*
Accepted: *29 June 2020*
Published: *17 July 2020*

Citation:
Rienties B, Køhler Simonsen H and Herodotou C (2020) Defining the Boundaries Between Artificial Intelligence in Education, Computer-Supported Collaborative Learning, Educational Data Mining, and Learning Analytics: A Need for Coherence. Front. Educ. 5:128. doi: 10.3389/feduc.2020.00128

This review aims to provide a concise overview of four distinct research fields: Artificial Intelligence and EDucation (AIED), Computer-Supported Collaborative Learning (CSCL), Educational Data Mining (EDM), and Learning Analytics (LA). While all four fields are focused on understanding learning and teaching using technology, each field has a relatively unique or common perspective on which theoretical frameworks, methods, and ontologies might be appropriate. In this review we argue that researchers should be encouraged to cross the boundaries of their respective field and work together to address the complex challenges in education.

Keywords: artificial intelligence in education, computer-supported collaborative learning, educational data mining, learning analytics, review

INTRODUCTION

In the last 20 years a range of disciplines have been developed in the broad field of education and technology. Since the early 1980s the broad field of Artificial Intelligence and EDucation (AIED) emerged that aimed to use a combination of Artificial Intelligence (AI), learning theory, and educational practice to improve learning outcomes for learners using computers (Boyd et al., 1982; Holmes et al., 2019). Within AIED various subfields of research emerged based upon the power of computing and machine learning, such as intelligent tutoring systems (Aleven and Koedinger, 2002), adaptive hypertext systems (Eysink et al., 2009; Romero et al., 2009), and Computer-Supported Collaborative Learning (CSCL). Since the early 1990s a range of CSCL publications appeared exploring how learners and teachers could work together online using computers. A vast number of CSCL studies (e.g.,Gunawardena, 1995; Roschelle and Koschmann, 1996; Fischer and Mandl, 2005; Rienties et al., 2009) have found that scaffolding, self-regulation, task design, and teaching presence are important concepts that can encourage learners to effectively work together.

In the mid-2000s a third stream of researchers (e.g., Baker and Yacef, 2009; Rosé et al., 2014) using Educational Data Mining (EDM) started to explore learning processes using bigger data sets and increased interconnections between data. Since 2011 a fourth research field of Learning Analytics (LA) emerged, which is specifically focused on understanding the complex learning processes and learning outputs, using a multi-disciplinary combination of computer-science, educational psychology, engineering, and learning sciences (Ferguson, 2012; Papamitsiou and Economides, 2014). In this contribution we aim to define what the potential boundaries and synergies are between AIED, CSCL, EDM, and LA, and how a combined interdisciplinary perspective can help to maximize the potential of these four research fields to understand the complexities of learning and teaching using technology. This might be particularly relevant for researchers and practitioners who may be new to these research fields. For a more detailed and deeper analysis of these fields, we encourage readers to connect to the respective journals in **Table 1**.

FOUR PERSPECTIVES ON COMPUTING, LEARNING, AND EDUCATION

The boundaries between AIED, CSCL, EDM, and LA are rather blurred. In part, this is because researchers and practitioners from these respective fields look at similar, yet slightly distinct phenomena, and in part, this is because researchers often work in interdisciplinary research groups across the boundaries of their specific research focus (Jeong et al., 2014; Aldowah et al., 2019; Dormezil et al., 2019). Therefore, the characterisations of the four research fields below are by definition an oversimplification of their complex, inter-linked, and fluid perspectives, relations, methodologies, and ontologies. Given that these fields emerged, faded, merged, and re-emerged at various points of time, rather than giving a historical overview of these fields, we will describe these fields in alphabetical order and in relation to the following aspects (see **Table 1**): (a) main aim/target, (b) educational and other underpinnings, (c) techniques and approaches, (d) society, and (e) conferences and journals.

Artificial Intelligence in Education

Although there is not a single definition of what AI might be, AI broadly refers to "computers which perform cognitive tasks, usually associated with human minds, particularly learning, and problem-solving" (Baker et al., 2019, p. 10). It is an umbrella term used to describe several methods such as machine learning, data mining (DM), neural networks or an algorithm (Zawacki-Richter et al., 2019). Its roots can be traced back to computer science and engineering, with a strong relation to economics, cognitive science, philosophy, and neuroscience (Popenici and Kerr, 2017; Holmes et al., 2019; Zawacki-Richter et al., 2019). As indicated in **Table 1**, the main aim of AIED is to simulate and predict learning processes. In terms of philosophical underpinning, a crucial underlying assumption of AI, and AIED in particular, is that any aspect of learning or any other feature of intelligence can be described, and that a machine is able to simulate it (Zawacki-Richter et al., 2019). In the last 20 years, substantial progress has been made in machine learning, which allows researchers to understand, model and simulate the complex behaviors of humans, which are assumed to be rational. Popenici and Kerr (2017, p. 2) defined machine learning "as a subfield of artificial intelligence that includes software able to recognize patterns, make predictions, and apply newly discovered patterns to situations that were not included or covered by their initial design." With the incredible advances of AI in other sectors (e.g., automobile, health care, manufacturing), recently there has been a renewed interest in AIED (Tuomi, 2018; Zawacki-Richter et al., 2019).

For example, in a review of 146 studies conducted between 2007 and 2018 (Zawacki-Richter et al., 2019) a range of applications of AI in higher education were identified, including making admission decisions and course scheduling (Andris et al., 2013), assessment and feedback (Adamson et al., 2014), intelligence tutoring systems (Aleven and Koedinger, 2002), profile and prediction of students dropping out (Rizvi et al., 2019), and student models and academic achievement (Rizvi et al., 2019). As identified by Zawacki-Richter et al. (2019), although substantial progress has been made in AIED, most studies are quantitative in nature, make use of human intervention studies (Blanchard, 2012), with a control and experimental group, lack reflection on risks, challenges and ethical implications, and present a weak connection to relevant educational theories.

Computer-Supported Collaborative Learning

A main aim of CSCL is to understand the complex interactions in and outside class settings. While AIED assumes that all learning can be described and simulated by machines, in CSCL literature there is often a recognition that learning is complex, and socially constructed. McKeown et al. (2017, p. 439) argued that "(r)esearch in CSCL focuses on learning as a cognitive and/or social process and studies learning designs, learning processes, and pedagogic practices that support technology-mediated collaborative processes in communities of practice." Given its focus on people working together, there are complex and dynamic interactions that may, or may not, be easily identifiable by computers (e.g., body language, cultural differences, emotions, linguistic styles). In order to develop and maintain a successful CSCL culture, Jeong et al. (2014) theorized that technology used for collaboration in CSCL needs to include: (1) a joint task, (2) communication, (3) sharing of resources, (4) engagement in productive processes, (5) engagement in co-construction, (6) monitoring and regulation, and (7) finding and building groups and communities. In face-to-face and blended learning scenarios, this maintenance of successful discourse might be difficult to achieve, while in online settings there is a wealth of research showing complexities in online collaboration (Fischer and Mandl, 2005; Rienties et al., 2009). For example, in a review of 180 articles published in CSCL conferences in the period 2005–2017, Xia and Borge (2019) found that most studies focused on interaction in classrooms (47%), technology implemented in classrooms (13%), technology implemented in informal settings (15%), and in labs (11%). This strong focus on in-class analysis seems substantially different to AIED. Furthermore, CSCL seems to have strong experimental and learning science roots (Wise and Schwarz, 2017), whereby approximately half of recent studies identified by Jeong et al. (2014) used a methodologically strong

TABLE 1 | Overview of the four research fields of education and technology.

	AIED	CSCL	EDM	LA
(A) Main aim/target	Simulate and predict learning processes	Understand learning processes in/outside classroom settings	Analyze data from educational systems	Improve learning processes
(B) Educational, theoretical, and philosophical underpinning	Any form of learning can be described and machines are able to simulate these processes. Learners are rational. Educationally/pedagogically neutral	Focused on collaboration and interaction between two or more people. Communication theories, social constructivist, sociocultural, social psychology	Neutral	A range of pedagogical theories used, including connected learning, self-regulated learning, socio-constructivist.
(C) Techniques and approaches	Machine learning, human intervention studies	Discourse analysis, content analysis, questionnaires, social network analysis.	Computational modeling (human–computer interaction, machine learning, AI), data mining, psychometrics statistics, visualization	Discourse analysis, natural language processing, machine learning, predictive modeling, qualitative research methods, social network analysis, visualization.
(D) Society	The International AIED Society (1997)	International Society of the Learning Sciences (2002)	International Educational Data Mining Society (2008)	Society for Learning Analytics Research (2011)
(E) Main conference and journal	AIED conference International Journal of AI in Education (no IF)	CSCL conference International Journal of Computer Supported Collaborative Learning (IF: 2.206)	Educational Data Mining conference International Journal of Educational Data Mining (no IF)	LAK conference Journal of Learning Analytics (no IF)

design. At the same time, several meta reviews indicated a need for CSCL researchers to embrace more analytics and multi-level approaches to extend their methodological toolbox as well as the rigor of their studies beyond a single classroom or context (Jeong et al., 2014; Wise and Schwarz, 2017; Xia and Borge, 2019).

Educational Data Mining

The main aim of EDM could be succinctly described as analyzing data from educational systems. With the rise of educational data, EDM has been going from strength to strength (Koedinger et al., 2015; Dutt et al., 2017; Aldowah et al., 2019). Early literature reviews (Romero and Ventura, 2007, 2010) noted the need for considering pedagogical aspects when mining data from educational systems, and identified benefits for students and teachers when recommender systems are used. Building on the first EDM conference in 2008, EDM has been defined (Baker and Yacef, 2009) as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." By using a range of DM techniques, EDM researchers aim to discover novel and potentially useful information from large amounts of data. As argued by a range of EDM researchers, while DM techniques are useful in big data contexts, in education there is a need to adjust algorithms to specific contexts (Dutt et al., 2017). Koedinger et al. (2015) explained that EDM focuses on a range of research questions in the psychology of learning: (a) assessment of cognition and learning, (b) transfer of learning, and discovery of cognitive models, (c) affect, motivation, and metacognition (Rosé et al., 2014), and (d) language and discourse analytics.

A desirable sequence of EDM research is to start off with DM leading to new statistical models of data, followed by building an (adaptive) automated system, and finally, closing the loop, by running an evidence-based experiment (Koedinger et al., 2015). In a review of 166 EDM studies, Dutt et al. (2017) identified five common clusters of studies: (1) analyzing student motivation, attitude and behavior; (2) understanding learning style; (3) e-learning; (4) collaborative learning; (5) EDM using clustering. A particular notable distinction between EDM, CSCL, and LA is the lack of specific reliance on educational theory. Most EDM research is considered pedagogically and educational theory-neutral, as the focus is on data discovery, testing of interventions, and optimizing models.

Learning Analytics

The Journal of Learning Analytics defines LA as ". . . research into the challenges of collecting, analyzing, and reporting data with the specific intent to improve learning." We define the main aim of LA as to improve learning processes. Several higher education institutions and distance learning providers have started to explore the use of LA dashboards that can display learner and learning behavior to teachers and instructional designers in order to provide more real-time or just-in-time support to students (Jivet et al., 2018; Herodotou et al., 2020). Furthermore, several institutions have developed predictive LA approaches to help identify, as early as possible, students who may be considered "at risk" of failing, and which of those students may need additional support (Viberg et al., 2018; Herodotou et al., 2020). Some institutions are also currently experimenting with providing LA data directly to students in order to support their learning processes and self-regulation (Winne, 2017; Rienties et al., 2019).

As argued by a range of authors, the distinction between EDM and LA is rather unclear, as leading researchers from both fields contribute to similar themes and debates across the two fields (Aldowah et al., 2019; Dormezil et al., 2019). According to Papamitsiou and Economides (2014), both EDM and LA communities share compatible goals and focus where learning science and data-driven analytics intersect. However, there are some subtle and more explicit differences in their ontological origins, techniques used, and perhaps most importantly the specific topics of interest. As argued by Papamitsiou and Economides (2014, p. 50) "LA adopts a holistic framework, seeking to understand systems in their full complexity. On the other hand, EDM adopts a reductionistic viewpoint by analyzing individual components, seeking for new patterns in data and modifying respective algorithms." In a review contrasting 1,952 LA articles with 783 EDM articles by Dormezil et al. (2019), several common themes were identified, such as "educational computing" and "student performance." LA focuses mostly on instruction and communication, student learning objectives and natural language processing. In contrast, EDM is focused on student performance and the technical specifications of respective predictive approaches, in particular "learning algorithms" and "student models." Nonetheless, there is more common overlap than distinct differences; Dormezil et al. (2019) argued that LA is probably best described as one domain with one prominent subset, that of EDM.

DISCUSSION

This review has briefly explored the intersection between education and technology in four fields: AIED, CSCL, EDM, and LA. In the last decade tremendous progress has been made to better understand the complexities of learning and teaching with technology. With the rise and availability of big data in education and AI, substantial leaps in the conceptual, theoretical, and evidence-based understanding of learning and teaching have been made in the four fields discussed. However, as highlighted by a range of reviews, most of these innovations have been localized in small lab studies, or in a single course, or specific context, with limited large-scale adoption within and across institutions (Viberg et al., 2018; Herodotou et al., 2020).

In order to truly make substantial leaps in the actual adoption of technology in large educational settings, achieve wide-spread uptake in educational institutions, and improve our understanding of the complexities of learning that can advance our theoretical models, we argue that the four research fields need to break down some of the artificial barriers between the respective communities, and jointly work together as one interdisciplinary research field. This can be achieved via a web of inter-related activities. First of all, national and international funding bodies should explicitly embrace and fund interdisciplinary research that cuts across the four (and other) fields. Second, by building cross-disciplinary network opportunities for researchers to learn from different disciplines might help to cross-fertilize and cross-pollinate different research ideas, methods and approaches. This can be "formally" achieved by including specific tracks in conference programs, joined special issues, and running some events together, as well as informally by encouraging research visits and invited seminars. Third, as highlighted in **Table 1**, there are substantial synergies that are possible in terms of theoretical, empirical and methodological advancement between the four fields. We argue that by bringing the best research minds together across the four fields, substantial progress can be made to address some of the large challenges in education and society at large. Toward this direction, in the last few years we have seen several initiatives that attempt to bring those fields closer, including the Festival of Learning and the creation of the International Alliance to Advance Learning in the Digital Era[1] that brings the various societies included in **Table 1** together. In terms of next steps following this work, and given the short-length nature of this article, a systematic and exhaustive review across the four fields would be particularly beneficial and help establish how exactly these fields differ and overlap.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

This article had received funding from the Horizon 2020 Research and Innovation Program ERASMUS+ (KA203-2019-002).

[1] http://www.alliancelss.com/

REFERENCES

Adamson, D., Dyke, G., Jang, H., and Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *Int. J. Artif. Intel. Educ.* 24, 92–124. doi: 10.1007/s40593-013-0012-6

Aldowah, H., Al-Samarraie, H., and Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: a review and synthesis. *Telemat. Inform.* 37, 13–49. doi: 10.1016/j.tele.2019.01.007

Aleven, V. A. W. M. M., and Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cogn. Sci.* 26, 147–179. doi: 10.1207/s15516709cog2602_1

Andris, C., Cowen, D., and Wittenbach, J. (2013). support vector machine for spatial variation. *Transact. GIS* 17, 41–61. doi: 10.1111/j.1467-9671.2012.01354.x

Baker, R. S., and Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *JEDM J. Educ. Data Min.* 1, 3–17. doi: 10.5281/zenodo.3554657

Baker, T., Smith, L., and Anissa, N. (2019). *Educ-AI-Tion Rebooted? Exploring the Future of Artificial Intelligence in Schools and Colleges* (London: Nesta). Available online at: https://www.nesta.org.uk/report/education-rebooted

Blanchard, E. G. (2012). "Intelligent tutoring systems," in *On the WEIRD Nature of ITS/AIED Conferences*, eds S. A. Cerri, W. J. Clancey, G. Papadourakis, and K. Panourgia (Berlin).

Boyd, G., Keller, A., and Kenner, R. (1982). Remedial and second language English teaching using computer assisted learning. *Comput. Educ.* 6, 105–112. doi: 10.1016/B978-0-08-028111-7.50020-9

Dormezil, S., Khoshgoftaar, T., and Robinson-Bryant, F. (2019). "Differentiating between educational data mining and learning analytics: a bibliometric approach," in *LABBEC Workshop (Learning Analytics: Building Bridges Between the Education and the Computing Communities)*, 1–6.

Dutt, A., Ismail, M. A., and Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access* 5, 15991–16005. doi: 10.1109/ACCESS.2017.2654247

Eysink, T. H. S., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., Wouters, P., et al. (2009). Learner performance in multimedia learning arrangements: an analysis across instructional approaches. *Ame. Educ. Res. J.* 46, 1107–1149. doi: 10.3102/0002831209340235

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *Int. J. Technol. Enhan. Learn.* 4, 304–317. doi: 10.1504/ijtel.2012.051816

Fischer, F., and Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: the role of external representation tools. *J. Learn. Sci.* 14, 405–441. doi: 10.1207/s15327809jls1403_3

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferencing. *Int. J. Educ. Telecommun.* 1, 147–166.

Herodotou, C., Rienties, B., Hlosta, M., Boroowa, A., Mangafa, C., and Zdrahal, Z. (2020). The scalable implementation of predictive learning analytics at a distance learning university: insights from a longitudinal case study. *Internet High. Educ.* 45:100725. doi: 10.1016/j.iheduc.2020.100725

Holmes, W., Bialik, M., and Fadel, C. (2019). *Artificial Intelligence In Education: Promises and Implications for Teaching and Learning*. Boston, MA: Center for Curriculum Redesign.

Jeong, H., Hmelo-Silver, C. E., and Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *Int. J. Comput. Supp. Coll. Learn.* 9, 305–334. doi: 10.1007/s11412-014-9198-3

Jivet, I., Scheffel, M., Specht, M., and Drachsler, H. (2018). "License to evaluate: preparing learning analytics dashboards for educational practice," in *Proceedings of the 8th International Conference on Learning Analytics & Knowledge (LAK'18)*, Sydney.

Koedinger, K., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., and Rosé, C. P. (2015). Data mining and education. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 333–353. doi: 10.1002/wcs.1350

McKeown, J., Hmelo-Silver, C. E., Jeong, H., Hartley, K., Faulkner, R., Emmanuel, N., et al. (2017). "A meta-synthesis of CSCL literature in STEM education," in *Computer Supported Collaborative Learning*, Philadelphia, PA.

Papamitsiou, Z., and Economides, A. (2014). Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Educ. Technol. Soc.* 17, 49–64.

Popenici, S. A. D., and Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res. Pract. Technol. Enhan. Learn.* 12:22. doi: 10.1186/s41039-017-0062-8

Rienties, B., Tempelaar, D. T., Nguyen, Q., and Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Comput. Human Behav.* 100, 345–357. doi: 10.1016/j.chb.2019.07.007

Rienties, B., Tempelaar, D. T., Van Den Bossche, P., Gijselaers, W. H., and Segers, M. (2009). The role of academic motivation in computer-supported collaborative learning. *Comput. Human Behav.* 25, 1195–1206. doi: 10.1016/j.chb.2009.05.012

Rizvi, S., Rienties, B., and Khoja, S. A. (2019). The role of demographics in online learning; a decision tree based approach. *Comput. Edu.* 137, 32–47. doi: 10.1016/j.compedu.2019.04.001

Romero, C., and Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert Syst. Appl.* 33, 135–146. doi: 10.1016/j.eswa.2006.04.005

Romero, C., and Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Trans. Syst. ManCybernet. Part C* 40, 601–618. doi: 10.1109/TSMCC.2010.2053532

Romero, C., Ventura, S., Zafra, A., and Bra, P. D. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. *Comput. Educ.* 53, 828–840. doi: 10.1016/j.compedu.2009.05.003

Roschelle, J., and Koschmann, T. (1996). "Learning by collaborating: convergent conceptual change," in *CSCL: Theory and Practice of an Emerging Paradigm*, ed. T. Koschmann (New Jersey: Lawrence Erlbaum Associates, Inc), 209–248.

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., et al. (2014). "*Social factors that contribute to attrition in MOOCs*," in *Proceedings of the first ACM Conference on Learning@scale Conference*, Atlanta, GA.

Tuomi, I. (2018). "The impact of artificial intelligence on learning, teaching, and education," in *Policies for the Future*, eds M. Cabrera, R. Vuorikari, and Y. Punie (Luxembourg: Publications Office of the European Union). Available online at: http://repositorio.minedu.gob.pe/bitstream/handle/MINEDU/6021/The%20Impact%20of%20Artificial%20Intelligence%20on%20Learning,%20Teaching,%20and%20Education.pdf?sequence=1

Viberg, O., Hatakka, M., Bälter, O., and Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Comput. Hum. Behav.* 89, 98–110. doi: 10.1016/j.chb.2018.07.027

Winne, P. H. (2017). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teach. Coll. Record* 119, 1–24.

Wise, A. F., and Schwarz, B. B. (2017). Visions of CSCL: eight provocations for the future of the field. *Int. J. Comput. Supp. Coll. Learn.* 12, 423–467. doi: 10.1007/s11412-017-9267-5

Xia, Y., and Borge, M. (2019). "A systematic review of the quantification of qualitative data in proceedings of international conferences on CSCL from 2005 to 2017," in *13th International Conference on Computer Supported Collaborative Learning - A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, CSCL 2019*, eds K. Lund, G. P. Niccolai, E. Lavoue, C. Hmelo-Silver, G. Gweon, M. Baker et al. Lyon.

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int. J. Educ. Technol. Higher Edu.* 16:39. doi: 10.1186/s41239-019-0171-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

METHODS
published: 17 July 2020
doi: 10.3389/feduc.2020.00097

Check for updates

Statistical Methods in Transdisciplinary Educational Research

Alfred Lindl[1]*, *Stefan Krauss*[2], *Anita Schilcher*[3] and *Sven Hilbert*[1]

[1] *Faculty of Human Sciences, University of Regensburg, Regensburg, Germany,* [2] *Faculty of Mathematics, University of Regensburg, Regensburg, Germany,* [3] *Faculty of Languages, Literature, and Culture, University of Regensburg, Regensburg, Germany*

A central task of educational research is to examine common issues of teaching and learning in all subjects taught at school. At the same time, the focus is on identifying and investigating unique subject-specific aspects on the one hand and transdisciplinary, generalizable effects on the other. This poses various methodological challenges for educational researchers, including in particular the aggregation and evaluation of already published study effects, hierarchical data structures, measurement errors, and comprehensive data sets with a large number of potentially relevant variables. In order to adequately deal with these challenges, this paper presents the core concepts of four methodological approaches that are suitable for the analysis of transdisciplinary research questions: meta-analysis, multilevel models, latent multilevel structural equation models, and machine learning methods. Each of these approaches is briefly illustrated with an example inspired by the interdisciplinary research project FALKE (subject-specific teacher competencies in explaining). The data and analysis code used are available online at https://osf.io/5sn9j. Finally, the described methods are compared, and some application hints are given.

Keywords: transdisciplinarity, meta-analysis, multilevel model, linear mixed model, structural equation model, machine learning, explaining, instructional quality

OPEN ACCESS

Edited by:
Matthias Stadler, Ludwig Maximilian University of Munich, Germany

Reviewed by:
Antoine Fischbach, University of Luxembourg, Luxembourg
Jan Dörendahl, University of Luxembourg, Luxembourg

****Correspondence:***
Alfred Lindl
alfred.lindl@ur.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *22 April 2020*
Accepted: *28 May 2020*
Published: *17 July 2020*

Citation:
Lindl A, Krauss S, Schilcher A and Hilbert S (2020) Statistical Methods in Transdisciplinary Educational Research. Front. Educ. 5:97. doi: 10.3389/feduc.2020.00097

INTRODUCTION

Interdisciplinarity is a key feature of empirical educational research. However, while this defining characteristic was for a long time primarily related to the participation and cooperation of various academic disciplines (e.g., pedagogy, psychology, sociology, or educational studies; see Deutscher Bildungsrat [German Education Council], 1974; Gräsel, 2015), in recent years, it has gained a within-field content-related dimension with regard to the diverse school subjects under investigation. The validity of findings from mathematical and scientific contexts, on which instructional research has mainly focused so far, is being questioned with regard to disparate teaching and learning conditions and subject-specific cultures in the human and social sciences–and their generalizability, in principle, is doubted (e.g., Praetorius et al., 2018; Schlesinger et al., 2018; Wisniewski et al., 2020). So, the school subject becomes an information-bearing grouping variable at a higher level, which must be adequately considered in the data analysis. The term transdisciplinary educational research is accordingly understood here as research in different school subjects in order to analyze subject-specific peculiarities and interdisciplinary differences on the one hand, and transdisciplinary similarities and generalizable effects on the other. Four different methodological approaches are suitable for this purpose, namely meta-analysis, multilevel models, (latent) multilevel structural equation models, and machine learning, which will be briefly presented individually below. In each case, the underlying theoretical model will be explained and possible applications in transdisciplinary research will be concisely illustrated using a reduced data set from the multidisciplinary research project FALKE (**Fa**chspezifische **L**ehrer**k**ompetenzen im **E**rklären; English: Subject-specific teacher competencies in explaining) as an example.

FALKE involved educational scientists of eleven different school subjects (Art, Biology, Chemistry, English, German, History, Mathematics, Music, Physics, Primary School Education, and Protestant Religious Education) and scientists of German linguistics as well as of speech science and training (see also Schilcher et al., 2020b). Using a joint study design, they investigated the quality of teaching explanations in the participating school subjects. For this purpose, five transdisciplinary criteria (structuredness, addressee orientation, linguistic comprehensibility, speech and body expression, personality effect) and one domain-specific criterion per subject (e.g., the importance of causality structures in History) were conceptualized and operationalized with corresponding items in an online questionnaire. In addition, six explanatory videos (seven in the case of the school subject Music), with varying didactical approaches (e.g., inductive vs. deductive) were created for each subject and shown to school students as typical addressees of explanations and (student) teachers as (prospective) experts in explaining. These two groups ($N = 3.116$ participants) first rated the videos globally and then according to the six criteria mentioned, each of which was represented by an individual scale. One of the main transdisciplinary research questions was, e.g., which of the criteria are relevant for the global rating of teaching explanations as being of high quality and whether the relationships are similar across all school subjects or whether there are differences between subjects.

Since a complete presentation of the FALKE project is beyond the scope of this paper (for details see Schilcher et al., 2020a), the investigation of this research question will be limited in the following to the correlation between structuredness and global rating for didactic and illustrative purposes. However, it will be examined under four different methodological approaches (meta-analysis, multilevel models, multilevel structural equation models, and machine learning). The data and script of these exemplary analyses, which were carried out using the statistical software R (R Core Team, 2019), are available online at https://osf.io/5sn9j.

META-ANALYTICAL APPROACHES IN INTERDISCIPLINARY STUDIES

With the aim of recording previous research in a certain area as comprehensively and systematically as possible and reporting its state of the art and core results concisely (e.g., Seidel and Shavelson, 2007; Hattie, 2009), meta-analytical procedures have long been part of the methodical inventory in educational research. Primary effects are summarized and weighted according to mathematically defined, objectifiable criteria and publication bias, content, as well as methodological quality and, in particular, sample size of primary studies can be taken into account as influencing variables. Thus, meta-analyses can reduce the distracting effects of sampling errors, measurement errors, and other artifacts that create the impression of extreme, sometimes even contradictory results of primary studies, and at the same time provide a measure of their consistency (Borenstein et al., 2009; Schmidt and Hunter, 2015). Which kind of effect size is applied in a meta-analysis is of secondary importance, as long as they are independent of study design aspects (such as sample size, covariates used, etc.), easy to calculate from the typically reported statistical information, and have good technical properties for further processing (e.g., known distribution; Borenstein et al., 2009). Accordingly, meta-analyses commonly use standardized distance measures (e.g., Cohen's d or Hedges' g) or standardized correlation measures (e.g., Pearson's product-moment correlation r).

The estimated meta-effect $\hat{\Phi}$ is nothing other than a weighted average, whereby its meaning and the weighting of the individual studies depend on two different theoretical assumptions about their distribution: a fixed effect model or a random effects model. In the fixed effect model, it is assumed that the same true population effect Φ underlies each individual study ($i = 1, \ldots, k$; k number of primary studies), which means that all analyzed effects are the same, and that the observed effect Z_i deviates only by sampling error ε_i with $Z_i = \Phi + \varepsilon_i$. Since these sampling errors depend largely on the sample size of the primary studies, the weight w_i of the respective effects is calculated as a function of the sample size N_i, so that more precisely estimated effects receive larger weights, while more roughly estimated ones receive smaller weights when determining the estimated population effect:

$$\hat{\Phi} = \frac{\sum_{i=1}^{k} w_i \times Z_i}{\sum_{i=1}^{k} w_i}.$$

The only source of variance is thus the sampling error of the studies ε_i with assumed $\varepsilon_i \sim N(0; \sigma^2)$.

However, since research designs of primary studies, even if they are identical, are sometimes carried out with varying details and because target populations differ (e.g., in terms of age, education, socioeconomic status, or subject-specific culture), the assumption of the fixed effect model is rarely correct. The true effect sizes Z_i in all studies ($i = 1, \ldots, k$; k number of primary studies) may be similar but are not likely to be identical. Accordingly, a random effects model assumes that the true effect sizes are a random sample from the population of all possible study effects and (normally) distributed around the true overall effect Φ. The true effects of the individual studies deviate from this by a study-specific value ζ_i and by a sampling error ε_i with $Z_i = \Phi + \zeta_i + \varepsilon_i$. Thus, the variance comprises two components: an inter-study variance τ^2 and an intra-study variance σ^2, both of which are included in the weighting (w_i^*) for the estimation of the meta-effect–on the one hand, in accordance with the random distribution assumption, and on the other hand, to take into account the precision (sample size) of each individual study i:

$$\hat{\Phi} = \frac{\sum_{i=1}^{k} w_i^* \times Z_i}{\sum_{i=1}^{k} w_i^*}$$

(for details Borenstein et al., 2009; Schmidt and Hunter, 2015).

This not only shows that the fixed effect model is a special case of the random effects model when the inter-study variance τ^2 is zero, and the use of a random effects model is generally recommended. Rather, attention shifts from the overall effect to the distribution of study effects when these vary substantially, and the meta-analytical procedures are functional continuations of analyses used in primary studies (e.g., analysis of variance, multiple regression; Borenstein et al., 2009). Thus, in analogy to one-way analysis of variance, the measure Q for the weighted square sums, which follows a central χ^2 distribution with $df = k\text{–}1$ degrees of freedom, and a corresponding null hypothesis significance test are used to check whether the heterogeneity of the individual study effects differs from zero. The variance of the effect size parameters of the primary studies is denoted as τ^2 with the corresponding standard deviation $\tau = \sqrt[2]{\tau^2}$. In addition, the parameter I^2 expresses the proportion of the total variance (= inter- and intra-study variance) that is actually due to the heterogeneity of the study effects. Thus, I^2 is a measure of the inconsistency within the study effects and is comparable with the coefficient of determination of classical variance-analytical procedures R^2. According to Higgins et al. (2003), tentative benchmarks or conventions for the proportion of true inter-study variance in the total variance are 25% low, 50% medium, and 75% high. Even small values for I^2, however, may present good reasons for the inter-study variance to be elucidated, for example by subgroup analyses or meta-regressions (see Borenstein et al., 2009; Schmidt and Hunter, 2015).

In the transdisciplinary educational context, meta-analytical procedures can be applied as usual to combine the results of several studies on one or more subjects (e.g., Seidel and Shavelson, 2007; Praetorius et al., 2018). On the other hand, however, their application is particularly suitable when, within an interdisciplinary research approach, several subject-specific studies with the same study design are to be compared and generalized. This specific usage is finally illustrated by an example from the FALKE project, in which among many other things the relationship between structuredness and global rating of explanations in eleven different school subjects was investigated. The corresponding correlation results, including the precision of the respective estimates, which are represented in the forest plot with 95% confidence intervals, and the distribution of the subject-specific effects are shown in **Figure 1**.

In order to investigate the size of the correlation between structuredness and global rating across all school subjects, the meta-effect was determined using both the fixed and the random effects model, with both approaches leading to the same result ($r = 0.44$). **Figure 1** clearly shows the different weightings that correspond to the sample sizes of the primary studies in the fixed effect model. Also, based on the hypothetical assumption that the true effect is identical in each subject, the estimation of the meta-effect turns out to be rather precise ($CI_{0.95} = [0.43; 0.46]$).

However, it seems theoretically more sound to assume that the effects observed in the individual studies are only a random sample due to, among other factors, subject-specific practices, different explanatory themes and addressees, and heterogeneous sample compositions–clearly, a random effects model seems more suitable. In this model, the studies are weighted almost

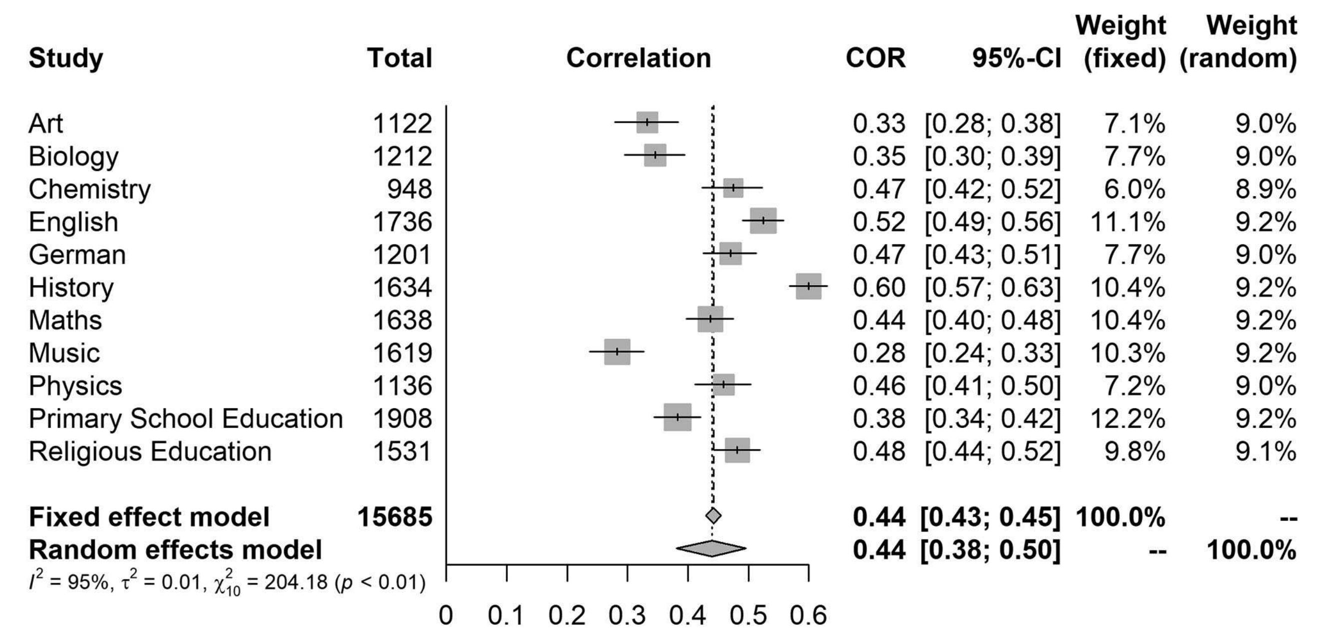

FIGURE 1 | Forestplot for the subject-specific relationships between structuredness and global rating and results of the fixed effect and random effects model.

equally (see also **Figure 1**) and the confidence interval of the meta-effect is larger ($CI_{0.95} = [0.38; 0.50]$), since the distribution of the subject-specific effects is also taken into account. As expected, this heterogeneity is significant ($Q \cong \chi^2_{10} = 204.18$, $p < 0.01$), and the inter-study variance is $\tau^2 = 0.01$ (standard deviation: $\tau = 0.10$). This variance can almost completely ($I^2 = 95\%$) be attributed to a true heterogeneity between the subject-related correlations and must be clarified in further analyses (Schilcher et al., 2020a).

HIERARCHICAL DATA STRUCTURES AND MANIFEST MULTILEVEL MODELS

While meta-analytical approaches for investigating transdisciplinary issues are based on published results data, for raw data structured according to studies (here: school subjects), multilevel models are used to simultaneously determine the (residual) variance of the study-related effect size parameters and the overall effect size (Raudenbush and Bryk, 2002). The hierarchical data structures to be considered here, in which analysis objects at the individual level can be assigned to one or more superordinate units, are well-known in educational research from a large number of applications and are accordingly widely discussed in the methodological literature (Ditton, 1998; Raudenbush and Bryk, 2002; Marsh et al., 2012; Beretvas et al., 2015; Nagengast and Rose, 2018). For example, students (level 1) are nested in classes (level 2), classes in schools (level 3), schools in administrative units (level 4), administrative units in countries (level 5), and so forth. The resulting potential similarity or dependence of measured values within the same category, the size of which can be determined by means of the intraclass correlation coefficient (*ICC*), violates the independence assumption of errors required by close to all classical models. This violation endangers the validity of statistical conclusions, since spurious correlations between variables, biased estimates of model parameters, underestimation of standard errors and, with regard to null hypothesis significance testing, inflated Type-1-error probabilities are some of the possible consequences (Ditton, 1998; Raudenbush and Bryk, 2002; Snijders and Bosker, 2012; Beretvas et al., 2015; Nagengast and Rose, 2018).

By specifying residual matrices at both the individual and the grouping levels (the mixing of the error terms is the reason for the often-used term “mixed models” instead of multilevel models), multilevel models explicitly consider hierarchical structures in the data. Also, these models allow for the straightforward inclusion of features and their relationships at different aggregation levels, since these are (mathematically) independent of each other (e.g., level 1: mathematics achievement, socio-economic status; level 2: classroom climate, class size; level 3: school track, school facilities; level 4: infrastructure, curriculum; level 5: gross domestic product, development level; cf. Snijders and Bosker, 2012; Beretvas et al., 2015; Nagengast and Rose, 2018). Compared to a conventional ordinary least squares regression model, the equation of a simple hierarchical model with two levels, for example, contains two additional random components (also with mean zero), which model the deviations u_{0j} from the group-specific regression intercepts from the overall intercept γ_{00} on the one hand, and the deviations u_{1j} of the group-specific regression slopes from the overall slope γ_{10} on the other hand:

$$Y_{ij} = \gamma_{00} + u_{oj} + (\gamma_{10} + u_{1j})X_{ij} + r_{ij}$$

with Y_{ij} representing the dependent variable, X_{ij} the value of the independent variable, and r_{ij} represents the error term of

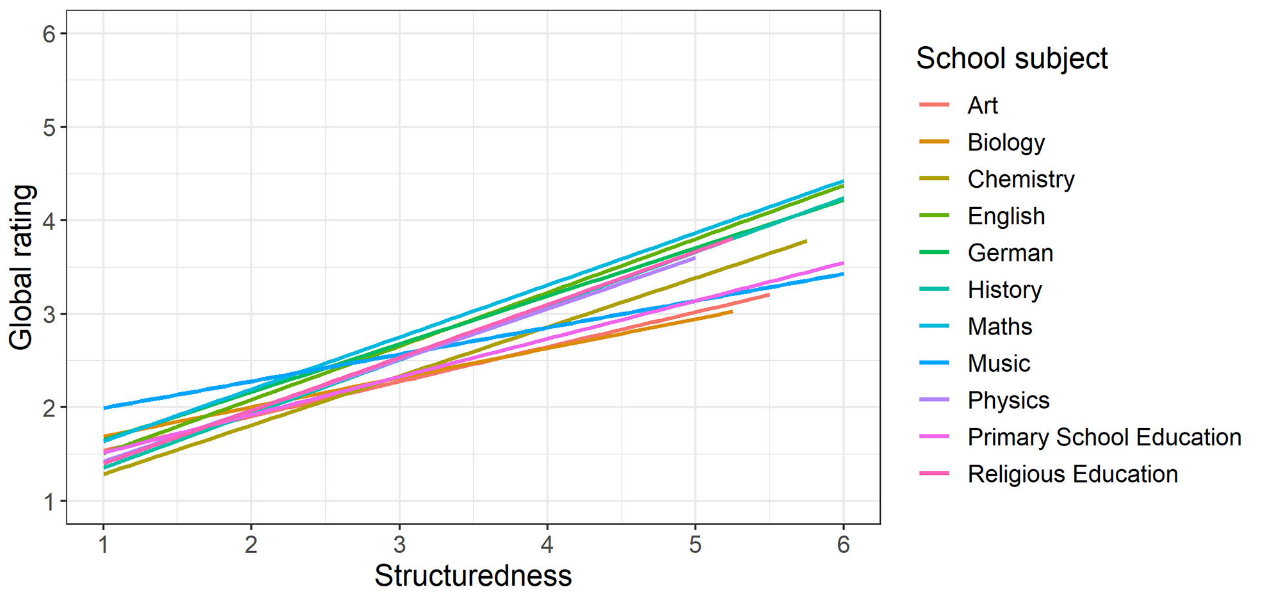

FIGURE 2 | Line graph of the subject-specific relationships between structuredness and global rating. Scaling of structuredness and global rating 1 = very good, ..., 6 = insufficient.

TABLE 1 | Random coefficient model with the dependent variable global rating for eleven school subjects.

***Obs.*: 15685**	**Fixed effects**			**Random effects**		
***ICC*: 2.26%**	***γ***	***SE γ***	**95% *CI γ***	**Per**	***SD***	**95% *CI SD***
Intercept	0.91	0.08	[0.75; 1.07]	Subject	0.26	[0.14; 0.37]
Structuredness	0.52	0.03	[0.45; 0.59]	Subject	0.11	[0.06; 0.15]
Marginal R^2		0.19		Conditional R^2	0.22	

Obs., number of observations; ICC, intraclass correlation; γ, (unstandardized) regression coefficient; SE, standard error; SD, standard deviation; CI, confidence interval (on 1,000 bootstrapping samples); R^2, coefficient of determination.

the entity i, with $i = 1,..., n_j$, in group j, with $j = 1,..., k$. The application of this so-called 'random coefficient model', in which regression constants as well as the predictors' regression weights vary freely over superordinate levels, is illustrated below in simplified examples with only transdisciplinary (2 levels) or with longitudinal and transdisciplinary data structure (3 levels).[1]

Multilevel Models (Considering Context-Related Data Structures)

With reference to the example in section Meta-Analytical Approaches in Interdisciplinary Studies, the correlation between structuredness and global rating of explanations for the eleven school subjects involved in the FALKE project will be investigated, taking into account the overall (transdisciplinary) correlation as well as the variance of the subject-specific relationships shown in **Figure 2**. To this end, a simple random coefficient regression with the dependent variable global rating and the independent variable structuredness is used to model the nested data structure arranged by school subject, in which the regression intercepts and slopes are variably modeled at subject level. The unstandardized results of this estimation, with both variables were measured on the same six-category response scale, are shown in **Table 1**.

As can be seen from **Table 1**, the global (transdisciplinary) regression coefficient for structuredness is $\gamma_{10} = 0.52$ and is significant. This means that, starting from the global intercept of $\gamma_{00} = 0.91$ (intersection of the overall regression line with the ordinate axis; cannot be interpreted in a meaningful way here), the global rating, on average, increases by about half a unit for each rating unit by which the structuredness increases. In terms of content, this shows that there is a positive correlation between the structuredness and the global rating of an explanation, that is, on average, the better structured an explanation is perceived the better it is rated overall. But this correlation is not the same in all subjects. In the present case of only one predictor variable, the intercepts ($SD = 0.26$) as well as slopes ($SD = 0.11$) not only vary significantly between the school subjects, so that in individual subjects there may be lower or higher starting levels and smaller or larger correlations between global rating and structuredness, which are visualized in **Figure 2** (for numerical details see **Table 2**). Rather, there is a significant correlation of $r = -0.86$ ($CI_{0.95}[-0.97; -0.50]$) between intercepts and slopes: the smaller the intercept, the greater the slope between global rating and structuredness or, in other words, the better very well-structured explanations are globally rated in an interdisciplinary comparison, the worse are very poorly structured explanations. On the one hand, this can be seen numerically from **Table 2**, which is additionally presented here for illustration purposes and contains the subject-specific model coefficients. On the other hand, the effect is shown graphically in **Figure 2**.

The variance explained by the present hierarchical model is acceptable for both the fixed effects (marginal $R^2 = 0.19$) as well as the fixed and random effects together (conditional $R^2 = 0.22$). In conclusion, it should be noted that with previous z-standardization of the variables global rating and structuredness per school subject, the reported random coefficient model (apart from small deviations and discrepancies due to different estimation procedures and rounding) leads to the same results as the random effects model of the meta-analysis (section Meta-Analytical Approaches in Interdisciplinary Studies), thus highlighting the obvious parallels between these two approaches.

Mixed Linear Models (Considering Longitudinal Data Structures)

Longitudinal data structures are a fairly regular case in educational research, for example when investigating the effectiveness of teaching methods with a pre- and a post-test, offer a specific application situation for multilevel models. Each person is assigned at least two measurement values (e.g., the pre- and the post-test results). The data can therefore be thought of as 'nested within persons'. At the same time, the persons are often divided into different groups (e.g., control and experimental group) at random or systematically according to different test conditions. According to Hilbert et al. (2019), mixed linear models with dummy-coded predictor variables are particularly suitable for analyzing studies with this type of design, since they are superior to traditional methods such as repeated measurement ANOVAs or OLS regressions with regard to less stringent model assumptions and higher statistical power (see also Raudenbush and Bryk, 2002). The approach proposed by Hilbert et al. is easily applicable to a transdisciplinary context by extending the nesting of the model to take different school subjects into account. For an exemplary case, data from the FALKE project will again be used to illustrate the model.

In (almost) all school subjects, two explanatory videos present the same teaching content using two didactically different approaches (A vs. B). These video pairs were shown to students on the one hand, and to teachers on the other, and both groups were asked to give their global rating (Schilcher et al., 2020b). An illustration of the results is provided in **Figure 3**, which shows differences in the rating depending on the group, didactical method, and subject.

In order to analyze the differences shown in **Figure 3** with a linear mixed model, the variable for the didactical method of

[1]In order to make the examples clear and comprehensible, the modelling of further levels that may be contained in the data (e.g., class, school) is avoided for didactical reasons.

TABLE 2 | Subject-specific coefficients for the random coefficient model in **Table 1**.

	Art	Bi	Ch	En	Ge	Hi	Ma	Mu	Ph	PSE	Re
Intercept	0.96	1.22	0.67	0.72	0.98	0.65	0.90	1.45	0.73	0.97	0.77
Structure	0.43	0.36	0.55	0.64	0.55	0.61	0.59	0.36	0.60	0.42	0.58

Art, art; Bi, biology; Ch, chemistry; En, English; Ge, German; Hi, history; Ma, mathematics; Mu, music; Ph, physics; PSE, Primary School Education; Re, Protestant Religious Education.

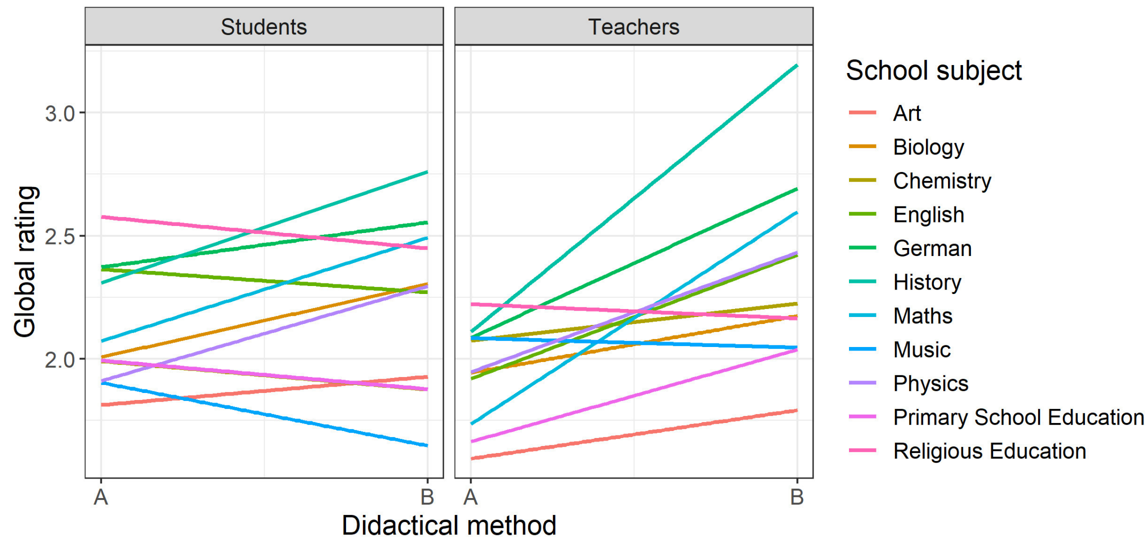

FIGURE 3 | Line graph of the subject-specific relationships between method and global rating, separated by students, and teachers.

a video pair (A: 0 vs. B: 1) as well as the variable for group membership (students: 0 vs. teachers: 1) are dummy-coded. The model includes both main effects as well as the interaction effect of the variables. Importantly, the interaction effect represents the additional rating difference between didactical method A and B for teachers compared to the students. Since the data are nested within persons, a person-specific residual term is included on the second level. In addition, school subject grouping is modeled as a third level, by which the regression intercept and slope parameters of all predictors may vary to obtain estimates for both the generalized effects and the transdisciplinary distribution of effects. The (non-standardized) coefficients of the corresponding linear mixed model are shown in **Table 3**.

Across all school subjects, students rate didactical method A on average with $\gamma_{00} = 2.12$, although this value varies significantly between disciplines ($SD = 0.23$; **Table 3**). The corresponding rating of the teachers is on average significantly lower by $\gamma_{10} = -0.18$ than compared to the students' and shows a significant variation from discipline to discipline ($SD = 0.18$). While there is no significant overall tendency among students across school subjects in favor of the didactical variant B ($\gamma_{01} = 0.10$, but the 95% *CI* includes the value 0), the significant transdisciplinary interaction effect between group and method ($\gamma_{11} = 0.30$) is: Compared to method A, the teachers assigned didactical method B a significantly higher average rating than the students (for an exhaustive description of the different model parameters and their interpretation, see Hilbert et al., 2019). In the present model, the variance that is explained by the fixed effects is small (marginal $R^2 = 0.02$), that explained by fixed and random effects is appropriate (conditional $R^2 = 0.36$). Thus, an interdisciplinary generalization of the results only

TABLE 3 | Linear mixed model with the dependent variable global rating for eleven school subjects.

***Obs.*: 5957**	**Fixed effects**			**Random effects**		
***ICC*: 27.62%**	**γ**	***SE* γ**	**95% *CI* γ**	**Per**	***SD***	**95% *CI SD***
Intercept	2.12	0.07	[1.98; 2.27]	Id	0.50	[0.47; 0.53]
				Subject	0.23	[0.11; 0.35]
Group	−0.18	0.06	[−0.31; −0.06]	Subject	0.18	[0.07; 0.28]
Method	0.10	0.08	[−0.06; 0.26]	Subject	0.24	[0.11; 0.35]
Group × Method	0.30	0.07	[0.16; 0.44]	Subject	0.20	[0.07; 0.32]
Marginal R^2		0.02		Conditional R^2	0.36	

Obs., number of observations; ICC, intraclass correlation; γ, (unstandardized) regression coefficient; SE, standard error; SD, standard deviation; CI, confidence interval (on 1,000 bootstrapping samples); R^2, coefficient of determination.

appears to make sense regarding the transdisciplinary variance of the effects.

LATENT MULTILEVEL STRUCTURAL EQUATION MODELS

The multilevel models described above are based on manifest scale values for each construct such as sum or mean values or the proportion of correctly solved tasks. However, any multiple indicators of the constructs, their factor structure and particularly measurement errors are not considered in manifest models (Marsh et al., 2012; Beretvas et al., 2015). This implies the assumption that all relevant variables are directly observable (and measured without errors), which hardly seems possible–in particular regarding typical target variables in the social sciences and educational research, such as (cognitive) abilities, knowledge, competence, skills, attitudes, or motivation. In contrast, structural equation models take up the basic idea of latent modeling, that is to capture a feature which is not directly observable only by means of various indicators, in whose manifestations this feature is reflected. Latent models split the variance of the manifest indicators into the measurement error component and the component of the latent variable on which the scale values are based. At the same time, the use of latent structural equation models allows the analysis of complex variable systems with several exogenous and endogenous elements (Kline, 2011; Beretvas et al., 2015; Nagengast and Rose, 2018).

By extending the multilevel approach, these advantages can also be used in latent multilevel structural equation models in which features can be measured and analyzed simultaneously at different levels of analysis (e.g., students, classes, school, subject; Raudenbush and Bryk, 2002). Possible applications of such models, for example in the context of instructional quality research, are shown by Baumert et al. (2010), Kunter et al. (2013) as well as Wisniewski et al. (2020) and their particular merit is underlined by Marsh et al. (2012). Because of the specific methodological requirements of educational research, in which manifest variables mostly reflect influences from several levels, these authors suggest the use of double latent models, which will be illustrated below using a simplified example.

Analogous to sections Meta-Analytical Approaches in Interdisciplinary Studies and Multilevel Models (Considering Context-Related Data Structures), the transdisciplinary relationship between structuredness and global rating of explanations is examined, taking into account individual differences (level 1) and heterogeneous subject cultures (level 2). For this purpose, a (latent) multilevel structural equation model, in which the structuredness is simultaneously indicated at levels 1 and 2 by the four items belonging to this latent construct, is estimated (**Figure 4**). The manifest value of the global rating indicator is decomposed into latent variance components at levels 1 and 2 as endogenous variables in each

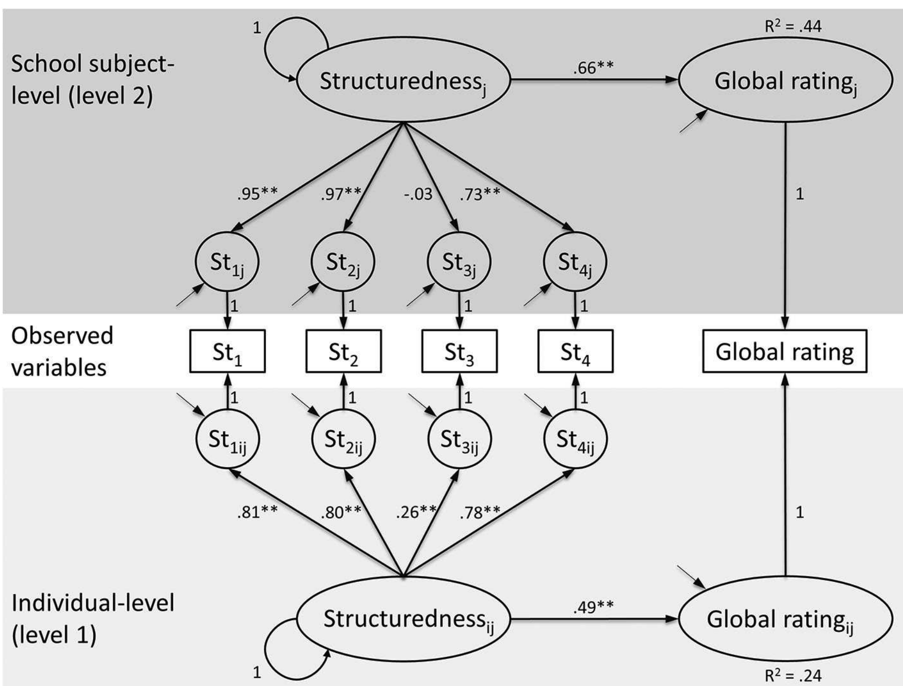

FIGURE 4 | (Doubly) Latent structure equation model for the correlation between structuredness and global rating at individual and school subject level. Latent constructs are represented as circles and indicators of these variables as squares. The boxes representing the observed variables are associated with both individual and subject-specific constructs (see Marsh et al., 2012); residual variances are not reported. The subscripts *ij* indicate that these variables take on different values for each student *i* in each classroom *j*. Number of observations: 15,639, number of clusters: 11; $\chi^2(10) = 173.13^{**}$, *CFI* = 0.99, *RMSEA* = 0.03, *SRMR* (within) = 0.02, *SRMR* (between) = 0.09; $^*p < 0.05$, $^{**}p < 0.01$.

case. **Figure 4** shows the corresponding measurement and structure models including the standardized factor loadings, variances, and regression coefficients (without residuals). The proportion of variance that can be explained by the school subject structure (*ICC* 1) is 2.27% [see section Multilevel Models (Considering Context-Related Data Structures), **Table 1**], the reliability of the subject-specific group means is 0.97 (*ICC* 2; Bliese, 2000) and the local and global fit values of the model are acceptable (**Figure 4**; Hu and Bentler, 1999). The standardized correlation between structuredness and global rating is $\beta_1 = 0.49$ ($p < 0.01$) at individual level and $\beta_2 = 0.66$ ($p < 0.01$) at subject level. Thus, due to the high factor reliability, the latent transdisciplinary effect of $R^2 = 0.44$ ($= \beta_2^2$) corresponds to the (measurement error-afflicted) estimates of the meta-analysis (section Meta-Analytical Approaches in Interdisciplinary Studies) and the multilevel model with standardized coefficients [section Multilevel Models (Considering Context-Related Data Structures)].

MACHINE LEARNING METHODS

Although the methods presented so far are suitable and proven for a large number of applications in the field of educational science, they require stringent distributional and model assumptions and can only handle a relatively restricted number of variables and constructs. This makes it difficult to adequately analyze large, weakly structured or short-lived datasets, which are summarized under the collective term "big data," increasingly available due to digitalization and also necessary to investigate the multifaceted complexity of many educational phenomena. In order to meet these methodological challenges, various data mining methods have been applied for many years and are constantly being further developed, which has been particularly favored by the rapid increase in computing power over the last two decades (Romero and Ventura, 2020; for an overview, see Fischer et al., 2020). These include machine learning methods, which enable an effective analysis of enormous amounts of data and complex data structures almost without distributional assumptions. So far, machine learning approaches have only rarely been used in empirical educational research (e.g., Kotsiantis, 2012), but they represent a promising alternative for the analysis of national and international large scale studies, such as PISA (Programme for International Student Assessment) or TIMSS (e.g., Depren et al., 2017; Yoo, 2018; Trends in International Mathematics and Science Study), for secondary analyses (e.g., Pargent and Albert-von der Gönna, 2018) or, as will be outlined in the following, for the investigation of transdisciplinary analyses.

From a theoretically unlimited number of variables, those relevant for predictions are automatically selected by machine learning algorithms and overfitting is prevented by a strict distinction between training and test data with resampling methods using multiple loops. This method is called (nested) resampling, because the entire sample of cases is recursively split into a training set, typically comprising two thirds of the data, and a test set, comprising the remaining third. The model is then trained with the training data only until the most predictive variables are selected (or equipped with large weights) and their interaction is modeled. The accuracy of the resulting model, however, is estimated through the performance of the test data, which has not been used to train the model. Overfitting the model to the training data therefore results in worse fit on the test data (because even random aspects of the training data enter the model, which have no bearing in the test data; Efron and Hastie, 2016). This procedure requires the data to be labeled before training, so that the prediction accuracy can be determined by the percentage of correctly predicted labels in the test data. These labels may be categorical or numerical. For categorical data, the percentage of the correct category is usually used as a measure of prediction accuracy, while for numerical data, the mean squared error is often employed. Machine learning with labeled data is termed "supervised learning," because the correctness of the result can be supervised through comparison of the labels with the predictions of the model.

This means that the models can be more easily generalized than conventional analysis methods, even though they are typically more exploratory and less theory-driven than classical statistical models (Efron and Hastie, 2016). A widespread criticism regarding machine learning techniques lies within the data-driven inherently exploratory approach of these models, which is partly simply the downside of their greatest strength, namely the lack of model assumptions. However, several techniques have been developed to look into the former blackbox that machine learning used to represented. Feature engineering has become a more and more prominent part of machine learning. It refers to the preparation of predictor variables (typically called "features" in the context of machine learning) to pre-process variables in a usually meaningful way to make them more valuable for the model. Goerigk et al. (2020), for example, extracted factor scores from structural equation models to use them as features in their models. The rapidly growing field of interpretable machine learning uses various techniques to infer the effect of single variables on the prediction accuracy, usually graphically illustrated through variable importance plots, partial dependence plots, or accumulated local dependence (Molnar, 2019). As will be illustrated below in an exemplary analysis of the FALKE data, using sum scores of scales and variable importance plots can lead to interpretable, theory-based results, even though this is not the core-strength of the machine learning techniques.

To provide a simple example, a random forest (Breiman, 2001) was used to analyze the FALKE data. One of the advantages of this (and most other common) machine learning model(s) is that it is not based on distributional and linearity assumptions. Random forest models simply randomize and average a large number of mathematical trees that split the sample according to the most suitable splitting points in the most suitable variable. In this example, the random forest model was used to predict the school subject of a video through the ratings on the six constructs operationalized in FALKE (including the global rating). In addition, feature importance (see Molnar

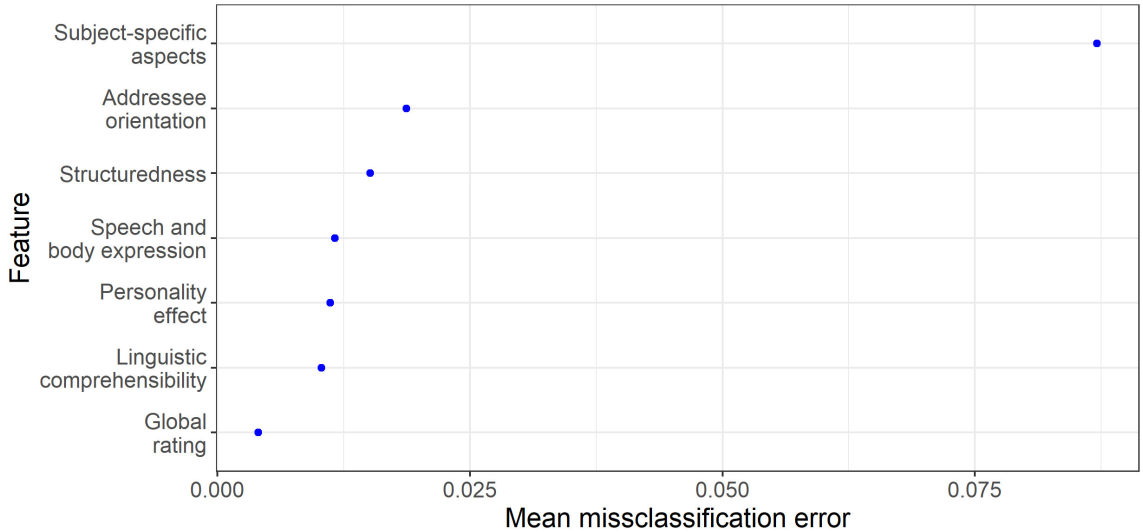

FIGURE 5 | Feature importance plot regarding the prediction of school subject membership by the six constructs and the global rating operationalized in FALKE.

et al., 2018) was estimated by sampling to determine which construct is most valuable for the prediction of the school subject. Despite the low number of predictors and high number of categories, this model already assigns 58.1% of all test set cases to the correct school subject. Notably, in contrast to the performance estimates presented in the previous sections (such as R^2), this is the accuracy for the testing sample, meaning cases the model has not been trained with. As shown in the representation of the variable importance (**Figure 5**), as expected, the subject-specific construct that operationalizes aspects typical for explanations in this subject (e.g., substance-particle level in Chemistry or acoustic vs. visual approaches in Music; Schilcher et al., 2020a) clearly has the greatest predictive power.

COMPARATIVE CONCLUSION AND FURTHER RECOMMENDATIONS

In the preceding sections, four different methods were presented for adequately dealing with methodological challenges such as meta-analytical approaches, hierarchical data structures, large measurement errors, or big and complex amounts of data, which are often present in transdisciplinary empirical educational research. The first three of these approaches–meta-analyses, multilevel models and latent multilevel structural equation models–are based, as cross-references between the respective sections illustrate, on the same classical framework of the Generalized Linear Model, which has several limitations. For instance, the choice of model is not only limited by the level of measurement and distributional assumptions. Rather, the requirement of a particular (mostly linear) relationship between variables itself is by no means self-evident, especially in teaching and learning contexts, and complex relational structures can easily be missed or even interpreted in erroneous ways with linear models. Moreover, the number of variables that can be considered simultaneously in is typically rather small due to multicollinearity problems and this also restrict the mapping of more complex relationships. Since the models are typically fitted exclusively to the respective underlying sample and rarely cross-validated or re-evaluated on the basis of additional samples, the classically reported coefficient of determination R^2 usually substantially overestimates their predictivity and their generalizability must therefore be critically questioned. Machine learning methods, on the other hand, do not have these limitations of the classical General Linear Model and can take them into account in modeling (see section Machine Learning Methods). Due to their versatile application potential, they thus enrich the current inventory of methods in transdisciplinary educational research (but also in empirical educational research in general) and appear to be an integral part of the future state of the art methods, especially for the analysis of "big data" (Efron and Hastie, 2016; Stachl et al., 2020). Their primarily explorative approach can be monitored and verified by contemporary interpretable machine learning methods (Molnar, 2019). On the other hand, machine learning models have not been developed for theory-testing purposes, but to maximize model predictivity, often at the expense of interpretability. The strength of the three approaches based on (generalized) linear models is the focus on testable hypotheses and the direct and interpretable quantification of deviations from proposed model fit.

In conclusion, it should be noted that all of the presented methods require rather large samples (Marsh et al., 2012), although the recommendations for minimum sample sizes (per analysis level) vary depending on the type of analysis as well as the model type and complexity and are controversially discussed in the methodological literature (e.g., Borenstein et al., 2009; Hox, 2010; Marsh et al., 2012). For the aggregation and evaluation of already published study effects, the application of a meta-analysis with a random effects model is appropriate. Here, the number of underlying effects should be enough to obtain a meaningful estimate of the between-studies variance. Using the statistical software R (R Core Team, 2019), central packages for meta-analyses are "meta" (Balduzzi et al., 2019) and "metafor" (Viechtbauer, 2010), and further information about meta-analysis that could not be presented in this brief introduction is provided by Borenstein et al. (2009) and Schmidt and Hunter (2015). Multilevel analyses with manifest variables are suitable, however, if hierarchical data structures exist due to context variables, but also due to measurements at several time points. A ratio of 30 : 30 is often given as the minimum for simple two-level models, but this is only a vague benchmark that depends mainly on the concrete data situation. Also, even though theoretically possible, rarely can more than three levels be modeled meaningfully and estimated in hierarchical models. Useful R packages for multilevel models are "multilevel" (Bliese, 2016), "lme4" (Bates et al., 2015), "lmerTest" (Kuznetsova et al., 2017) as well as "MuMIn" (Barton, 2020). Further application notes are provided by Ditton (1998), Raudenbush and Bryk (2002), Hox (2010), and Snijders and Bosker (2012). If measurement errors or more complex relationships between variables are to be modeled additionally, the use of latent multilevel structural equation models is recommended. Besides an appropriate ratio of persons and parameters to be estimated (at least 10 : 1), from a multilevel perspective the effective sample size is the number of higher level units (at least 50), not just the number of individual level subjects. For the analysis of these models using R, the packages "lavaan" (Rosseel, 2012) and "sem" (Fox et al., 2017) are necessary and additional references to latent (multilevel) structure equation modeling can be found in Kline (2011) and Marsh et al. (2012). Finally, the benefits and efficiency of machine learning methods become more apparent the more extensive and confusing the data set to be analyzed is ($\gg$1,000 persons and/or variables). A basic R package for the application of machine learning methods is "mlr" (Bischl et al., 2016) and an in-depth introduction is provided by Efron and Hastie (2016).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Open Science Framework https://osf.io/5sn9j.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This publication is a result of the KOLEG project (**Ko**operative **L**ehrerbildung **G**estalten) at the University of Regensburg, which was funded by the German Federal Ministry of Education and Research as part of the joint quality offensive for teacher training by the federal and state governments (grant number: 01JA1512).

ACKNOWLEDGMENTS

We would like to thank the numerous students, student teachers, teachers, and teacher trainers of the different school subjects very much for their voluntary participation in the study. We also thank the editors and reviewers of Frontiers for their critical and very helpful comments on earlier versions of this paper.

REFERENCES

Balduzzi, S., Rücker, G., and Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* 22, 153–160. doi: 10.1136/ebmental-2019-300117

Barton, K. (2020). *MuMIn: Multi-Model Inference. R Package Version 1.43.17*. Available online at: https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *Am. Educ. Res. J.* 47, 133–180. doi: 10.3102/0002831209345157

Beretvas, S. N., Whittaker, T. A., and Stafford, R. E. (2015). "Statistical modeling methods for classroom management research," in *Handbook of Classroom Management, 2nd Edn*, eds E. T. Emmer and E. J. Sabornie (New York, NY: Routledge), 519–537.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., et al. (2016). mlr: machine learning in R. *J. Mach. Learn. Res.* 17, 1–5.

Bliese, P. (2000). "Within-group agreement, non-independence, and reliability," in: *Multilevel Theory, Research, and Methods in Organisations,* eds K. Klein and S. Kozlowski (San Francisco: Jossey-Bass), 349–381.

Bliese, P. (2016). *multilevel: Multilevel Functions. R package version 2.6*. Available online at: https://CRAN.R-project.org/package=multilevel

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons. doi: 10.1002/9780470743386

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Depren, S. K., Aşkin, Ö. E., and Öz, E. (2017). Identifying the classification performances of educational data mining methods: a case study for TIMSS. *Educ. Sci. Theory Pract.* 17, 1605–1623. doi: 10.12738/estp.2017.5.0634

Deutscher Bildungsrat [German Education Council] (1974). *Empfehlungen der Bildungskommission – Aspekte für die Planung der Bildungsforschung*. Bonn: Bundesdruckerei.

Ditton, H. (1998). *Mehrebenenanalyse.* Grundlagen und Anwendungen des Hierarchisch Linearen Modells. Weinheim/München: Beltz Juventa.

Efron, B., and Hastie, T. (2016). *Computer Age Statistical Inference. Algorithms, Evidence, and Data Science.* New York, NY: Cambridge University Press. doi: 10.1017/CBO9781316576533

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., et al. (2020). Mining big data in education: affordances and challenges. *Rev. Res. Educ.* 44, 130–160. doi: 10.3102/0091732X20903304

Fox, J., Nie, Z., and Byrnes, J. (2017). *sem: Structural Equation Models. R package version 3.1-9.* Available online at: https://CRAN.R-project.org/package=sem

Goerigk, S., Hilbert, S., Jobst, A., Falkai, P., Bühner, M., Stachl, C., et al. (2020). Predicting instructed simulation and dissimulation when screening for depressive symptoms. *Eur. Arch. Psychiatry Clin. Neurosci.* 270, 153–168. doi: 10.1007/s00406-018-0967-2

Gräsel, C. (2015). "Was ist Empirische Bildungsforschung?," in: *Empirische Bildungsforschung. Strukturen und Methoden, 2nd Edn,* eds H. Reinders, H. Ditton, C. Gräsel, and B. Gniewosz (Wiesbaden: Springer), 15–30. doi: 10.1007/978-3-531-19992-4_1

Hattie, J. A. C. (2009). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. London: Routledge. doi: 10.4324/9780203887332

Higgins, J., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. doi: 10.1136/bmj.327.7414.557

Hilbert, S., Stadler, M., Lindl, A., Naumann, F., and Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *Test. Psychometrics Methodol. Appl. Psychol.* 26, 101–119. doi: 10.4473/TPM26.1.6

Hox, J. (2010). *Multilevel analysis. Techniques and Applications, 2nd Edn,* New York, NY: Routledge. doi: 10.4324/9780203852279

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model A Multidisc. J.* 6, 1–55. doi: 10.1080/10705519909540118

Kline, P. (2011). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artif. Intell. Rev.* 37, 331–344. doi: 10.1007/s10462-011-9234-x

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., and Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *J. Educ. Psychol.* 105, 805–820. doi: 10.1037/a0032583

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., and Köller, O. (2012). Classroom climate and contextual effects: conceptual and methodological issues in the evaluation of group-level effects. *Educ. Psychol.* 47, 106–124. doi: 10.1080/00461520.2012.670488

Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.* Morrisville: Lulu Press.

Molnar, C., Casalicchio, G., and Bischl, B. (2018). iml: An R package for interpretable machine learning. *J. Open Source Softw.* 3:786. doi: 10.21105/joss.00786

Nagengast, B., and Rose, N. (2018). "Quantitative Bildungsforschung und Assessments," in *Handbuch Bildungsforschung,* eds R. Tippelt and B. Schmidt-Hertha *Vol. 2, 4th Edn*, (Wiesbaden: Springer), 669–688. doi: 10.1007/978-3-531-19981-8_28

Pargent, F., and Albert-von der Gönna, J. (2018). Predictive modeling with psychological panel data. *Zeitschrift für Psychologie* 226, 246–258. doi: 10.1027/2151-2604/a000343

Praetorius, A.-K., Klieme, E., Herbert, B., and Pinger, P. (2018). Generic dimensions of teaching quality. The German framework of three basic dimensions. *ZDM Mathe. Educ.* 50, 407–426. doi: 10.1007/s11858-018-0918-4

R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage.

Romero, C., and Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. *WIREs Data Mining Knowledge Discovery* 10 (3). doi: 10.1002/widm.1355

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Schilcher, A., Krauss, S., Lindl, A., and Hilbert, S. (2020a). *FALKE – Fachspezifische Lehrerkompetenzen im Erklären: Untersuchungen zur Beurteilung und zu Kriterien unterrichtlicher Erklärqualität aus der Perspektive von 13 Fachbereichen*. Weinheim/Basel: Beltz Juventa.

Schilcher, A., Lindl, A., Hilbert, S., Krauss, S., Asen-Molz, K., Ehras, C., et al. (2020b). Experiences from transdisciplinary research. *Front. Educ*. 5, in preparation.

Schlesinger, L., Jentsch, A., Kaiser, G., König, J., and Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM Math. Educ.* 50, 475–490. doi: 10.1007/s11858-018-0917-5

Schmidt, F. L., and Hunter, J. E. (2015). *Methods of Meta-Analysis. Correcting Error and Bias in Research Findings,* 3rd Edn, Thousand Oaks, CA: Sage. doi: 10.4135/9781483398105

Seidel, T., and Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Rev. Educ. Res.* 77, 454–499. doi: 10.3102/0034654307310317

Snijders, T. A. B., and Bosker, R. J. (2012). *Multilevel analysis. An Introduction to Basic and Advanced Multilevel Modelling, 2nd Edn*. London: Sage.

Stachl, C., Pargent, F., Hilbert, S., Harari, G., Schoedel, R., Vaid, S., et al. (2020). Personality research and assessment in the era of machine learning. *Eur. J. Personal.* doi: 10.1002/per.2257. [Epub ahead of print].

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03

Wisniewski, B., Zierer, K., Dresel, M., and Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: two-level structure and measurement invariance. *Learn. Instruct.* 66:101303. doi: 10.1016/j.learninstruc.2020.101303

Yoo, J. E. (2018). TIMSS 2011 student and teacher predictors for mathematics achievement explored and identified via Elastic Net. *Front. Psychol.* 9:317. doi: 10.3389/fpsyg.2018.00317

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

ORIGINAL RESEARCH
published: 20 August 2020
doi: 10.3389/feduc.2020.00150

Interdisciplinary Doctoral Training in Technology-Enhanced Learning in Europe

Viktoria Pammer-Schindler[1], Fridolin Wild[2]*, Mikhail Fominykh[3], Tobias Ley[4], Maria Perifanou[5], Maria Victoria Soule[6], Davinia Hernández-Leo[7], Marco Kalz[8], Ralf Klamma[9], Luís Pedro[10], Carlos Santos[10], Christian Glahn[11], Anastasios A. Economides[5], Antigoni Parmaxi[6], Ekaterina Prasolova-Førland[3], Denis Gillet[12] and Katherine Maillet[13]*

[1] Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria, [2] Institute of Educational Technology, The Open University, Milton Keynes, United Kingdom, [3] Department of Education and Lifelong Learning, Norwegian University of Science and Technology, Trondheim, Norway, [4] School of Educational Sciences, Tallinn University, Tallinn, Estonia, [5] SMILE Lab, University of Macedonia, Thessaloniki, Greece, [6] Language Centre, Cyprus University of Technology, Limassol, Cyprus, [7] ICT Department, ICREA Academia, Universitat Pompeu Fabra, Barcelona, Spain, [8] Department of Technology-Enhanced Learning, Institute for Arts, Music and Media, Heidelberg University of Education, Heidelberg, Germany, [9] Advanced Community Information Systems (ACIS), RWTH Aachen University, Aachen, Germany, [10] DigiMedia, University of Aveiro, Aveiro, Portugal, [11] Zurich University of Applied Sciences, Zurich, Switzerland, [12] School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [13] Telecom SudParis, Institut Polytechnique de Paris, Paris, France

OPEN ACCESS

Edited by:
*Matthias Stadler,
Ludwig Maximilian University
of Munich, Germany*

Reviewed by:
*Stefan Krauss,
University of Regensburg, Germany
Nicole D. Anderson,
MacEwan University, Canada*

***Correspondence:**
*Viktoria Pammer-Schindler
viktoria.pammer@tugraz.at
Fridolin Wild
f.wild@open.ac.uk*

Specialty section:
*This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Education*

Received: *14 April 2020*
Accepted: *29 July 2020*
Published: *20 August 2020*

Citation:
Pammer-Schindler V, Wild F, Fominykh M, Ley T, Perifanou M, Soule MV, Hernández-Leo D, Kalz M, Klamma R, Pedro L, Santos C, Glahn C, Economides AA, Parmaxi A, Prasolova-Førland E, Gillet D and Maillet K (2020) Interdisciplinary Doctoral Training in Technology-Enhanced Learning in Europe. Front. Educ. 5:150. doi: 10.3389/feduc.2020.00150

Technology enhanced learning (TEL) research connects Learning Sciences, Educational Psychology, and Computer Science, in order to investigate interventions based on digital technologies in education and training settings. In this paper, we argue that doctoral training activity for TEL needs to be situated at the intersection of disciplines in order to facilitate innovation. For this, we first review the state of disciplinarity in TEL, reviewing existing meta-studies of the field. Then, we survey 35 doctoral education programs in Europe in which doctoral students working on TEL topics are enrolled. Findings indicate that most doctoral schools are associated with a single discipline and offer methodological rather than content-specific modules. TEL-specific content is provided only in exceptional cases, creating a potentially isolating gap between master-level education and scientific conferences. On this background, we argue that cross-institutional doctoral training is important to progress TEL as a field. In this article, we study and share the approach of an international doctoral summer school organized by the European society EA-TEL over the past 15 years. The summer school provides foundational methodological knowledge from multiple disciplines, content-specific topical knowledge in TEL, access to cutting edge scientific discourse, and discussion of horizontal issues to doctoral students. We further provide an analysis of shifting program topics over time. Our analysis of both, institutional as well as cross-institutional doctoral training in TEL, constitutes this paper's core contribution in that it highlights that further integration of perspectives and knowledge is to be done in TEL; together with codification and explication of knowledge in the intersection of disciplines.

Keywords: technology-enhanced learning, doctoral training, doctoral education, educational technology, learning technology, survey, case study

INTRODUCTION

Research fields are, in many ways, set up as communities of shared knowledge and practice (Lave and Wenger, 1995; Latour, 2005), typically geographically distributed. Communities differentiate themselves from each other with regards to what the agreed objects of interest are, and what are to be considered valid ways of contributing and gaining seniority (Lave and Wenger, 1995; Kuhn, 2012). This includes specific methodological commitments in extension of a generally shared agreement across disciplines that the generation of new knowledge is the goal. Moreover, this also involves an often unspoken agreement as to which publishing venues are considered acceptable and reputable. Doctoral training is often considered an academic rite of passage (Amran and Ibrahim, 2012).

Research fields tend to cascade into Higher Education over time, for instance in the form of doctoral schools, as a way to commodify recruitment and training of future community members. Doctoral education is thereby commonly implemented in non-interdisciplinary academic structures (Lindvig, 2018), while at the same time aiming to establish a transdisciplinary view of science ("mode 2 science"), driven by grand challenges (O'Rourke et al., 2016) that do not regard disciplinary boundaries (Carr et al., 2018).

In principle, one could discuss that "doctoral-level education" (as in "doctoral training program" or "Ph.D. studies") is an oxymoron, as any such expression pretends that the key principles of education could be directly applied to research. Any common definition of 'education' includes the idea of giving and receiving systematic instruction to motivate the re-construction or re-development of existing knowledge, skills, abilities, and other characteristics by the recipient of education, the learner, of course adapted to given context. Even more thought-provoking, ideas of academic knowledge exchange suggest that skills should be transferred from a knowledgeable scholar (and their academic outputs such as textbooks, journal articles, or online course materials). "Research" on the other hand requires systematic investigation, with the aim to discover or develop a novel insight, previously unknown. Delineating it from bachelor (level 6) and master (level 7), the International Standard Classification of Education speaks in this context for its definition of level 8 of requiring submission of "written work of publishable quality that is the product of original research and represents a significant contribution to knowledge in the respective field of study" (ISCED, 2011, p. 60, § 264).

The review of the state of the art, however, has become and will become increasingly more complex, as the amount of codified knowledge (publications, research data) grows continuously year after year. In parallel, methods evolve to take up new possibilities to analyze data, and to do so in a more complex manner. For example, public betas ("facebook as a testbed"), open test collections, online crowdsourcing, and participatory approaches such as citizen science promise to lower barriers to research (regarding access, replication, and reuse, see Cleverdon, 1960; Kittur et al., 2008; Shneiderman, 2008; Herodotou et al., 2014). New requirements emerge regarding ethics, research and research data documentation, and accessibility. From this position, one could argue that as both methodology, practice, and existing knowledge exhibit increased complexity when operated on, there is a need for additional training and guidance beyond the prerequisite bachelor and master levels.

Nevertheless, doctoral training is widely accepted to be a key activity of research communities. Technology Enhanced Learning (TEL) is no exception to this. This article therefore sets out to deliver both an analysis of the current governing structure of doctoral education in TEL, particularly in Europe, and a proposal for a common core of doctoral-level training in TEL. We break this further down into the following research questions:

- RQ1: What is the current practice of institutional doctoral training in TEL in Europe?
- RQ2: How could cross-institutional doctoral education be organized, and which topics are relevant?

We first investigate the state of affairs with regards to the disciplinarity in the field of TEL as background to our present discussion (see section "Technology Enhanced Learning as an Interdiscipline"). Then we describe our methodological approach to answering the above two research questions (see section "Methodology"). In section "Ph.D. programs in a Single Department or Doctoral School" we describe the heterogeneity of current doctoral training in TEL at European universities based on a survey, and present an overview on the past 15 years of the historical development of cross-institutional and interdisciplinary doctoral school in the framework of what is now the European Association of Technology Enhanced Learning (EA-TEL[1]). Finally, we conclude, also with an outline a vision for further development of the framework and connected measures of success (see section "Dedicated Doctoral Training in an International Society").

TECHNOLOGY ENHANCED LEARNING AS AN INTERDISCIPLINE

Technology enhanced learning is an interdisciplinary field that connects Computer Science with the Learning Sciences, Psychology, and other Social Sciences, Humanities, or Engineering Sciences (Meyer, 2011; Tchounikine, 2011; Meyer et al., 2013; Kalz and Specht, 2014). Wild (2016) defines TEL as being directed at "human development of competence [. . .] with tools that afford isolated or collaborative endeavors in formal and informal situations", deliberately defining TEL as inclusive for both educational as well as professional contexts.

While TEL is a standing term in European research, sometimes its related expressions are preferred internationally, such as *Educational Technology, Digital Education,* and *Learning Engineering* (see **Figure 1** for a depiction of regional preferences in terminology). The expression of Learning Engineering recognizes the need for technical competence as an essential requirement for learning and development initiatives in fields

[1] www.ea-tel.eu

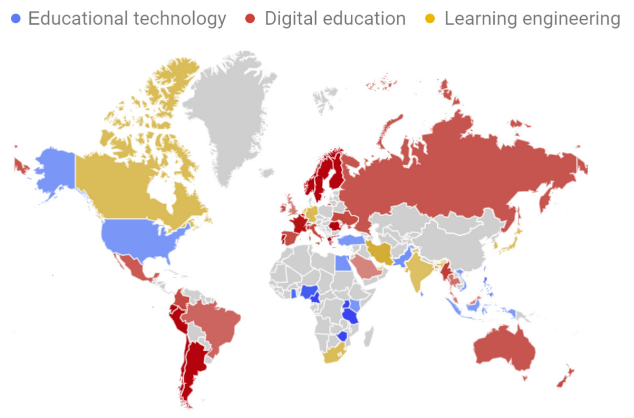

FIGURE 1 | Google search trends of competing terms by region (07/2010–07/2020). Data source: Google Trends (https://www.google.com/trends). Color intensity represents percentage of searches of most searched of the three terms (the higher the percentage, the more intense the color). Percentage means percentage of all searches of the three terms.

that methodologically depend on data science, Computer Science, and Learning Sciences.

All terms and definitions recognize the need of epistemic fluency to facilitate interdisciplinary dynamics, in which participating professionals have "the capacity to understand, switch between, and combine different kinds of knowledge and different ways of knowing" (Markauskaite and Goodyear, 2017).

Examples of Interdisciplinary Dynamics in TEL Research

To discuss aspects of interdisciplinarity for Technology enhanced learning (TEL), we first need to operationalize the different terms commonly used to describe collaboration between scientific actors. For this purpose, we build on the work of Wagner et al. (2011) who provided the definitions listed in **Table 1**.

Furthermore Kalz and Specht (2014) differentiate four distinct approaches to interdisciplinary research, based on the work of Aboelela et al. (2007):

- *Interaction-oriented*: How do members of a scientific community cooperate?
- *Communication-oriented*: How do members talk to each other?
- *Conceptual*: How are concepts, ideas and models integrated in the inquiry-phase for problem-solving?
- *Outcome-oriented*: What are the products of the cooperation?

TABLE 1 | Definitions based on Wagner et al. (2011).

Term	Definition
Transdisciplinarity	Cooperation between scientists and practitioners.
Crossdisciplinarity	Any form of scientific cooperation between scientific disciplines without any further explication of shared methods, goals, and mutual interest.
Interdisciplinarity	Collaboration where various disciplines retain autonomy (i.e., without becoming a serving discipline), but solve a given problem jointly, which cannot be solved by one discipline alone.

In this article, we understand and discuss TEL as interdisciplinary, since actors jointly address the question how technologies can be used to make learning more effective, efficient, enjoyable, or accessible. In addition, we follow in this study an *outcome-oriented approach* combined with a *conceptual approach*.

Technology Enhanced Learning (TEL) research is often connected to practical problems or "grand challenges" of education, a theory, or technological affordances. For example, it is well known that the most effective way for humans to acquire domain knowledge is by 1:1 tutoring. At the same time, however, it is simply not possible to offer a private tutor to each student, posing a grand challenge. Such practical problems very often make integration of knowledge from different disciplines necessary. In this sense, TEL is an interesting case study for an analysis of interdisciplinarity, since the work profits from mono-disciplinary research of the contributing domains while at the same time problems of TEL can only be addressed by joint work. The below examples illustrate TEL interdisciplinarity, and the feedback to the disciplines from a conceptual and outcome-oriented perspective.

As a solution to the tutoring problem, *Personalized Adaptive Learning Systems* have been conceived in the Computer Science field, using models of learning and cognitive science as input to their design. As a popular example, the "cognitive tutor" (e.g., Ritter et al., 2007) has been built around models of cognitive psychology derived from ACT-R, a general purpose cognitive architecture that explains the working of different cognitive functions like perception, memory and learning. Interdisciplinarity goes even a step further, namely when the results of the cognitive tutor's evaluation in practice is fed back to the contributing disciplines: In math education, the construction and ordering of problems and the creation of curricula has been influenced (e.g., Ritter et al., 2007). In cognitive psychology, data from large scale evaluation can now be used to validate models derived from the cognitive psychology that are commonly only studied in the laboratory, such as how concepts are formed in self-directed learning (Seitlinger et al., 2020). For intelligent systems, some general implications have been derived in terms of what models are involved (e.g., knowledge base, learner model, adaptation model), how they can be formalized and applied, e.g., by providing adaptive prompts for reflection (Fessl et al., 2017).

Another example for interdisciplinary dynamics in TEL can be found in research on *organizational workplace collaboration and learning*. By analyzing technologies used for collaboration and the resulting collaborative artifacts (e.g., shared notes, wiki pages, ontologies etc.), the knowledge maturation model was established in the Information Systems domain to describe goal directed collective learning processes and how knowledge materialized and matured in a systematic manner in organizational settings, e.g., how initial ideas are transformed into organizational improvements or new products (Maier and Schmidt, 2015). Extending the model from the perspective of the Learning Sciences, the knowledge appropriation model explained how healthcare professionals and construction workers were learning in such settings when they co-created new working practices (e.g., on treating diabetes or on applying sustainable construction techniques) by scaffolding and guiding learning at the workplace (Ley et al., 2019). Finally, the model found application in learning analytics, where it was used to analyze traces in an online collaborative learning design environment (Rodríguez-Triana et al., 2019). This allowed insights into collaborative learning and design processes that otherwise would have been difficult to observe, namely that the more teachers build on others' work, the higher was the likelihood they used the final outcome in the classroom.

Last, but not least, *wearable enhanced learning* (Buchem et al., 2019) combines technical disciplines, most notably represented through the topics of wearable computing, augmented and virtual reality, artificial intelligence, and machine learning, with social sciences, arts, and humanities, most notably represented through education, design, and social impact studies. It is only in this combination, that research and innovation in learning with wearable technology becomes possible. For example, the sensor-based Augmented Reality system for experience capture and re-enactment documented in Limbu et al. (2019) combines innovative wearable hardware and software technology based on the four-component instructional model 4C/ID that proposes to design for learning by connecting background knowledge with procedural information, and learning and practice tasks (Van Merriënboer, 2019). Another example of interdisciplinary work in wearable enhanced learning can be found in Hall et al. (2019), which combines insights about post-stroke rehabilitation from the health sciences with a feedback model from the neuro and learning sciences in its computer science and audio engineering implementation of a real-time auditory biofeedback system for (re-)learning arm trajectories.

Meta-Analysis of the Dynamics of Interdisciplinarity in the TEL Research Community

Several outcome-oriented studies have been conducted in the TEL field that have focused on publications, sometimes combined with an interaction-oriented approach. Dynamics in the research field were investigated more broadly in the past, by applying scientometric analyses or the analysis of research collaboration and funding. While these studies are now dated and were conducted in a comparatively short time-frame (publication dates 2012-2014), an artifact of funding policy, they provide evidence of how the field emerged and developed over time.

Kalz and Specht (2014) applied a *publication analysis* using 3,746 TEL publications indexed in the Web-of-Science. By comparing within-domain with outside-domain citations as the measure of diversity (weighted by disparity/variety of disciplines participating), the authors conclude that the field operates on a high level of interdisciplinarity.

Meyer et al. (2013) studied interdisciplinarity and research practice in TEL using a *survey* (N = 123), complemented with a social network analysis over publicly available information on research collaboration of the participants (who were not anonymized). The authors found diverse disciplinary backgrounds among the respondents, including social sciences, engineering, multidisciplinary backgrounds, and backgrounds in other disciplines such as life and natural sciences. A cluster analysis identified key groups among the respondents, differentiating along two axis, namely degree of TEL participation and disciplinary orientation. The motor of the community, i.e., those groups with high participation in TEL and an interdisciplinary view, is identified in three groups: established computer scientists (5%), TEL interdisciplinarians (21%), and progressive social scientists (10%). The social network analysis added to that picture that the TEL interdisciplinarians show the highest betweenness centrality value, indicating that "many others are dependent on this group in order to reach indirect contacts" (Meyer et al., 2013).

Pham et al. (2012) analyzed five major TEL conferences with the help of *social network analysis*, ICALT, AIED, EC-TEL, ITS, and ICWL. AIED and ITS exhibit mature, so-called 'focused' author communities with stable, in parts hierarchical structure and few isolates. Both conferences bridge disciplines and bring together artificial intelligence research with research in education. The publications at ICWL and ICALT are less connected than those of ITS, AIED, and ECTEL. Both ICWL and ICALT are inclusive and open to a wide range of perspectives, at the same time reflecting fragmentation of their global constituencies. This is supported by analyzing the development of the maximum betweenness values. This indicates the existence of more common core references in the scientific communities of ITS, AIED, and ECTEL. The diameters of ECTEL and AIED have begun to shrink very early, indicating that the body of literature of these communities is relatively stable and the themes of the communities are settled, reflecting the common ground that exists by now. Overall, for most TEL conferences (not ICWL) at the time, more than 35% authors continued to publish at the same venue.

Derntl and Klamma (2012) analyzed European *project funding* in TEL, and found in particular European funding for larger research networks (Integrated Projects, and Networks of Excellence in the then current funding program) served to shape the research agenda of the field, and to create strong collaborative ties between research institutions, a characteristic that reflects on the doctoral training offered in these projects.

This snapshot of the community provides evidence for the field's emergence, maturation, and its interdisciplinarity at the time. It is compelling, especially in times of a Black Swan event, the COVID-19 pandemic, where TEL is more important than ever before, that up-to-date systematic meta-analyses of the TEL research communities' interactions and collaboration are missing, and therefore, sadly, there are no up-to-date expert and expertise directories. The past analyses, however, establish enough evidence to claim the field as interdisciplinary, even if we may not fully know the current state of affairs with regards to knowledge and community integration, and research communication.

METHODOLOGY

In order to answer the first guiding research question (RQ1) about current practices of institutional doctoral training in TEL in Europe, we sampled TEL-oriented programs offered at 35 Higher Education Institutions from eleven countries in Europe. Sampling was expert-driven convenience sampling, collecting data and recommendations from collaborators. Data were extracted from websites and direct communication, looking at content, teaching methodologies, resources, and the administrative context of the programs each. The sample is non representative, but spread out enough to allow for exploration and qualitative insights into the nature of the programs offered. The 35 cases were classified inductively into the three types “Ph.D. programs in a single department or school”, “postgraduate programs (master programs) in a single department or school which offers a TEL specialization”, and “Cross-departmental or multi-disciplinary program.” This analysis investigates to what degree the field of TEL has commodified and institutionalized in form of dedicated doctoral training programs. It is an outcome-oriented analysis, where the unit of analysis is not publications, but educational curricula and administration (association to departments). This analysis is described in section “Ph.D. programs in a Single Department or Doctoral School.”

To answer the second research question (RQ2) about how cross-institutional programs can complement such institutional doctoral training, we study set-up and development of a doctoral summer school, the joint TEL summer school (JTELSS) of the European Association for Technology-Enhanced Learning (EA-TEL). We conduct an in-depth case study of this cross-institutional doctoral training program. The summer school is part of a set of doctoral training activities, complemented by a 2-day Doctoral Consortium and a Ph.D. student best paper award. Both are attached to the annual academic conference of the society, the European Conference on Technology-Enhanced Learning (EC-TEL). This analysis is described in section “Methodology.”

We analyze the programs of 15 years from the perspective of the types of sessions (keynotes, thematic sessions, methodology sessions, soft-skills sessions, informal learning sessions, career-development sessions, see section “EA-TEL Summer School Activity Framework”), and the topics of sessions (see section “Shifting of Program Topics Over Time”). The topics were identified by manual inductive coding (two experts, mediating agreement), with statistical clustering applied over the coded data to identify themes, that were subsequently intelligently labeled. Details of the clustering procedure are described in section “Shifting of Program Topics Over Time” together with the presentation of results. Again, this analysis constitutes an outcome-oriented analysis (distribution session types and topics over time, emergent framework), as well as a conceptual analysis (clustering constituting research themes).

PERTINENT INSTITUTIONAL DOCTORAL TRAINING ON TEL IN EUROPE

To answer the first research question about current practice in institutional TEL doctoral training, we summarize below the findings from the survey. The programs that were overall analyzed are listed and categorized are provided as **Supplementary Material** to this article.

Ph.D. Programs in a Single Department or Doctoral School

Ph.D. Programs in TEL (3 Cases)

An example is the Ph.D. program “Education and ICT (e-learning)” offered by the Open University of Catalonia. This program combines study and research, such that students first get training, and only in a second stage set up and carry out their doctoral research plan. Offered courses in the first phase include both methods (e.g., qualitative and quantitative research methods, or data analytics), and foundations in technology-enhanced learning. In the second phase, an additional personalized study plan is drawn up, while up to five additional blocks of training support progressing the research project (seminars, bridging courses, research/transfer/entrepreneurship courses, workshops).

Monodisciplinary (Ph.D. in Computer Science – 7 Cases; Or in Education – 6 Cases)

In these cases, groups that host TEL doctoral students do research in TEL, but the doctoral programs are not specific to TEL. An example of such a program is at Graz University of Technology, the doctoral school of Computer Science which offers a Ph.D. program in Computer Science and a number of mandatory courses (such as “methods of scientific work”) and elective courses (can be chosen from all university courses at master level, agreed by supervisor/director of studies). Doctoral students carry out their work as part of research groups.

Postgraduate Programs (Master Programs) in a Single Department or School Which Offer a TEL Specialization

Postgraduate Program in TEL (8 Cases)

In these cases, courses are particularly suitable as foundation for a Ph.D. in TEL, and courses connected to relevant research groups / dedicated Ph.D. programs. An example case is the postgraduate course “Educational Technology” offered by the University of Tartu, which provides professional development to people who teach (or plan to teach), and are interested in how to use educational technology in their work. After successful completion, students have the possibility to continue to a Ph.D. program with the same specialization.

Monodisciplinary Postgraduate Program (Computer Science: 7 Cases; Education: 5 Cases)

Cyprus University of Technology, for example, offers a postgraduate program on “Interaction Design” and its graduates can continue their studies for a Ph.D. in various TEL-related areas such as Embodied Play and Learning using Technology, Interaction Design and Creative Collaboration Spaces, Inclusive Design and Social Change using Technology, Design for social change and innovation, and Computer-Assisted Language Learning.

Cross-Departmental or Multidisciplinary Programs

Cross-departmental or multidisciplinary programs (Ph.D.: 2 cases; Postgraduate: 7 cases): An example is at the University of Aveiro which offers a Ph.D. program in "Multimedia in Education," a joint degree offered by the Communication and Arts and the Department of Education and Psychology.

Summary of the Pertinent Institutional Doctoral Training in Europe

Dedicated Ph.D. and postgraduate programs in TEL [see sections "Ph.D. Programs in TEL (3 Cases)" and "Postgraduate Program in TEL (8 Cases)"] are examples of institutionalization. They are interdisciplinary. Both Ph.D. and postgraduate programs in other fields [see sections "Monodisciplinary (Ph.D. in Computer Science – 7 Cases; or in Education – 6 Cases)" and "Monodisciplinary Postgraduate Program (Computer Science: 7 Cases; Education: 5 Cases)] are examples of non-institutionalized TEL, meaning that they are not institutionalized at all in the respective Higher Education Institution, and training is mono-disciplinary. There are also cross-departmental or multidisciplinary Ph.D. program instances where TEL has been operationalized as cooperation or collaboration between departments and disciplines.

Most doctoral programs investigated focus on methodological courses, rather than on TEL-specific topics or topics specific to another discipline. The programs studied, however, differ widely, meaning they provide so heterogeneous foundational knowledge. From the insights on doctoral training within Higher Education Institutions studied, we have to conclude that the creation of common ground for the field of TEL is not happening from inside these institutions (with limited exceptions). Furthermore, not all Ph.D. programs are available in English, further limiting the sharing of existing resources.

DEDICATED DOCTORAL TRAINING IN AN INTERNATIONAL SOCIETY

To answer the second research question on how cross-institutional doctoral training could be organized and which topics would be relevant, we have analyzed the program development of a joint European summer school on TEL, organized by EA-TEL. The summer school is typically organized in a rather remote location, so as to underline its retreat character of providing a protective, low-exposure environment for next-generation researchers (compared to a big scientific conference).

The summer school itself is evidence that interdisciplinary Ph.D. training works, managing now for 15 years to bring together next generation researchers with very heterogeneous training, as the application and evaluation processes year after year reveal. Some Ph.D. candidates have no training, while some have a lot of course work. Most Ph.D. candidates attending the summer school are in the early stage of their Ph.D. work (main target group), some are late stage (often co-organizing workshops/part of the organization team). The doctoral consortium is more oriented-toward late stage Ph.D. candidates getting ready for their Ph.D. exam.

In the application process, all Ph.D. candidates are required to submit a summary of their research work. These summaries are peer-reviewed by an international committee of established researchers, following criteria similar to those at academic conferences, such as evaluation of the related work, theoretical framework, methods, and progress. The reviews do not only provide Ph.D. candidates with unbiased feedback, but often urge them to update their own understanding of all components of their work. Most Ph.D. candidates demonstrate a good level of awareness of their selected topic, but many (and not those in the early-stage) are struggling with defining their theoretical frameworks. The evaluation of the methods vary greatly from excellent rates to questioning the overall research design. Some are trying to run a project, lacking research questions, only pursuing development work. Others are misguided to study and reflect upon local TEL efforts in their own institution only, rather than working in a manner conducive to receiving the international recognition (and impact) required for a doctoral degree.

While the doctoral training targets directly Ph.D. candidates, it also indirectly supports Ph.D. advisors. A computer science expert in machine learning could benefit from having a Ph.D. candidate investigating, for example, the design and the impact of educational chatbots liaising with the TEL community. This is a way to stimulate frontier research by supporting Ph.D. advisors in tackling challenges outside their comfort zone.

All three EA-TEL doctoral training instruments share the objective of the society to establish a universal concept of what a Ph.D. in TEL should look like, connecting the community by establishing a review and quality assurance process, reinforcing reflection on how the developed technologies actually serve learning. Review thereby includes both peer review from candidates at other universities as well as from established researchers, using the main conference, EC-TEL, as a recruiting ground. The best student paper award serves as a showcase of what excellence in TEL research looks like.

EA-TEL Summer School Activity Framework

The program of activities offered at the summer school changes annually, reacting to evaluation results of the previous edition, while also implementing new and experimental ideas. Nevertheless, over the years, a stable common framework has emerged, which covers six distinct session types (**Figure 2**). The program of the summer school is compiled combining sessions selected from submissions to an open call for instructors with 'standard' sessions from the framework, added by the

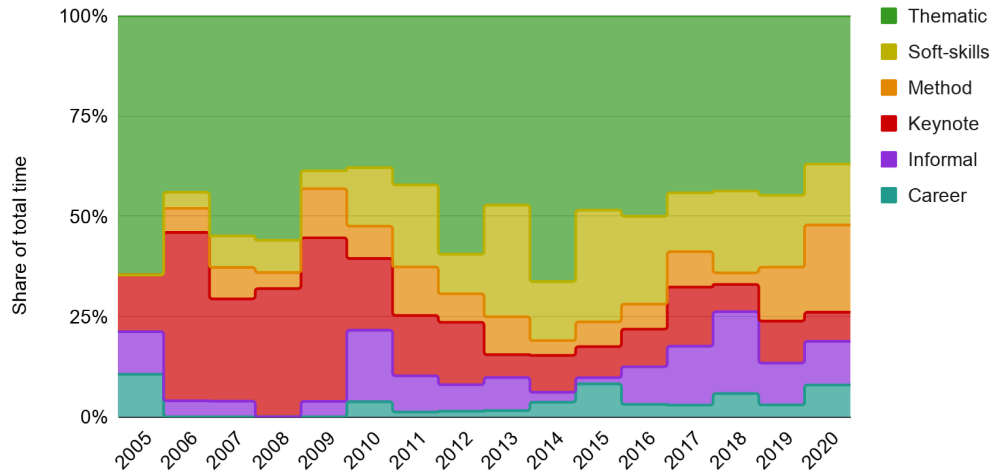

FIGURE 2 | Share of total time of different session types against years.

organizers. All thematic workshops are proposed via the call for instructors, whereas, typically, all keynotes and informal sessions are added by the organizers. Methodological, soft-skills, and career workshops are mixed-initiative. Some workshops come through via the open call, some are added by the organizers. Ph.D. candidates co-design the program, submitting workshops, often collaboratively with supervisors or peers.

Thematic workshops change year by year, driven by community interest. These workshops serve as indicators not only for the development of the program, but for the TEL field at large. The share of sessions dedicated to these workshops in the program has been stable over the years: in average, 47.8% of the time of the program is spent on thematic workshops (see **Figure 2**).

Methodological workshops are mostly proposed by the community. They focus on different research methodologies that can be applied in various contexts of TEL research, such as systematic reviews, resign-based research, statistics for TEL, field studies, and many others. They make up for 8.9% of the program in average (time-wise). The participants highly appreciated these workshops and their number grew to 13% in 2019 and 22% in 2020.

Soft-skills workshops remain relatively stable over time, even though they are proposed via the open call. They cover topics such as academic writing, dissemination and communication, or presentation skills, making up for 15.5% of the total time in average.

Established researchers present *keynotes* covering central themes as well as frontiers topics. In early years, the summer school used "lecture" type sessions submitted to the open call (these sessions were categorized as keynotes for the subsequent analysis). From the early 2010s, instructors were encouraged to focus on interaction rather than lecturing. In 2018, it was decided to accept only interactive workshops via the open call, removing the lecture category from the open submission process, while at the same time increasing the number of keynotes and managing the speakers drafting process centrally through invitation. Since 2018, keynotes are rated higher by the participants than any other session type. Keynotes make up for 15.6% of the total time.

Informal learning sessions have been refined over the years, staying relatively stable in recent years. They include Ice-breaker, Pecha Kucha, Pitch and Poster Session, Fish Bowl, Game Night, and Speed Mentoring. They encourage active participation and allow participants to present their work, bring up their questions and challenges, without any restriction to specific topics. These sessions play a key role in developing strong ties in the community, contributing to the social atmosphere of the event. Informal sessions make up 8.8% of the program.

Career workshops usually target late-stage Ph.D. candidates and focus on opportunities for Ph.D. graduates in both academia and business. They make out 3.4% of the program.

Overall, the activity framework provides a mix of structured regular activities combined with a dynamic community-driven curriculum. Moreover, it offers a networking venue for the TEL research community. Instructors value the opportunity to disseminate research results, promote publications and projects (and write new ones), and share knowledge. Ph.D. candidates value networking with peers at the *informal learning sessions* and between the sessions. In the past three years, 70–77% of Ph.D. candidates named "Discovering topics of other Ph.D. candidates" among the top three most beneficial aspects of the event in terms of learning, followed by "learning about TEL state of the art" from *keynotes* and *thematic workshops* (55–66%). In the past 3 years, 89–100% of Ph.D. candidates named "Networking with other Ph.D. candidates" among the top three benefits in terms of professional development, followed by "Networking with TEL experts" (74–78%).

Shifting of Program Topics Over Time

To investigate how topics shift in the program, we expert coded (two experts, mediating agreement) all thematic workshops and keynote sessions from the past 15 years. Each session could have multiple codes. Coding was performed inductively, starting with the first session in the first year, and adding

TABLE 2 | Clusters of topics from the thematic workshops and keynotes.

#	*Cluster*
1	Personalized, Contextualized, and Adaptive Learning
2	Pervasive, Immersive, and Social Learning
3	Organizational Learning
4	Learning Environments
5	Wearable Enhanced Learning
6	Open Education

new codes (or extending existing ones) as we went along in chronological order. The resulting matrix is sparse, and therefore was tabulated by years in order to allow for cluster analysis. The full distribution of codes over the 15 years is provided in the **Supplementary Material**.

We excluded topics that occurred only up to three times, based on the assumption that their low appearance frequency will inevitably lead to artifacts of a cluster analysis. We converted the tabulated data for topics by years to distances, testing Jacquard distance against binary distance measures, and testing clustering structure by inspecting the agglomerative coefficients for average, single, complete, and ward clustering methods for agglomerative nesting (agnes, package "cluster," Maechler et al., 2019). Binary distances and Ward's method came out top. Depending on granularity aimed for, the cluster prediction measures we consulted (Charrad et al., 2014; Kassambara and Mundt, 2019) favored – after the initial peak of 2 or 3 clusters for overview – between 6 and 8 (gap statistics), 4 and 8 (average silhouette), 6 and 8 (second differences Dindex). We ran multiple combinations, inspecting the homogeneity of the clusters via their dendrogram height, and settled on six clusters as a useful level of analysis. For the full cluster dendrogram, see the second plot in the **Supplementary Material**. The resulting clusters were labeled in agreement by the two human analysts (**Table 2**).

Overall, the topics of the thematic workshops change year after year, and – by their interactive nature, enforced particularly in recent years – they are more catalysts to community building than knowledge exchange. In the end, you cannot teach something that has not been invented yet. Below, we first describe each cluster on its own, and then show how it is possible to distinguish the clusters along two axes, from personal to organizational, and from knowledge to behavior in a cluster plot (**Figure 3**); and how the topics have developed over time (**Figure 4**).

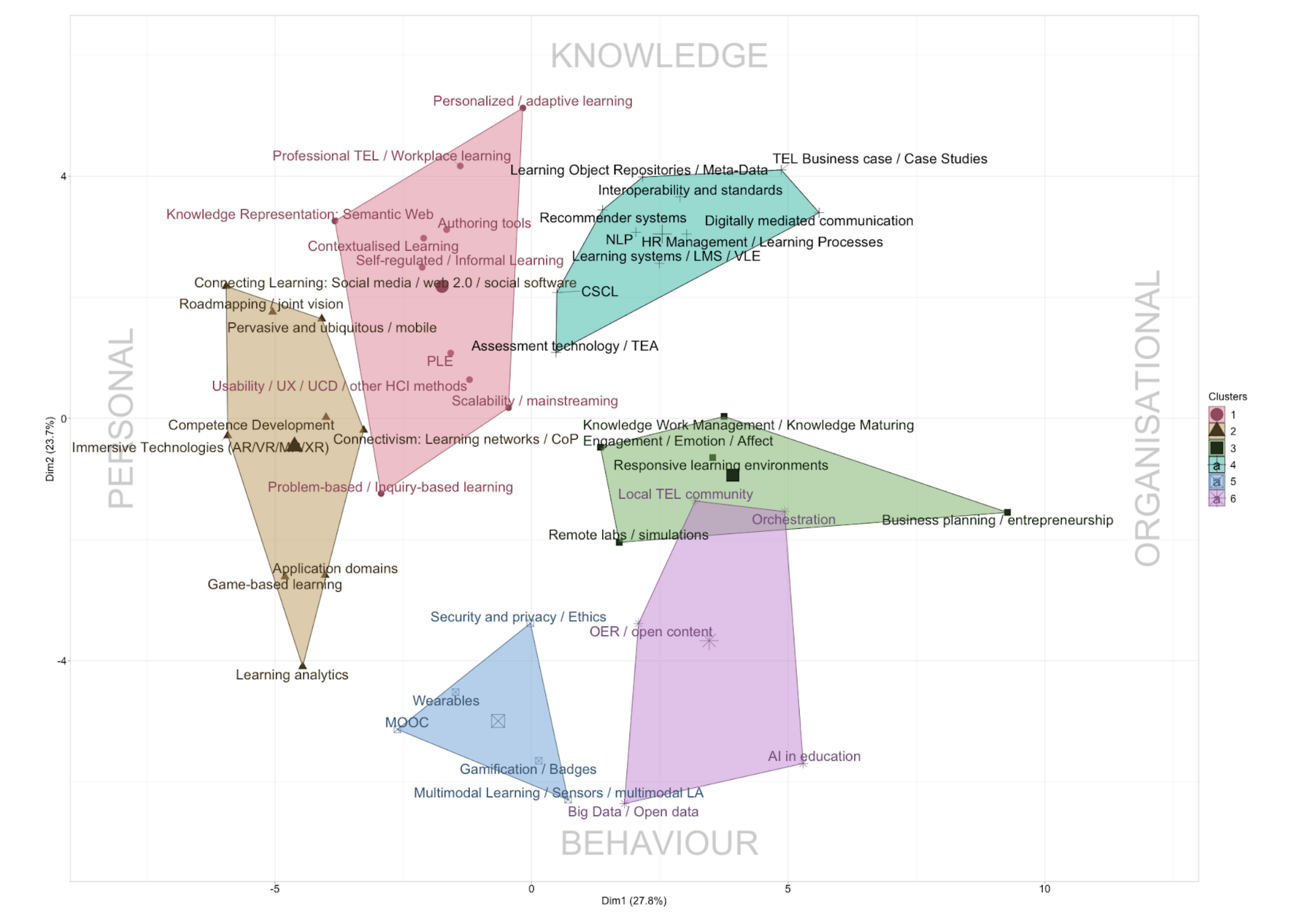

FIGURE 3 | Cluster plot of the session topics.

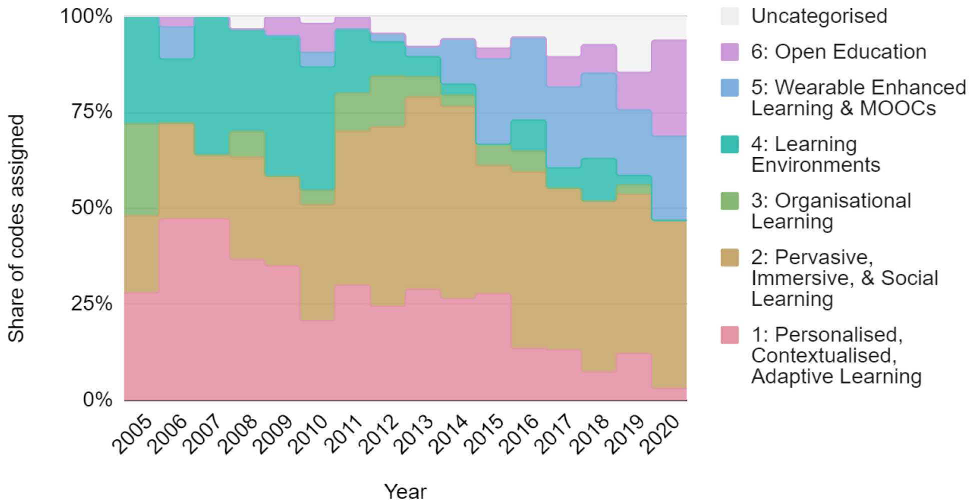

FIGURE 4 | Shift of session topics over the years (Uncategorized comprises those sessions that were coded with codes that appeared overall only up to three times).

Cluster 1 – Personalized learning makes up for 25.4% of all session codes (see **Figure 4**). It is the dominant topic in the decade from 2005 to 2015, less prominent though in later years. The initial focus in early years on adaptation, authoring tools, and knowledge representation is extended with personal learning environments and contextualized learning, extending approaches with a behavioral (associationist) perspective. Self-regulated learning replaces the debate around informal learning.

Cluster 2 – Pervasive learning contains 36.0% of all session codes (see **Figure 4**). The topics in the cluster appear most often in the summer school programs, but change character over time. While early years focus on mobile learning, middle years emphasize the social character of learning, acknowledging the connectivist perspective, consequently adding learning analytics as a strong subtopic around 2012. From 2010, game-based and immersive learning became very popular themes at the summer school.

Cluster 3 – Organizational Learning contains 4.5% of all session codes (see **Figure 4**). In this cluster, Knowledge Maturing replaces Knowledge Work Management from early years, adding a new approach. In both cases, the focus on knowledge is complemented with a behavior perspective, looking at management of the people producing knowledge. Consequently both topics are located almost half way both from the knowledge axis extreme and its behavior counterpart. Responsiveness as a new principle is added in middle years, leveraging engagement and emotion/affect to particularly support professional contexts.

Cluster 4 – Learning environments contains 15.7% of all session codes (see **Figure 4**). The theme was very popular from 2005 to 2009. It started to decline in 2010 and almost disappeared as a topic, even after meta-data information extraction and learning object repositories naturally led to recommender systems, and natural language processing to technology-enhanced assessment. Human resource management as a topic was taken over in early years by Learning Processes.

Cluster 5 – Wearable Enhanced Learning contains 9.2% of all session codes (see **Figure 4**). A new focus on *wearable enhanced learning* (using accessories, headworn devices, and smart garments with embedded sensors) and on multimodal learning emerges in the last few years as a cluster on its own. The agglomerative clustering merges this with the added focus on massive open online courses (MOOCs) and gamification/badges growing from 2011. Both reflect the renewed interest to observe learner behavior beyond the cognitive, but we disagree with the grouping of the automated analysis and argue, also from the positions in the cluster plot (**Figure 3**), that MOOC and Gamification/Badges should be grouped together with Cluster 6, Open Education. We could imagine that wearable enhancing learning and multimodal learning/LA could additionally be grouped together with immersive technologies. Future years will reveal how the cluster structure changes, and – in our view – undoubtedly remediate the clustering artifact.

Cluster 6 – Open Education contains 4.5% of all session codes (see **Figure 4**). Open Content and, later, Open Data focus on *Open Education* for all. Recent years add a renewed focus on the use of artificial intelligence (Alexa/Siri/Cortana like learning assistants). As mentioned above, we propose to move MOOCs and Gamification/Badges into this cluster, as they are closer to the Open Education theme.

CONCLUSION AND DISCUSSION

We established that technology enhanced learning (TEL) is a complex field with a plethora of perspectives that benefits from disciplinary dynamics. A major challenge in advancing the field is therefore to provide suitable community spaces in which these dynamics can unfold. It is necessary for researchers to have epistemic fluency and understand sufficiently the field in order to profoundly contribute to these dynamics, while at the same time contributing rigorously to the state of the art. On this background, establishing a viable frame of reference for doctoral training in TEL is a key puzzle piece, required to drive forward the commodification of the field. Such a frame of reference is supposed to secure the interdisciplinary common ground within the field for the next generation researchers, while building shared understanding among the already established researchers.

Up to now, this frame of reference does not exist; and we perceive that the integration of interdisciplinary knowledge and ways of knowing is still ongoing. We found that this is reflected particularly in the way TEL doctoral training is organized on the institutional level, i.e., on the level of universities; answering our first research question (RQ1). Our non-exhaustive study provides evidence that TEL doctoral training is fragmented: While some next generation researchers receive interdisciplinary training already in their home institution or via the cross-institutional doctoral school, others remain trained in a monodisciplinary way. We believe that to overcome this fragmentation, the prerogative must be to not only connect isolated Ph.D. students better, but also their supervisors, directors of studies, and institutions. It requires community efforts to build and sustain professional social connections. Conferences, workshops and symposia are the traditional networking events. Existing networks, however, and the existing review criteria pose barriers for early-stage researchers for building their own social network. Doctoral summer schools and doctoral consortia are established instruments and effective tools to remediate that and support next generation researchers in developing their own social and intellectual links within the TEL community, even before publishable results are available. Both formats allow early stage as well as senior members of the community to connect over their work and sustain a continuous discourse.

Over the past 15 years, the summer school was the catalyst for moderating this perpetual discussion about the core and emerging topics, explicitly reflecting the interdisciplinary foundations of the field. Our insight from these past summer schools is that maintaining this dialogue community-driven and bottom-up is possible. The summer school provides a clear structure using an overarching activity framework to integrate the organically grown thematic structure into a complex learning experience. Despite the flexibility of this framework, it does not offer an *a priori* definition of basic and elective subjects (yet). Therefore, we propose that the identified structure could serve as input to a further refined cross-institutional curriculum. This structure requires expert agreement found in curricular commissions involving all key stakeholders in order to clarify the mandatory and optional elements for training researchers in the field of TEL. Such an offer can help institutions to overcome potential local shortcomings, while preserving the bottom-up prevalent community-driven, culture. This answers our second research question (RQ2) about the need and characteristics of complementary cross-institutional training activity for TEL doctoral-level research.

Beyond cross-institutional doctoral education activities, we identified the scarcity of shared educational resources as a key gap in current TEL doctoral training. Earlier efforts regarding TEL OER do exist, resulting in a TEL dictionary; and a collection of educational resources at doctoral level, albeit with a stronger emphasis on general learning sciences than having a TEL-specific focus. These existing efforts need to be updated and extended for adapting to latest developments of a constantly changing field. OER are useful tools for doctoral education beyond their educational use, for example by involving early stage researchers as authors in the participatory development of resources and concepts. This aims at lowering the barriers for early stage researchers to leave their mark in our interdisciplinary community, by codifying and preserving the established common ground. In parallel to creating a stronger base in OERs, the element of openness can be extended toward open science in the broader sense. The TEL field needs more open data gathering, curation, sharing, and re-use activities that strengthen evidence-based research, while complying with data protection regulations. Doctoral education is the perfect place to promote and discuss the practices of open science within TEL.

Finally, beyond TEL doctoral training, strategic and integrative activities exist of course that contribute to bringing TEL forward as an interdisciplinary field by bridging across communities. Examples are roadmapping and observatory initiatives like, e.g., the "Innovating Pedagogy" reports of the Open University of the United Kingdom, the "Horizon Reports" formerly of the New Media Consortium, now as part of Educause, or the "Emerging state of XR and Immersive Learning" report of the Immersive Learning Research Network. Complementing such observatory activities as integrative across institutions and communities with overlapping interests, a liaison across a number of scientific societies in TEL and closely related fields, the International Alliance to Advance Learning in the Digital Era (IAALDE), has been set-up, which fosters cross-fertilization across a broad range of research communities by exchanging best papers between conferences.

This study was limited several ways. First, the review of scientometric meta-analysis of the field is dated and more of historic value, shedding light on the foundation of TEL. We hope that with this article, we contribute to a long-overdue update, but we also have to acknowledge that the scope of our analysis was rather limited, sacrificing breadth (TEL community at large) for depth (TEL doctoral training). Moreover, we limited our analysis of educational doctoral training programs geographically to Europe. To overcome these limitations, we propose to study in more depth, also in quantitative ways, the existing TEL doctoral training *globally*, including ways of promoting cross-institutional structures and resources. Future investigations should also help account for career paths of TEL alumni as academics, EdTech entrepreneurs, or executives in the knowledge-oriented economy. Additionally, further analysis as to how to teach students with diverse educational backgrounds and how to overcome inevitable problems would serve useful to the field.

This would not only help to document another snapshot of the field with a focus on doctoral education. It would additionally serve to inform professional societies, Higher Education Institutions, as well as beneficiaries of TEL alike about the composition of the field, the current state of the art, and about the human talent available. Such stock-taking would ensure the world is better prepared for lock-down imposed by a pandemic such as COVID-19, where technology enhanced learning is the only viable option for education and training at large. Ultimately, this could also help TEL research to have a higher impact on curricula in teacher training: In the end, TEL research is essential in supporting policy makers with the ambitious goal of the digitization of society.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: The data are partially available in the article/**Supplementary Material**, in particular, the full list of all summer school sessions is available on request. Requests to access these datasets should be directed to mikhail.fominykh@ntnu.no.

AUTHOR CONTRIBUTIONS

VP-S contributed to lead manuscript-writing, overall edits, including structure and main line of argumentation, and discussions. FW contributed to overall manuscript-writing, manuscript strategy, including structure and main line of argumentation, discussions, and summer school analysis. MF contributed to overall manuscript-writing and 15 years of EATEL summer school analysis. TL contributed to overall manuscript-writing and discussions, and focus on TEL as an interdiscipline and examples. MP contributed to coordination of the research/writing of section "Methodology" TEL Doctoral Training in European Universities. MS contributed to discussions, focus on section "Methodology" TEL Doctoral Training in European Universities. DH-L contributed to sections "Introduction", "Technology Enhanced Learning as an Interdiscipline", and "Conclusion". MK contributed to discussions, focus on definitions of interdisciplinarity, and meta-analysis of the dynamics of the field. RK contributed to discussions and meta-analysis of the dynamics of the field. LP and CS contributed to describing the relevance of doctoral training. CG and AP contributed to doctoral education in Europe in TEL. AE contributed to discussions and description of section "Methodology" TEL Doctoral Training in European Universities. EP-F contributed to discussion. DG and KM: contributed to 15 years of EATEL summer school analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the European Commission under the Erasmus+ Program, as part of the DE-TEL project (Grant Agreement No. 2019-1-NO01-KA203-060280) and under the Horizon 2020 Program, as part of the ARETE project (Grant Agreement No. 856533).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2020.00150/full#supplementary-material

REFERENCES

Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., et al. (2007). Defining interdisciplinary research: conclusions from a critical review of the literature. *Health Serv. Res.* 42, 329–346. doi: 10.1111/j.1475-6773.2006.00621.x

Amran, N. N., and Ibrahim, R. (2012). Academic rites of passage: reflection on a PhD journey. *Procedia Soc. Behav. Sci.* 59, 528–534. doi: 10.1016/j.sbspro.2012.09.310

Buchem, I., Klamma, R., and Wild, F. (2019). "Introduction to wearable enhanced learning (WELL): trends, opportunities, and challenges," in *Perspectives on Wearable Enhanced Learning (WELL): Current Trends, Research, and Practice*, eds I. Buchem, R. Klamma, and F. Wild (Cham: Springer).

Carr, G., Loucks, D., and Blösch, G. (2018). Gaining insight into interdisciplinary research and education programmes: a framework for evaluation. *Res. Policy* 47, 35–48. doi: 10.1016/j.respol.2017.09.010

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36.

Cleverdon, C. W. (1960). The aslib cranfield research project on the comparative efficiency of indexing systems. *ASLIB Proc.* 12, 421–431. doi: 10.1108/eb049778

Derntl, M., and Klamma, R. (2012). "The European TEL projects community from a social network analysis perspective," in *Proceedings of the 21st Century Learning for 21st Century Skills. 7th European Conference on Technology Enhanced Learning, EC-TEL 2012*, eds A. Ravenscroft, S. Lindstaedt, C. D. Kloos, and D. Hernández-Leo (Berlin: Springer Verlag), 51–64. doi: 10.1007/978-3-642-33263-0_5

Fessl, A., Wesiak, G., Rivera-Pelayo, V., Feyertag, S., and Pammer, V. (2017). In-app reflection guidance: lessons learned across four field trials at the workplace. *IEEE Trans. Learn.Technol.* 1, 488–501. doi: 10.1109/tlt.2017.2708097

Hall, S., Wild, F., and Olde Scheper, T. (2019). "Real-time auditory biofeedback system for learning a novel arm trajectory: a usability study," in *Perspectives on Wearable Enhanced Learning (WELL) Current Trends, Research, and Practice*, eds I. Buchem, R. Klamma, and F. Wild (Cham: Springer).

Herodotou, C., Villasclaras-Fernández, E., and Sharples, M. (2014). "The design and evaluation of a sensor-based mobile application for citizen inquiry science investigations," in *Open Learning and Teaching in Educational Communities. EC-TEL 2014. Lecture Notes in Computer Science*, Vol. 8719, eds C. Rensing, S. de Freitas, T. Ley, and P. J. Muñoz-Merino (Cham: Springer).

ISCED (2011). *International Standard Classification of Education*. Paris: UNESCO Institute for Statistics.

Kalz, M., and Specht, M. (2014). Assessing the cross-disciplinarity of technology-enhanced learning with science overlay maps and diversity measures. *Br. J. Educ. Technol.* 45, 415–427. doi: 10.1111/bjet.12092

Kassambara, A., and Mundt, F. (2019). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.6*.

Kittur, A., Chi, E. D. H., and Suh, B. (2008). "Crowdsourcing user studies with mechanical turk," in *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, New York, NY: ACM, 453–456.

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions. 50th Anniversary Edition*. Chicago, IL: University of Chicago press.

Latour, B. (2005). *Reassembling the social. An introduction to actor-network-theory*. Oxford: Oxford University Press.

Lave, J., and Wenger, E. (1995). *Situated Learning - Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Ley, T., Maier, R., Thalmann, S., Waizenegger, L., Pata, K., and Ruiz-Calleja, A. (2019). A knowledge appropriation model to connect scaffolded learning and knowledge maturation in workplace learning settings. *Vocat. Learn.* 13, 91–112. doi: 10.1007/s12186-019-09231-2

Limbu, B., Vovk, A., Jarodzka, H., Klemke, R., Wild, F., and Specht, M. (2019). "WEKIT.one: a sensor-based augmented reality system for experience capture and re-enactment," in *Transforming Learning with Meaningful Technologies*, eds M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, and J. Schneider (Cham: Springer).

Lindvig, K. (2018). The implied PhD student of interdisciplinary research projects within monodisciplinary structures. *Higher Educ. Res. Dev.* 6, 1171–1185. doi: 10.1080/07294360.2018.1474343

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). *cluster: Cluster Analysis Basics and Extensions. R Package Version 2.1.0*.

Maier, R., and Schmidt, A. (2015). Explaining organizational knowledge creation with a knowledge maturing model. *Knowl. Manag. Res. Pract.* 13, 361–381. doi: 10.1057/kmrp.2013.56

Markauskaite, L., and Goodyear, P. (2017). *Epistemic Fluency and Professional Education*. Dordrecht: Springer.

Meyer, P. (2011). *Technology-Enhanced Learning' as an Interdisciplinary Epistemic Community*. master thesis, University of Augsburg, Augsburg.

Meyer, P., Kelle, S., Ullmann, T. D., Scott, P., and Wild, F. (2013). "Interdisciplinary cohesion of TEL – an account of multiple perspectives," in *Scaling up Learning for Sustained Impact. EC-TEL 2013. Lecture Notes in Computer Science*, Vol. 8095, eds D. Hernández-Leo, T. Ley, R. Klamma, and A. Harrer (Berlin: Springer).

O'Rourke, M., Crowley, S., and Gonnerman, C. (2016). On the nature of cross-disciplinary integration: a philosophical framework. *Stud. History Philos. Biol. Biomed. Sci.* 56, 62–70. doi: 10.1016/j.shpsc.2015.10.003

Pham, M. C., Derntl, M., and Klamma, R. (2012). Development patterns of scientific communities in technology enhanced learning. *Educ. Technol. Soc.* 15, 323–335.

Ritter, S., Anderson, J. R., Koedinger, K. R., and Corbett, A. (2007). Cognitive tutor: applied research in mathematics education. *Psychon. Bul. Rev.* 14, 249–255. doi: 10.3758/BF03194060

Rodríguez-Triana, M. J., Prieto, L. P., Ley, T., de Jong, T., and Gillet, D. (2019). "Tracing teacher collaborative learning and innovation adoption : a case study in an inquiry learning platform," in *Proceedings of the CSCL 2019 - International Conference of Computer Supported Collaborative Learning*, France, doi: 10.22318/cscl2019.432

Seitlinger, P., Bibi, A., Uus, Õ, and Ley, T. (2020). "How working memory capacity limits success in self-directed learning: a cognitive model of search and concept formation," in *Proceedings of the 10th International Conference on Learning Analytics & Knowledge*, New York, NY: AMC.

Shneiderman, B. (2008). Science 2.0. *Science* 319, 1349–1350.

Tchounikine, P. (2011). *Computer Science and Educational Software Design*. Berlin: Springer.

Van Merriënboer, J. J. G. (2019). *). The Four-Component Instructional Design Model: An Overview of its Main Design Principles*. Maastricht: Maastricht University.

Wagner, C., Roessner, J., Bobb, K., Thompson Klein, J., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. *J. Infometr.* 5, 14–26. doi: 10.1016/j.joi.2010.06.004

Wild, F. (2016). *Learning Analytics in R*. Berlin: Springer.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 August 2020
doi: 10.3389/fpsyg.2020.02190

Contrasting Classical and Machine Learning Approaches in the Estimation of Value-Added Scores in Large-Scale Educational Data

Jessica Levy[1], Dominic Mussack[2], Martin Brunner[3], Ulrich Keller[1], Pedro Cardoso-Leite[2] and Antoine Fischbach[1]*

[1] Luxembourg Centre for Educational Testing, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [2] Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [3] Department of Education, University of Potsdam, Potsdam, Germany

There is no consensus on which statistical model estimates school value-added (VA) most accurately. To date, the two most common statistical models used for the calculation of VA scores are two classical methods: linear regression and multilevel models. These models have the advantage of being relatively transparent and thus understandable for most researchers and practitioners. However, these statistical models are bound to certain assumptions (e.g., linearity) that might limit their prediction accuracy. Machine learning methods, which have yielded spectacular results in numerous fields, may be a valuable alternative to these classical models. Although big data is not new in general, it is relatively new in the realm of social sciences and education. New types of data require new data analytical approaches. Such techniques have already evolved in fields with a long tradition in crunching big data (e.g., gene technology). The objective of the present paper is to competently apply these "imported" techniques to education data, more precisely VA scores, and assess when and how they can extend or replace the classical psychometrics toolbox. The different models include linear and non-linear methods and extend classical models with the most commonly used machine learning methods (i.e., random forest, neural networks, support vector machines, and boosting). We used representative data of 3,026 students in 153 schools who took part in the standardized achievement tests of the Luxembourg School Monitoring Program in grades 1 and 3. Multilevel models outperformed classical linear and polynomial regressions, as well as different machine learning models. However, it could be observed that across all schools, school VA scores from different model types correlated highly. Yet, the percentage of disagreements as compared to multilevel models was not trivial and real-life implications for individual schools may still be dramatic depending on the model type used. Implications of these results and possible ethical concerns regarding the use of machine learning methods for decision-making in education are discussed.

Keywords: value-added modeling, school effectiveness, machine learning, model comparison, longitudinal data

OPEN ACCESS

Edited by:
Arthur C. Graesser,
The University of Memphis,
United States

Reviewed by:
Jesús-Nicasio García-Sánchez,
Universidad de León, Spain
Joni Tzuchen Tang,
National Taiwan University of Science and Technology, Taiwan

****Correspondence:***
Jessica Levy
jessica.levy@uni.lu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *12 May 2020*
Accepted: *04 August 2020*
Published: *21 August 2020*

Citation:
Levy J, Mussack D, Brunner M, Keller U, Cardoso-Leite P and Fischbach A (2020) Contrasting Classical and Machine Learning Approaches in the Estimation of Value-Added Scores in Large-Scale Educational Data. Front. Psychol. 11:2190. doi: 10.3389/fpsyg.2020.02190

INTRODUCTION

Value-Added Modeling

Value-added (VA) models are statistical models designed to estimate school (or teacher) effectiveness based on students' achievement. More specifically, they intend to estimate the "value" specific schools (or teachers) add to students' achievement, independently of students' backgrounds (e.g., Amrein-Beardsley et al., 2013). VA scores are estimated by juxtaposing the actual achievement attained by students attending a certain school with the achievement that is expected for students with the same starting characteristics (e.g., pretest scores).

The use of VA models is highly consequential because VA scores are often used for accountability and high-stakes decisions to allocate financial or personal resources to schools (for an overview from a more economical point of view, see, Hanushek, 2019). These high stakes can make estimating VA scores a politically charged topic, especially in the US, where many states have implemented VA-based evaluation systems (Amrein-Beardsley and Holloway, 2017; Kurtz, 2018). In 2015, 41 states recommended the use of VA scores or other student growth measures for human resource decisions (Kurtz, 2018). However, in recent years, the consequential use of VA models seems to be decreasing again in many states (Close et al., 2020).

Despite both the practical and political relevance, there is currently no consensus on how to best estimate VA scores (Everson, 2017; Levy et al., 2019). This lack of consensus can be observed for various aspects of the VA model, including the applied statistical model, methodological adjustments (e.g., for measurement error or missing data), and the selection of covariates used to compute the VA score. In the present study, we focused on how the choice of the statistical model affects VA scores, which are used to evaluate the effectiveness of schools, as VA scores and thus measures of school effectiveness may vary greatly depending on the statistical method used (e.g., Sloat et al., 2018). In particular, we compare classical models for prediction with those drawn from machine learning.

Statistical Models for the Estimation of School VA Scores

While VA models stem originally from economics (Hanushek, 1971), they consist of statistical methods that are common in educational or psychological sciences. Even though there are many different possible statistical models, all VA models follow the same logic. As shown in Eq. 1, this expected achievement $\hat{y}$ is estimated for every student i in school j as a function f (e.g., linear regression) of their initial characteristics x_{ij} at an earlier time point (e.g., prior achievement) and an error term e_{ij}.

$$\hat{y}_{ij} = f\left(x_{ij}\right) \tag{1}$$

In a second step and as demonstrated in Eq. 2, the VA score for each school j is estimated by calculating the mean difference (i.e., residuals) between the expected achievement $\hat{y}$ and the actual achievement y for all n students in this school j. This is equal to the average error term e of all students in school j.

$$VA_j = \frac{\sum_i^j \left(y_{ij} - \hat{y}_{ij}\right)}{n_{ij}} = \frac{\sum_i^j \left(e_{ij}\right)}{n_{ij}} \tag{2}$$

Positive $\mathrm{VA_j}$ values mean that students in school j achieved better than expected, while negative $\mathrm{VA_j}$ values mean that they achieved worse than expected. The aim is to statistically eliminate all factors that cannot be influenced by a school, such that everything that is left (i.e., the residuals) will be attributed to the effect of a certain school. Hence, the quality of the initial prediction step (i.e., Eq. 1) is crucial for the estimate.

Classical Approaches in the Estimation of Value-Added Scores

There are currently two main classical models to compute VA scores: linear regression and multilevel models (Kurtz, 2018; Levy et al., 2019). These models are often claimed to be interpretable for most researchers and practitioners (see, e.g., Molnar, 2020). However, they make strong assumptions (e.g., linearity), which may limit their accuracy. Intuitively, most people would agree that learning does not happen linearly (e.g., as illustrated in this blog entry, McCrann, 2015). This is underlined by findings that at least in some cases, non-linear models fit the data better than linear models, implying that the typical linearity assumption might not be warranted (Lopez-Martin et al., 2014). However, this does not necessarily mean that non-linear models are also more appropriate for the estimation of VA scores. For example, one finding from a national project on VA modeling was that even though the data fit was better when using a curve rather than a straight line, this had almost no effect on VA scores (Fitz-Gibbon, 1997). In situations with high noise, low model complexity can have better performance (Friedman et al., 2001), but as data quality and amount improve more complex methods may be more appropriate.

Machine Learning Approaches for the Estimation of Value-Added Scores

In educational research, as in many other domains, the amount of available data is consistently growing (as reflected in the development of the new domain of "educational data mining"; see, e.g., Romero and Ventura, 2010; Baker, 2019). Although big data is not new in general, it is relatively new in the realm of social sciences and education, requiring new data analytical approaches. This is both a challenge and an opportunity, as it is becoming feasible to use the strength of interdisciplinary approaches and combine expertise from the domains of Education, Psychology, and Computational Sciences to apply machine learning methods to estimate VA scores. While machine learning seems to be promising for practices within the classroom (see, e.g., Kaw et al., 2019; Moltudal et al., 2020), the focus of the present study is on their potential use for the estimation of school VA scores.

The collaboration of educational, psychological, and data scientists offers an alternative approach to the classical models: machine learning methods. Machine learning (ML) has yielded spectacular results in numerous fields, such as automated face identification (Taigman et al., 2014) or beating human players at the game Go (Silver et al., 2017). The fields of statistics and machine learning highly overlap in terms of tools and methods, primarily differing on focus of problems and applications (for a discussion, see, e.g., Harrell, 2018). With larger data sets, the social sciences are now drawing more on machine learning approaches (e.g., as in computational social sciences, Lazer et al., 2009), which may provide higher prediction accuracies. Often these approaches draw from more general non-linear function fitting approaches (e.g., kernel methods) or combine several weaker models to improve performance (e.g., boosting or random forests). However, these approaches often require large datasets and involve models that can be more difficult to interpret (i.e., black boxes).

Recent research in learning analytics and educational data mining has applied, with great success, machine learning to a wide set of educational problems (Romero and Ventura, 2020), such as predicting student performance in higher education (for recent reviews of the literature, see, e.g., Hellas et al., 2018; Papadogiannis et al., 2020). While a wide variety of models have been tested (e.g., decision trees, neural networks, support vector machines, and linear regression; Rastrollo-Guerrero et al., 2020), there is no consensus yet on which model is the "best" (Papadogiannis et al., 2020). This state of affairs is partially due to studies using different covariates or model performance metrics (Hellas et al., 2018). Such differences between studies make direct comparison of their results difficult, and may lead to inconsistencies and confusion (Papadogiannis et al., 2020). The lack of standardized benchmarks means that while it is clear that these machine learning methods may overall perform well in predicting student performance, determining which specific model to privilege requires direct statistical comparison on a given dataset. In addition, one should note that most predictive models are influenced by their modelers (see, e.g., Kuhn and Johnson, 2013), which begs the question of how far the VA scores from these predictive models differ from those obtained via more classical approaches.

The specific fruitfulness of machine learning methods for the application of school VA models is supported by recent research reporting higher accuracy and more reliable estimates of school VA scores when comparing "random forests" regression to a classical linear regression (Schiltz et al., 2018). To the best of our knowledge, this is the only study that has compared machine learning methods to a classical approach for the estimation of school VA scores. In Schiltz et al. (2018), simulated data and population data from Italy were used to investigate the application of random forests for the estimation of school VA scores. They reported that random forest models predicted outcomes more accurately than linear regression models. Not only did VA scores differ numerically depending on the model type used, the ranking of VA scores across schools differed as well (in particular among schools that ranked very high or very low) which implies that the choice of model type may have substantial practical consequences. The authors recommended the use of random forests over linear regressions when estimating school VA scores, especially when using VA scores for high-stakes decisions, as higher accuracy may prevail over transparency. Random forests methods can capture complex non-linear relationships between dependent and independent variables and are far more flexible than linear regression models; if the data deviates from linearity and the dataset is large enough, techniques like "random forests" can grasp patterns that classical linear models cannot.

This means that these random forests have an advantage over the classical linear regression model, as they do not assume a linear relationship. However, random forests only represent one type of machine learning approach, and so it is unclear whether the improved performance is due to either the non-linearities or the ensemble nature of the method. Additionally, linear regressions are only one of the two typically used model types in the estimation of school VA scores (Kurtz, 2018; Levy et al., 2019); the other one, multilevel models, was not considered in their study. Finally, it is unclear how their result will generalize across other datasets, in particular given differences in covariates and populations.

Hence, we expand on this work by considering a broader class of predictive models, which will be described in detail in the method section. In brief, we compare: linear, multilevel, and polynomial regression, random forest, neural networks, support vector machines, and boosted approaches (see also **Table 1** for an overview).

The Present Study

As mentioned above, there is currently no consensus on how to best estimate school VA scores (e.g., Levy et al., 2019, but see also Schiltz et al., 2018). One previous study has sought to analyze systematically different covariate combinations in school VA models (Levy et al., 2020), with one limitation of this study being the use of only one model type (i.e., multilevel model). The present study thus aims to expand the study by Levy et al. (2020) by examining different model types for the estimation of school VA scores by the interdisciplinary approach of adding methods typically used in computational sciences to the typically psychometric approaches.

We aim to extend the study from Levy et al. (2020) by examining different model types for the estimation of school VA scores, and the study from Schiltz et al. (2018) by using a different data set with population data, by adding multilevel models, by adding non-linear "classical" models, and by adding different types of machine learning methods (e.g., with and without the assumption of linearity) to the comparison.

A common and appropriate way of comparing predictive models is by using a class of methods called cross-validation (Hastie et al., 2009). Cross-validation allows us to estimate a model's out-of-sample performance, that is performance on predicting data that the model was not fit on. It achieves this by randomly splitting the data into "train" and "test" subsets. The model is then fit on the training set, and performance (e.g., R^2) is evaluated on the testing set. This process can then be repeated, either by randomly subsampling or by an initial partitioning, allowing the results to be averaged. For all models used in our analysis, VA scores were computed based on average residuals per school, in the same way as the linear model.

For our analysis, we used the same selection of covariates across all statistical models. This allows for a fair comparison

TABLE 1 | Description of the applied models.

Model	Relationship	Specifics	Package and function	Hyper parameters
Classical approaches				
Linear regression	Linear		*stats* (R Core Team, 2019)*: lm*	/
Multilevel model	Linear	Hierarchical structure taken into account	*lme4* (Bates et al., 2015)*: lmer*	/
Polynomial regression	Non-linear	Third degree to all continuous variables	*stats* (R Core Team, 2019)*: lm*	/
Machine learning approaches				
Random forest	Non-linear	Extension of decision trees	*ranger* (Wright et al., 2020)*: ranger*	– Randomly selected predictors: 2, 5, 8 – Splitting rule of variance, extra trees, maxstat – Minimum node size: 5, 8, 10.
Neural networks	Non-linear	Sequential logistic regression	*nnet* (Venables and Ripley, 2002)*: nnet*	– Number of hidden units: 1, 3, 5, 10 – Weight decay: 0, 0.001, 0.1, 0.5, 0.9
Linear support vector machines	Linear	Extension of regression approaches; combination of finding the minimal margin hyperplane and the kernel method	*kernlab* (Karatzoglou et al., 2004)*: svmLinear*	– Cost of constraint violation: 0.001, 0.01, 0.1, 0.5, 0.9, 1
Polynomial support vector machines	Non-linear		*kernlab* (Karatzoglou et al., 2004)*: svmPoly*	– Polynomial degree: 1, 2, 3 – Distance measure for kernel: 0.001, 0.010, 0.100 – cost of constraint violation: 0.001, 0.01, 0.1, 1
Radial support vector machines	Non-linear		*kernlab* (Karatzoglou et al., 2004)*: svmRadial*	– Distance measure (kernel): 0.01, 0.05, 0.1, 0.5, 1 – Cost of constraint violation: 0.001, 0.01, 0.1, 0.5, 0.9, 1
Boosting	Linear	Ensemble method; models sequentially trained based on performance of past models	*xgboost* (Chen et al., 2019)*: xgbLinear*	– Number of boosting iterations: 25, 50, 100 – L1 and L2 regularization: 0, 0.01, 0.01, 0.1, 1 – Learning rate: 0.05, 0.1, 0.3, 0.6

between models. The choice of covariates was made based on the basis of models of school learning (e.g., Haertel et al., 1983; Wang et al., 1993), findings on predictors of students' achievement and recent findings from systematic analyses on covariate selection in school VA models (Levy et al., 2020). More specifically, these results were obtained by using multilevel models and indicated that the inclusion of prior math achievement, prior language achievement, and covariates related to students' sociodemographic and sociocultural backgrounds (i.e., socioeconomic status of the parents, languages spoken at home, migration status, and sex) into school VA models can make a difference in controlling for between-school differences in student intake and in the resulting school VA scores. Hence, these covariates were included into all statistical models in the present study. One limitation of the study by Levy et al. (2020) was that only one model type was used (i.e., multilevel model); here we contrast several model types.

We addressed two main research questions:

(1) How is the predictive power of school VA models (in predicting student academic scores) affected by different types of classic and more modern models?

(2) How sensitive is schools' VA ranking to the selection of model types for the VA model?

MATERIALS AND METHODS

Participants

This study is a secondary analysis and uses longitudinal large-scale data obtained from the Luxembourg School Monitoring Programme *ÉpStan* (LUCET, 2019). *ÉpStan* assesses students' academic competencies (in math and languages), their subjective achievement motivation as well as information on their sociodemographic and sociocultural background at the beginning of the grade levels 1, 3, 5, 7, and 9. Every year, the entire student population in each of the concerned grade levels participates in the *ÉpStan*. In the present paper, we used longitudinal data from the student cohort that participated in *ÉpStan* in grade 1 in 2014.

For our analyses, we included only those N = 3,026 students attending 153 primary schools with complete cases on all variables (see **Table 2** for details on sample composition and excluded students). Excluded students (N = 1,977) were either

TABLE 2 | Details on the sample composition and excluded students.

	Included students[a]	**Excluded (at least one missing value)**	**Excluded (no participation in grade 3)**	**Excluded (students switched school)**
Number of students	3,026	577	1,068	332
Mean prior math ach. in grade 1	523 (*SD* = 91)	503 (*SD* = 90)	437 (*SD* = 103)	499 (*SD* = 82)
Mean prior language ach. in grade 1	523 (*SD* = 92)	495 (*SD* = 97)	441 (*SD* = 103)	489 (*SD* = 85)
Percentage of female students	50	48	47	49
First language of instruction not spoken at home	49	58	65	60
Mean HISEI score	50.1 (*SD* = 15.4)	47.0 (*SD* = 15.0)	42.6 (*SD* = 15.5)	44.9 (*SD* = 15.2)
Mean math ach. in grade 3	519 (*SD* = 103)	495 (*SD* = 108)	–	479(*SD* = 93)
Mean language ach. in grade 3	518 (*SD* = 101)	482 (*SD* = 101)	–	474 (*SD* = 103)

ach., achievement. [a]Based on the criteria described above.

absent on the day of testing in third grade (N = 1,068; e.g., due to illness or grade repetition), or they changed schools between grades 1 and 3 (N = 332), or they had at least one missing value in the relevant covariates (N = *577*). Excluded students had lower achievement values than included students, indicating among others that non-participation in grade 3 could most likely be due to repeating a grade between first and third grade. Treatment of missing data is a highly debated subject in many areas, also in VA research (e.g., Dearden et al., 2011). Here, we decided to analyze only complete cases, as the model comparisons would otherwise depend on assumptions made at the imputation process which could favor particular model types and hence prevent a clear interpretation of their results.

The *ÉpStan* has a proper legal basis and the national committee for data protection gave its approval. Appropriate ethical standards were respected (American Psychological Association, 2017). All participating children and their parents or legal guardians were duly informed before the data collection, and had the possibility to opt-out. To ensure students' privacy and in accordance with the European General Data Protection Regulation, collected data were pseudonymized with a so called "Trusted Third Party" (for more information see LUCET, 2019). For the present analysis, an anonymized dataset was used.

Measures

Academic Achievement

VA modeling requires a choice of academic achievement as outcome measure, and often uses previous academic achievement as a covariate. Since for our data we have two equally appropriate choices—namely math and language achievement—we computed VA scores for both and report all results (for a recent meta-analysis on the mutual relationship between language and mathematics, see Peng et al., 2020). These two achievement measures from grade 3 were used as outcome variables, while the same scores from grade 1 were used as covariates (i.e., as a measure of prior achievement). At the very beginning of grades 1 and 3, all achievement measures were assessed with standardized achievement tests, which were developed on the basis of the national curriculum standards (defined by the Ministry of National Education, Children and Youth, 2011) by interdisciplinary expert groups, thus assuring content validity (Fischbach et al., 2014). The tests were administered in the classroom setting, given in a paper-and-pencil format, and mostly based on closed-format items. To scale the items, a unidimensional Rasch model was used (Fischbach et al., 2014; see Wu et al., 2007; Nagy and Neumann, 2010). Weighted likelihood estimates (WLE; Warm, 1989) were used as measures of students' achievement (Fischbach et al., 2014). The reliability scores of all achievement scales were calculated using the function *WLErel* from the *TAM* package version 3.3.10 (Robitzsch et al., 2019), which estimates reliability scores based on WLE values and their standard errors.

Math achievement

The math tests in grade 3 were constructed in German because the language of instruction in grades 1 and 2 is German. Math items assessed children's competencies in three areas: "numbers and operations," "space and form," and "quantities and measures." The reliability of the math test scores in grade 3 was 0.90. *Math achievement in grade 1* (i.e., prior math achievement) was assessed in Luxembourgish (which is, although politically and culturally a language on its own, linguistically speaking a variety of German, see Dalby, 1999) because the language of instruction in preschool is Luxembourgish. Mathematics items assessed children's competencies in the domains "numbers and operations," "space and shape," and "size and measurement"[1]. The reliability of the math test scores in grade 1 was 0.75.

Language achievement

Language achievement in grade 3 was operationalized by the children's listening and reading comprehension in the German language. Listening comprehension was based on the subskills "identifying and applying information presented in a text" and "construing information and activating listening strategies." Reading comprehension was assessed with the subskills "identifying and applying information presented in a text" and "construing information and activating reading strategies/techniques"[1]. The reliability of the listening comprehension and the reading comprehension test scores in grade 3 were 0.81 and 0.88, respectively. The correlation between those two achievement scores was 0.69. We computed

[1]https://epstan.lu/en/assessed-competences-31/

a mean score across listening and reading comprehension in the German language to represent students' language achievement in grade 3 in order to have only one dependent variable. *Language achievement in grade 1* (i.e., prior language achievement) consists of "early literacy comprehension" and "listening comprehension" in Luxembourgish in grade 1 because the language of instruction in preschool is Luxembourgish. Listening comprehension was assessed with the two subskills "identifying and applying information presented in a text" and "construing information and activating listening strategies" with different kinds of texts, which were played from an audio recording. Early literacy comprehension was assessed with the subskills "phonological awareness," "visual discrimination," and "understanding of the alphabetic principle"[2]. The reliability of the listening comprehension and the reading comprehension test scores in grade 1 were both 0.70. Contrary to the language achievement measures in grade three, both the listening and the reading score were included into the models instead of averaging them in order to keep as much information as possible, as these two scores only correlated with each other at 0.50.

A note on psychometric quality of achievement measures

The present study is a secondary analysis and relies on archive data for which only limited information on psychometric quality of achievement measures is available (see Appelbaum et al., 2018). Of note the underlying data are already used in real-life and drive political decisions and hence psychometric data quality has been optimized in that regard. As noted above, the present domain-specific tests were developed by expert panels (i.e., teachers, content-specialists on teaching and learning, psychometricians) to ensure content validity of all test items. All test items have also undergone intensive pilot-testing and psychometric quality checks concerning their empirical fit to the Rasch-Model that was used to derive WLE estimates representing students' domain-specific achievements in grades 1 and 3. Further, all test items were examined whether they exhibit differential item functioning across student cohorts attending the same grade level to allow for commensurable measures across time. These psychometric quality measures helped to ensure structural validity of the test items within and across student cohorts. Additional analyses on their convergent and discriminant validity showed that domain-specific achievement test scores in both grade 1 and grade 3 followed the theoretically predicted pattern to academic self-concepts in matching and non-matching domains (Niepel et al., 2017; van der Westhuizen et al., 2019). Finally, the WLE-scores representing students' domain-specific achievement demonstrated score reliability (with score reliability ranging between 0.70 and 0.90) that suffices research purposes (Schmitt, 1996).

Sociodemographic and Sociocultural Background Variables

To obtain information on children's sociodemographic and sociocultural background, a parents' questionnaire was administered in grade 1. Parents were asked to locate their profession within a given list of occupational categories (e.g., academia or craft); these categories were based on the ISCO classification (International Standard Classification of Occupations). For each occupational category, the average value of the ISEI scale, which is a validated scale (International Socio-Economic Index of occupational status, see Ganzeboom, 2010), was computed to obtain a proxy for the socioeconomic status of the parents (SES). In our grade 1 dataset, ISEI values have a mean of 50.1 and a standard deviation of 15.4. In the first PISA tests in 2000, the average ISEI for all OECD countries was 48.8 (OECD and UNESCO Institute for Statistics, 2003). Parents were also asked where they and their child were born to indicate their immigration status, resulting in the immigration status categories "native," "first generation," and "second generation." In the present analyses, immigration status was coded as dummy variables with "native" being the reference category. In addition to the questionnaire filled out by the parents, grade 1 students also filled out a questionnaire on their own, where they were asked to indicate language(s) spoken with their father and their mother, respectively. As the first language of instruction is Luxembourgish, not speaking any Luxembourgish at home represents a challenge for the newly enrolled students. We thus created a dummy variable to differentiate between those students who do not speak any Luxembourgish at home and those who speak Luxembourgish with at least one parent (reference category). Students' sex was retrieved from the official database of the Ministry of National Education, Children and Youth. **Table 2** includes among others an overview of sociodemographic and sociocultural variables of all 3,026 students from 153 schools.

Analysis

All analyses were conducted using R version 3.6.1 (R Core Team, 2019); the scripts can be found online at https://osf.io/rgt8x/?view_only=752453b81cd243e0b4ebfe33e1a74c33. In order to run models as similarly as possible, the *caret* package version 6.0.85 (Kuhn, 2019) was used as a wrapper of most functions. For all models except for the multilevel model, we followed the steps in Eqs 1 and 2 for prediction and VA estimation. Unless otherwise stated, the function call of the model was defined as follows:

Achievement_in_grade_3 ∼ Prior_Math_Achievement + Prior_Reading_Achievement + Prior_Language_Achievement + SES + migration_status + language_spoken_at_home + sex.

Hence, achievement in grade 3 is our outcome y variable with the others as covariates (see the *caret* package for details on function calls). Note that the "+" operator is treated as selecting covariates from the data, where the model type determines how the covariates are combined (e.g., for random forest covariates are selected to form tree branches, hence the "+" is not literal addition).

As is standard in machine learning, the dataset was randomly split into a *training-set* which contains 70% of the data and is used to fit or "train" models and a *test-set* which contains the remaining 30% of the data and which is used to evaluate the fitted model's ability to predict new ("out-of-sample") data (prediction accuracy was estimated via R-squared). To prevent the results from being dependent on a particular split of the data, the above procedure (i.e., split, train, test) was repeated

[2]https://epstan.lu/en/assessed-competences-21/

100 times, thus creating 100 *training*- and *test*-sets and the prediction performance averaged across those repetitions (100 was chosen to balance estimation with computing limitations). The parameter ranges we specified for hyper parameter selection (i.e., grid search) were selected for each parameter in order to span reasonable values. These are generally based on suggested default ranges for each model that come from standard practice, while respecting computational limitations. We report all values tested below. We performed cross-validation for hyper parameter selection, comparing the resampled results across models for the best performing hyper parameter set (for more detail on resampling procedures, see, e.g., Hothorn et al., 2005; Eugster and Leisch, 2011). For models with no hyper parameters (e.g., linear regression) we performed the identical cross-validation resampling procedure for between model comparison. Model performance was then compared on the resampled results.

Model Comparison

Fundamentally, when creating a predictive model of the form $y = f(x)$, both statistics and machine learning practitioners would specify a function space to optimize over (e.g., polynomials) given some loss criterion (e.g., mean-squared error). Where they differ is in these details of function space, criteria, and fitting procedure. **Table 1** depicts an overview of the different model types used, which relationship they assume between dependent and independent variable(s), some specific criteria to each model type, which package and function was used, and which hyper parameters were defined. We use common models for prediction, including ensemble approaches (random forest and boosting) which combine across many weaker models to improve performance, and general function fitting models (neural networks and support vector machines) which transform the inputs into a potentially more useful space for prediction. A more detailed conceptual description and implementation can be found in the Online Supplement A1.

Estimation of VA Scores

For most statistical models used, school VA scores were calculated as the mean difference between the actual and the predicted achievement values from each student in a certain school (i.e., the residuals).

The only exception were multilevel models, where the VA score of a school was quantified in terms of an estimate of the random effect for a particular school at school level (i.e., the residual of a certain school; see Ferrão and Goldstein, 2009). School VA scores were thus estimated using the *ranef* function from the *lme4* package (Bates et al., 2015). Note that this is only the case for the VA score; the resampling results were estimated the same way across all models.

Operationalization of the Research Questions

To address Research Question 1, we evaluated the predictive power of the underlying VA model in terms of the total amount of variance (R^2s) explained. This was estimated with the *resamples* function from the *caret* package (Kuhn, 2019) by the same estimation process for all models used (i.e., based on the comparison between predicted and observed values of student achievement, using the model's *predict* function).

Further, we tackled Research Question 2 on how the VA ranking of schools depends on the model selection by computing correlations of school VA scores with each other as obtained from various school VA models and by analyzing the implications of model selection on benchmark classifications. Specifically, following current benchmarks (e.g., Marzano and Toth, 2013), we classified the best 25% of schools (in terms of VA scores) as "highly effective," the worst 25% as "needs improvement," and the remaining 50% of schools (i.e., between the 25th and 75th percentiles) as "moderately effective." For every school VA model, we computed the percentage of disagreements by calculating the percentage of schools identified at a different benchmark classification as the one resulting from the multilevel model, which is one of the two most commonly used school VA models and which in this analysis serves as a reference (Kurtz, 2018; Levy et al., 2019). More concretely, to get the percentage of disagreements, for every model, the number of schools ranking at a different benchmark than by the multilevel model was divided by the total number of schools and multiplied by 100. Smaller values represent a higher concordance with the benchmark classifications from the multilevel model; higher values indicate a higher rate of disagreements. While all preceding operationalizations include results from all 153 schools, real-life implications of benchmark classifications based on different model types will be illustrated based on five example schools.

RESULTS

Research Question 1: How Is the Predictive Power of School VA Models Affected by Different Types of Classic and More Modern Models?

School VA Models for Math Achievement

Figure 1A shows the mean and the confidence intervals of the amount of explained variance (R^2) for the 100 cross-validations of each statistical model with math achievement as a dependent variable. It can be observed that the values of the different models are close to each other, with the highest predictive power error for the multilevel model (mean R^2 of 0.51) and the lowest for neural networks (mean R^2 of 0.40). For all the other models, the mean R^2 was between 0.44 and 0.47.

School VA Models for Language Achievement

Figure 1B shows the confidence intervals of explained variance (R^2) for the 100 cross-validations of each statistical model with language achievement as a dependent variable. The results are analogously to the school VA models for math. It can be observed that the values of the different models are close to each other, with the highest predictive power for the multilevel model (mean R^2 of 0.54) and the lowest mean R^2 for neural networks (0.44) and linear boosting (0.46). For all the other models, the mean R^2 was between 0.48 and 0.49.

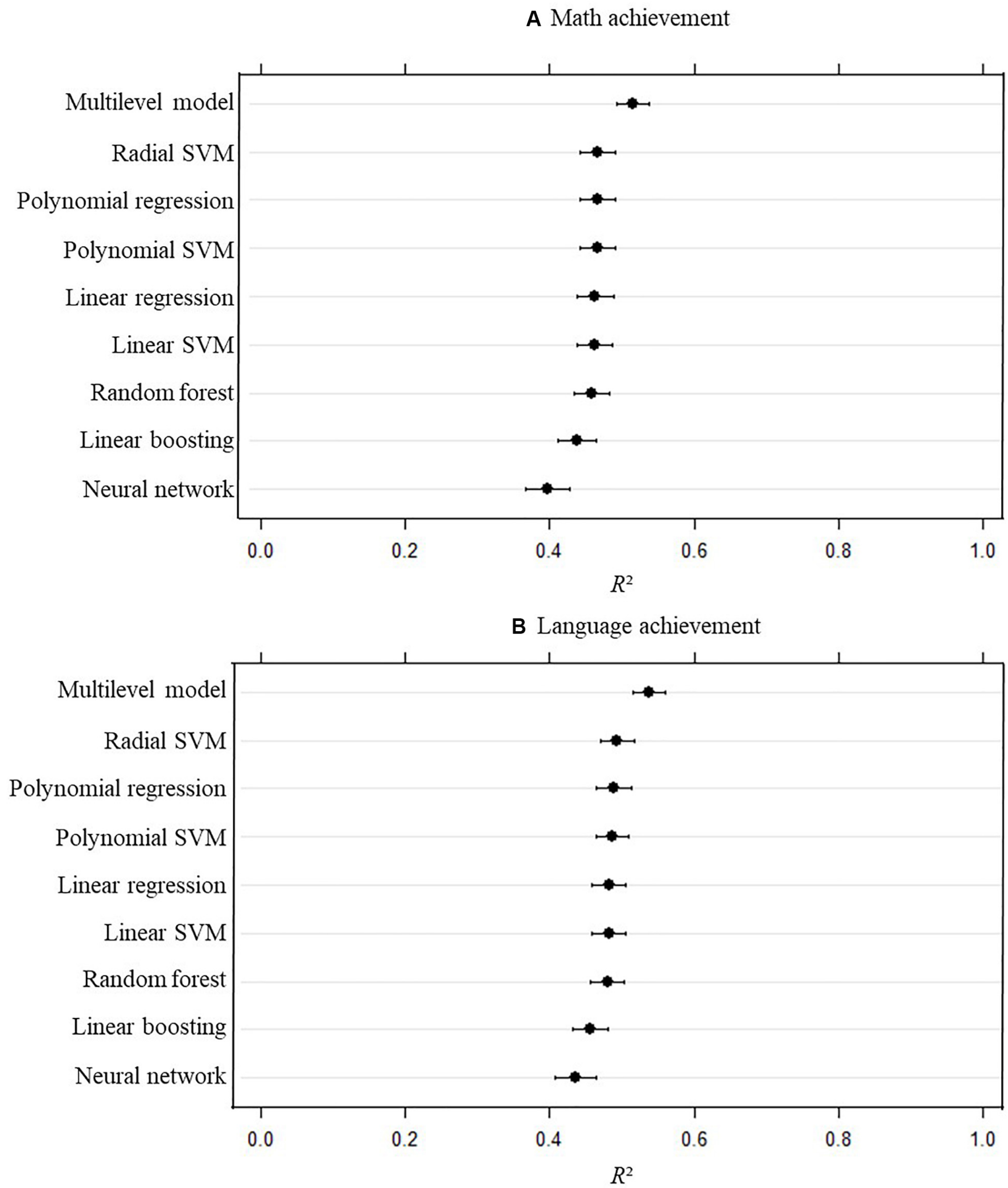

FIGURE 1 | Amount of explained variance (R^2) for each statistical model with math achievement **(A)** and language achievement **(B)** as a dependent variable with confidence intervals computed on the out-of-sample measures obtained via the 100 repetitions. Error bars represent 95% confidence intervals computed on the out-of-sample measures obtained via the 100 repetitions. SVM, support vector machines.

Research Question 2: How Sensitive Is Schools' VA Ranking to the Selection of Statistical Models Types for the VA Model?

Correlations Between School VA Scores by Model Type

School VA models for math achievement

Table 3 shows the correlations between the school VA scores resulting from the VA models for math achievement based on different statistical models. They range between 0.88 and 1.00 (*Mdn* = 0.98). The lowest correlations can be observed for those school VA scores resulting from linear boosting (ranging from 0.88 to 0.94) and neural networks (ranging from 0.94 to 0.97). For all other model types, the resulting school VA scores correlate with each other to at least $r = 0.98$.

School VA models for language achievement

A similar pattern can be observed for school VA models for language achievement (**Table 4**). Correlations between the resulting school VA scores range between 0.89 and 1.00 (*Mdn* = 0.98). The lowest correlations can be observed for those

TABLE 3 | Correlations between school VA scores resulting from different model types with math achievement as a dependent variable.

Model type	Linear regression	Multilevel model	Polynomial regression	Random forests	Neural networks	Linear SVM	Poly-nomial SVM	Radial SVM
Linear regression	–							
Multilevel model	0.98	–						
Polynomial regression	1.00	0.98	–					
Random forests	0.99	0.98	0.99	–				
Neural networks	0.96	0.94	0.96	0.97	–			
Linear SVM	1.00	0.98	1.00	0.99	0.96	–		
Polynomial SVM	1.00	0.98	1.00	0.99	0.96	1.00	–	
Radial SVM	1.00	0.98	1.00	0.99	0.96	1.00	1.00	–
Linear boosting	0.93	0.92	0.93	0.94	0.88	0.93	0.93	0.93

SVM, support vector machines. All correlations are significant with $p < 0.01$.

TABLE 4 | Correlations between school VA scores resulting from different model types with language achievement as a dependent variable.

Model type	Linear regression	Multilevel model	Polynomial regression	Random forests	Neural networks	Linear SVM	Poly-nomial SVM	Radial SVM
Linear regression	–							
Multilevel model	0.97	–						
Polynomial regression	1.00	0.96	–					
Random forests	0.99	0.95	0.99	–				
Neural networks	0.93	0.93	0.92	0.94	–			
Linear SVM	1.00	0.97	1.00	0.99	0.93	–		
Polynomial SVM	1.00	0.96	1.00	0.99	0.93	1.00	–	
Radial SVM	0.99	0.96	1.00	0.99	0.94	0.99	1.00	–
Linear boosting	0.96	0.92	0.96	0.97	0.89	0.96	0.96	0.96

SVM, support vector machines. All correlations are significant with $p < 0.01$.

school VA scores resulting from linear boosting (ranging from 0.89 to 0.97) and neural networks (ranging from 0.92 to 0.94). For all other model types, the resulting school VA scores correlate with each other to at least $r = 0.95$.

Percentage of Disagreement in Comparison to the Multilevel Model

In the following section, we evaluate to what extent the classification of schools into one of the three benchmark classifications "needs improvement," "moderately effective," and "highly effective" depends on the particular model used to compute the VA scores. More specifically, we will use the classification that results from the multilevel model as the reference against which to compare the classifications that results from all other VA scores estimation methods.

School VA models for math achievement

Figure 2 shows the percentage of disagreement as compared to the school VA scores based on the multilevel model. In **Figure 2A**, representing the school VA models with math achievement as a dependent variable, it can be observed that for most statistical models, the percentage of disagreement is under 10%. The only exceptions are school VA scores based on the neural network model (21% of disagreements) and the linear boosting model (17% of disagreements). A detailed overview of percentages of disagreement from school VA models for math achievement compared to those from the multilevel model can be found in **Table 5**.

School VA models for language achievement

Figure 2B shows how many schools' benchmark classifications would be in disagreement based on their language VA scores resulting from the different statistical models. Analogously to the results for the school VA models for math achievement, it can be observed that the percentages of disagreement of most statistical models are similar to each other. More specifically, the percentages of disagreement are around 10% for all models except for the neural network (24%). A detailed overview of percentages of disagreement from school VA models for language achievement compared to those from the multilevel model can be found in **Table 6**.

Real-Life Implications on the Example of Five Schools

Figure 3 illustrates the real-life implications that the use of different statistical models for the estimation of VA scores may have for five schools that were chosen as examples (see **Table 7** for descriptive data on these schools; these are the same schools that were presented in Levy et al., 2020). It shows the range of the VA percentiles resulting from the different statistical models for these schools and illustrates that, despite high correlations across schools, there is variation within individual schools. More specifically, for most schools, when comparing schools'

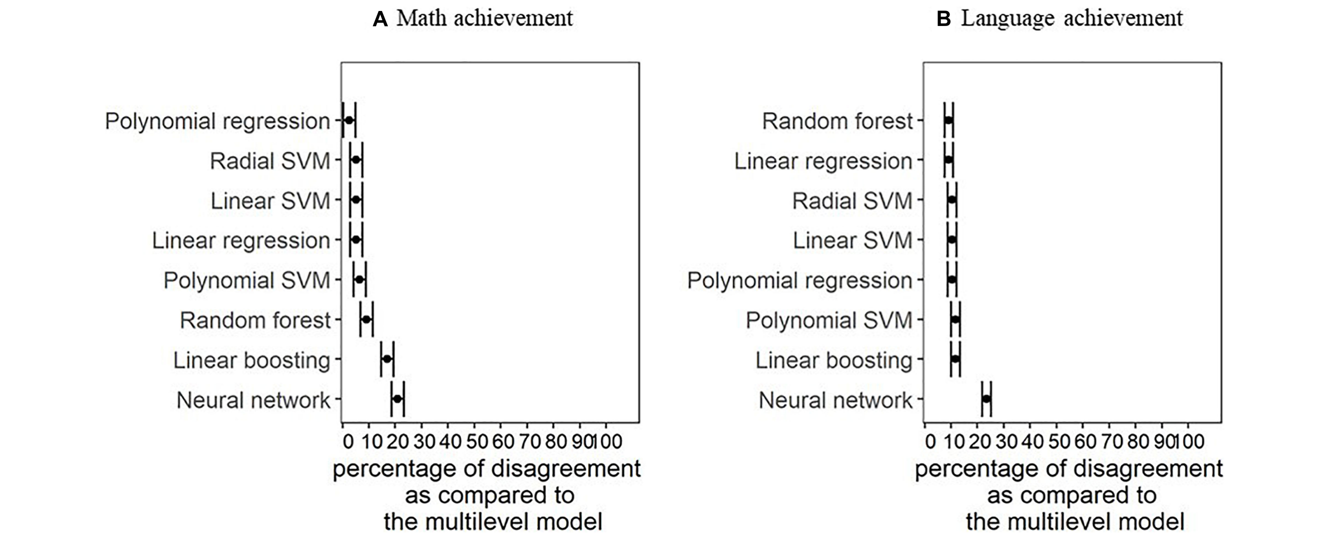

FIGURE 2 | Percentage of disagreement as compared to the benchmark reached based on VA scores from the multilevel model with math achievement **(A)** and language achievement **(B)** as a dependent variable. SVM, support vector machines.

TABLE 5 | Percentage of disagreement as compared to the benchmark reached based on VA scores from the multilevel model with math achievement as a dependent variable.

	Classified by multilevel model as:			
Model	**Needs improvement**[a]	**Moderately effective**[b]	**Highly effective**[c]	**Total percentage of disagreement**
Linear regression	7.89	5.19	2.63	5.23
Polynomial regression	2.63	2.60	2.63	2.61
Random forest	7.89	9.09	10.53	9.15
Neural network	23.68	20.78	18.42	20.92
Linear boosting	10.53	16.88	23.68	16.99
Linear SVM[d]	7.89	5.19	2.63	5.23
Polynomial SVM[d]	8.89	6.49	5.26	6.54
Radial SVM[d]	2.63	5.19	7.89	5.23

[a]Below the 25th percentile. [b]Between the 25th and the 75th percentiles. [c]Above the 75th percentile. [d]SVM, support vector machines.

TABLE 6 | Percentage of disagreement as compared to the benchmark reached based on VA scores from the multilevel model with language achievement as a dependent variable.

	Classified by multilevel model as:			
Model	**Needs improvement**[a]	**Moderately effective**[b]	**Highly effective**[c]	**Total percentage of disagreement**
Linear regression	7.89	9.09	10.53	9.15
Polynomial regression	7.89	10.39	13.16	10.46
Random forest	5.26	9.09	13.16	9.15
Neural network	21.05	23.38	26.32	23.53
Linear boosting	13.16	11.69	10.53	11.76
Linear SVM[d]	7.89	10.39	13.16	10.46
Polynomial SVM[d]	7.89	11.69	15.79	11.76
Radial SVM[d]	5.26	10.39	15.79	10.46

[a]Below the 25th percentile. [b]Between the 25th and the 75th percentiles. [c]Above the 75th percentile. [d]SVM, support vector machines.

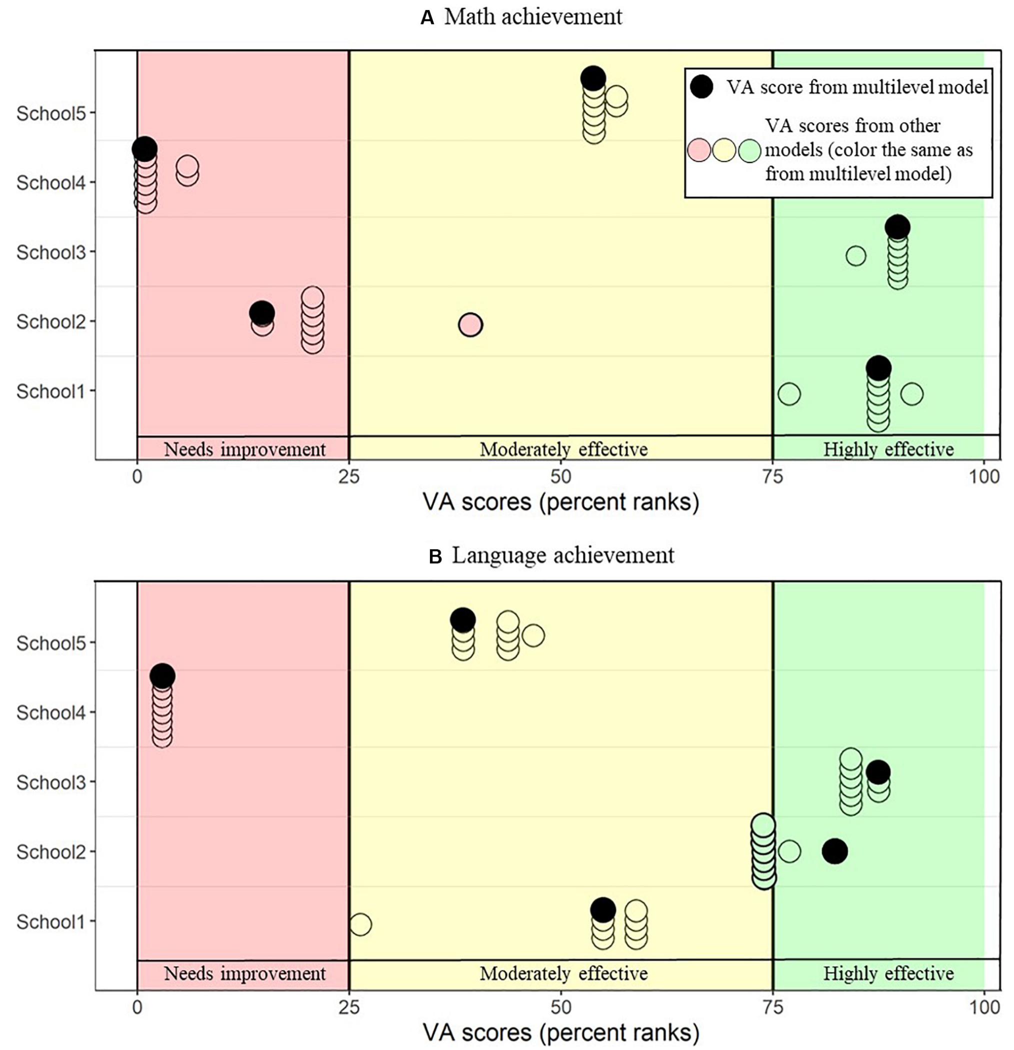

FIGURE 3 | Range of percentiles resulting from math **(A)** and language **(B)** VA scores for five example schools. Every dot represents the school VA percentile as obtained from a certain VA model. At the 25th and 75th percentiles, there are cut-off lines to define the border between schools classified as "needs improvement" (colored in red), "moderately effective" (colored in yellow), and "highly effective" (colored in green). VA percentiles resulting from the multilevel model are marked in black. All the other models are represented by dots in the same color as the area the VA percentile from the multilevel is situated, allowing to see how often schools' classifications would be in disagreement.

VA percentiles within the same dependent variable, schools would be categorized within the same benchmark (i.e., constantly within "needs improvement," "moderately effective," and "highly effective"), regardless of the statistical model used (the exact values can be found in **Table 8**). However, for school 2, depending on the type of model used, the school is classified differently. Interestingly, for every school except for school 1 the most extreme values of VA percentiles (i.e., highest or lowest) are reached with the multilevel as the underlying school VA model.

DISCUSSION

School VA models are statistical models designed to estimate school effectiveness (i.e., school VA scores) based on the evolution of students' achievement. These VA scores are often used for accountability and high-stakes decisions to allocate financial or personal resources to schools. However, despite their practical relevance, there is currently no consensus on how to best estimate VA scores. The two most commonly used statistical models are linear regression and multilevel models (Kurtz, 2018; Levy et al., 2019) and some researchers have applied non-linear models for the estimation of school VA scores (Fitz-Gibbon, 1997; Lopez-Martin et al., 2014). An alternative approach to these classical models involves machine learning methods, which social sciences are drawing on more as larger and more complex data sets become increasingly available, as new types of data require new data analytical approaches. Such techniques have already evolved in fields with a long tradition in crunching big data (e.g., gene technology). One

TABLE 7 | Descriptive data from the five example schools shown in **Figure 3**.

	School 1	School 2	School 3	School 4	School 5
Number of students	18	52	33	49	26
Mean prior math achievement in grade 1	459	575	509	592	534
Mean prior language achievement in grade 1	468	567	511	587	506
Percentage of female students	33%	54%	61%	53%	58%
Percentage of "First language of instruction not spoken at home"	66	33	55	80	46
Mean HISEI score	41.0	56.2	54.6	56.8	48.2
Mean math achievement in grade 3	504	540	543	498	517
Mean language achievement in grade 3	476	579	527	470	510

TABLE 8 | Percentiles resulting from different VA models for the five example schools.

	School 1		School 2		School 3		School 4		School 5	
	Math	Language	Math	Language	Math	Language	Math	Language	Math	Language
Linear regression	88	55	21	77	89	86	1	2	54	39
Multilevel model	88	53	14	82	91	88	0	1	53	39
Polynomial regression	88	57	20	75	90	85	1	2	57	43
Random forest	87	59	19	72	89	83	2	2	55	42
Neural network	77	26	39	75	85	84	5	5	55	45
Linear boosting	91	57	16	76	89	85	7	3	55	47
Linear SVM	88	57	20	74	89	87	1	2	53	38
Polynomial SVM	88	61	22	74	90	88	1	2	55	39
Radial SVM	88	56	22	74	90	86	1	2	57	44

SVM, support vector machines.

reason to investigate the use of machine learning methods for the estimation of VA scores is because they can take complex interactions between students' intake characteristics as well as complex functional forms how these characteristics are related to outcome variables into account. This is different to the classic approach of how VA scores are computed where typically only a linear function is specified that relates the pretest measure (and perhaps further covariates) to the outcome measure. Our goal in the present paper was to contribute to the discussion on how best to compute VA scores while at the same time evaluating the potential of modern machine learning methods in this discipline. More specifically, the present study aimed to address two main research questions:

(1) How is the predictive power of school VA models affected by different types of classic and more modern models?
(2) How sensitive is schools' VA ranking to the selection of model types for the VA model?

In the following, we discuss the implications from the results of the math and language school VA models together because the focus of the present paper is on the model choice rather than on the domain of the dependent variable. The findings are consistent across both domains, which suggests that they are robust and that the results from of the math and language school VA models can be grouped together in the discussion.

How Is the Predictive Power of School VA Models Affected by Different Types of Classic and More Modern Models?

The predictive power of school VA models was very similar for most model types. The only exceptions were the values from the multilevel model and from the neural network, which were significantly different from the other model types. More concretely, the multilevel model performed better than all the other models and the neural network worse. These findings were consistent across dependent variables (i.e., math achievement and language achievement).

The fact that most of the machine learning models were not significantly different from the linear, and the multilevel model outperforming, might seem surprising on the first sight. However, this is likely due to a few reasons. One is a simple model complexity tradeoff, and that out-of-sample performance penalizes overfitting the training data. The performance of these data driven approaches is always subject to basic statistical issues of model complexity and bias-variance tradeoffs (e.g., Hastie et al., 2009). Models with low complexity (e.g., low number of parameters) can perform better for out of sample (or "test") datasets, as complex models are prone to overfit (i.e., adjust the model to noise). While educational domains are likely highly structured, it is not *a priori* obvious if atheoretical non-linear or complex models will be able to capture this structure. Without formal theory that makes strong predictions in this domain, we must rely on statistical comparisons to select the machine learning models for estimating VA scores. Either the non-linear structure is not appropriate for the data (hence no benefit beyond a linear equation) or the noise in data is high enough to prefer the simpler models.

Additionally, the multilevel model's performance could be explained by the fact it took into account the nested structure of the data, which the other models did not. This indicates that fitting VA models only across schools is not enough, as there seems to be important information within schools that can add to explaining variance. Of course, it could be argued that the multilevel models received more information than the other models, as schoolID was added to the equation. However, the standard logic of VA estimation means such information cannot be appropriately included, unless one can estimate school-level error independent of non-school factors (which the multilevel model allows). They are the only one of the model types chosen for the present analysis that is able to take into account the nested structure of the data appropriately. Just adding school as a covariate to the other models would thus not contribute to a solution, as this would result in breaking the logic in estimation of VA scores, as they are calculated for schools without school information.

As opposed to the findings from Schiltz et al. (2018), the predictive power of the linear regression was not significantly different from the random forest. This could be due to differences in the datasets: Schiltz et al. (2018) used Italian population data, while we used Luxembourgish population data. Additionally, Schiltz et al. (2018) used secondary school students from Italian population data, while we used primary school students from Luxembourgish population data. These differences suggest that model performance can depend on the dataset used, and that caution should be given in generalizing beyond one dataset. Additionally, different covariates were included, such as our inclusion of language(s) spoken at home, a significant variable for the Luxembourgish population (e.g., Ugen et al., 2013; Martini and Ugen, 2018). Given our previous results (Levy et al., 2020), covariate choice is highly important in model performance, and therefore a critical concern in comparing different datasets.

However, the present results of the superior performance of multilevel models offer a suggestion on default model choice. The exploration of school VA scores on primary school aged children is especially relevant in heterogeneous populations such as Luxembourg, as socioeconomic disparities appear already within the two first grades of primary school (Hoffmann et al., 2018). For this specific context, multilevel models outperform classical linear and polynomial regressions, as well as different machine learning models. While in many domains linear regression is widely accepted as the default model, changing this default to multilevel modeling works well for hierarchically structured data (as discussed by McElreath, 2017). This seems to be the case for school VA models, as well. While this is sensible since the data clearly have a hierarchical structure (e.g., students nested within schools); the present results statistically demonstrate the multilevel model's performance.

How Sensitive Is Schools' VA Ranking to the Selection of Model Types for the VA Model?

School VA scores resulting from the different model types correlated highly with each other (ranging from 0.88 to 1.00 for school VA models in math and from 0.89 to 1.00 for school VA models in language). At first glance, this might suggest that the resulting school VA scores are similar to each other across schools, which could even lead to the—premature—conclusion that the least complex model, in terms of parsimony and transparency (i.e., the linear regression because of its intuitive interpretability, see, e.g., Molnar, 2020) should be chosen (see e.g., Cohen, 1990; Wilkinson and Task Force on Statistical Inference, 1999). However, high correlations between different school VA scores will not necessarily prevent disagreements of classifications from individual schools (e.g., Timmermans et al., 2011; Ehlert et al., 2014; Levy et al., 2020). This is why, in a second step, the school VA scores were transferred to percentiles and then benchmarks were used to classify schools (i.e., "needs improvement," "moderately effective," and "highly effective").

We compared the resulting benchmarks from all models to those obtained by the multilevel model by calculating the percentage of disagreement. The percentage of disagreement was mostly around 10%. The only exceptions were neural network for school VA scores in math and language and linear boosting for school VA scores in language. However, these two model types were also the ones with the lowest predictive power and it is thus not surprising that their resulting benchmark classifications deviate the most.

As for all the other models, the percentage of disagreement seems low. However, 10% of disagreement means that for most models, at least 15 out of these 153 schools would be classified differently if another model than multilevel models is used (assuming multilevel models provide the reference classification). Given that these benchmark classifications can have high-stakes consequences, the present results underline the relevance of model choice, as individual schools' VA rankings are sensitive to the selection of model types. To further illustrate the real-life implications the model selection can have on individual schools, we will discuss the example of five schools.

Real-Life Implications

Despite very high correlations between school VA scores across models, we can still see differences in benchmark classifications for some of our example schools depending on the model used. This raises the question "how high should a correlation be for it not to matter?" This question cannot be answered in a general way, as it depends on the very practical and political issue of how these VA scores are used in practice. Rather, it should be kept in mind by any researcher, practitioner, or politician when applying or interpreting results from school VA models. Most importantly, it should be kept in mind, especially when taking high-stakes or accountability decisions based on VA models, that any single value used to evaluate schools' effectiveness represents only one possible truth; a single point estimate. One alternative would be to include confidence intervals on VA score rankings and classifications, as well as a combination of interdisciplinary methods as was done in the present study. This would allow a range of possible VA scores for every school VA scores, incorporating uncertainty underlying the estimate. However, ethical implications of the results from the present study and of the use of machine learning methods for consequential decisions in a discipline they were not specifically designed for (i.e., education) should be discussed.

Ethical Implications

The idea of using machine learning to make better and more objective predictions than with conventional statistical methods sounds promising for the application on school VA models, and ideas on how to use machine learning methods for education have been around for decades (Romero and Ventura, 2007, 2010, 2020). However, even though machine learning techniques have driven progress in numerous other disciplines, such as automated face identification (Taigman et al., 2014) or beating human players at the game Go (Silver et al., 2017), their potential downsides and limits need to be discussed, too (see, e.g., Synced, 2018). For example, one study on the diagnostic analysis of medical images reported that out of 516 studies, only 6% tested their algorithms on datasets in different hospitals (Kim et al., 2019). This can result in false associations, such as the association between images from portable x-ray machines and illness (as described by Couzin-Frankel, 2019). This happened because these portable x-ray machines were only used when the patients were already too ill to get out of their bed and as images from these x-ray machines look different from the ones when a patient is not lying down and was thus a circular conclusion. This shows how important data collection processes are, as biased data will lead to biased results, regardless of the applied model.

Image recognition has the advantage that it is still comparably easy for humans to objectively judge whether the classification done by a certain algorithm is true or not. However, this becomes more challenging for concepts such as school VA scores, where the entire point is that we do not know schools' effectiveness, which is why we are estimating VA models in the first place, in order to approximate a measure of schools' effectiveness. Model performance is always limited by the model's assumptions and the data used to train it (e.g., Hastie et al., 2009; Alpaydin and Bach, 2014). This highlights the importance of transparency and clear communication in how these models are estimated, selected, and used, as is also underlined by recent lawsuits (Paige, 2020; Paige and Amrein-Beardsley, 2020).

Implications for Educational Practice

As discussed, a core concern with decisions based on estimates of VA scores is how to appropriately communicate limitations to stakeholders. For example, even if there were no differences across schools in "real" VA scores, a ranking of schools can still be constructed (based purely on randomness). These concerns, along with others presented above, leads to the suggestion that models used for the estimation of school VA scores should never be used alone for high-stakes decisions. This has been elaborated on a more general level in Barocas et al. (2019), where the authors stress the importance of the complementary use of these models together with observational, qualitative, and/or ethnographic studies. This goes in line with researchers recommending a combination of VA scores and observations for high-stakes decisions (e.g., Bacher-Hicks et al., 2019) or of using school VA scores only for informative purposes rather than accountability (e.g., Leckie and Goldstein, 2019). Additionally, even though multilevel models provided the best predictive power within the present dataset, this finding may not generalize to other contexts. We thus recommend that practitioners do not just implement the model suggested by the field, but instead follow a process for model selection with different model types which combine the expertise from different disciplines, as it has been done in the present study. Future work should develop standardized processes and benchmarks, such as following those from Hothorn et al. (2005) and Eugster and Leisch (2011). Optimally, a transparent process for model selection with different model types, combining expertise from multiple disciplines, should be implemented for the estimation of VA scores.

Limitations and Future Work

Treatment of missing data is a highly discussed subject in many areas, also in VA research (e.g., Dearden et al., 2011). We decided to analyze only complete cases, as the model comparisons would otherwise already depend on assumptions made at the imputation process and could lead to differences in VA scores. For future research, it would be interesting to include those with missing cases, possibly comparing different imputation methods and/or by dummy coding whether an entry is missing or not. However, this should be done after the principle differences between different model types have been investigated, hence the importance of the present study.

School VA scores were computed differently in multilevel models as compared to the other model types (i.e., estimated based on the random effects at school level). On the one hand, this might make the comparison between the resulting VA scores unfair in favor of multilevel models. On the other hand, the amount of explained variance was estimated in the same way for all model types. Additionally, this estimation of VA scores by the random effects at school level is specifically how school VA scores are estimated in most cases, thus representing a realistic representation of practice (Ferrão and Goldstein, 2009; Levy et al., 2019). Furthermore, in most other studies comparing classical and machine learning approaches, machine learning approaches have an advantage due to less strict assumptions. More specifically, it is not possible to get a comparison that is fair in every aspect. However, we tried to keep as many aspects constant across model as possible.

Furthermore, the data was obtained with pen and paper rather than using a computer. The latter would have allowed to compute response time and avoid transcription errors. However, particular steps were taken to maximize objectivity and consistency, for example by double coding a random set of answers (Fischbach et al., 2014). However, other measures such as discriminant of convergent validity do not exist, yet. Future studies should thus investigate these important quality criteria, for example by matching the achievement test results with school grades. Given that the present study is a secondary analysis and relies on archive data, only limited information on the psychometric quality of achievement test scores was available and presented here (see Appelbaum et al., 2018). Of note, the domain-specific achievement tests that we used in the present study were developed to support their use for real-life and political decisions in educational settings in Luxembourg. The achievement tests also demonstrated score reliability that is typically considered to suffice research purposes (see Schmitt, 1996). Nevertheless, future work should consider whether the underlying reliability of the measures has an impact on model selection.

Additionally, as the present data set was already prepared (and thus simplified) by classical psychometric methods (i.e., IRT and WLE), it would be interesting for future works to compare the different models when using the raw data instead, as it could be that the machine learning models can make use of the higher level of complexity in the data.

The Luxembourgish school system consists of learning cycles, which usually take 2 years but can be extended to 3 years. Thus, the number of students who took part in grade 1 in 2014 but not in grade 3 in 2016 was quite high. This can introduce biases into our dataset, since the excluded students had lower achievement, lower socioeconomic status values, and a higher percentage of students who did not speak the first language of instruction at home than those who met the inclusion criteria. However, this problem exists in most educational datasets, as students who switched schools or repeated a grade are generally excluded from VA models in prior research; given the VA estimates are typically used for accountability purposes. This means that our data and results largely reflect the reality of how VA scores are typically estimated. Additionally, biases in the dataset will impact any model trained on that dataset similarly.

As previously discussed, the dataset is important in generalizing claims about model performance. Luxembourg is a particularly diverse and multilingual educational context compared to other school systems. Additionally, most applications of VA models estimate performance based on a 1-year time difference, while for us the difference between time points was 2 years (representing one learning cycle in the Luxembourg school system). Future research should replicate the present study to more homogeneous settings, and longitudinal data sets with 1 year, to determine to what degree our results are specific to the particular setting in Luxembourg.

The present study only used data from a single student cohort to obtain school VA score estimates. However, previous research suggests there is a naturally high variability in VA scores across cohorts (e.g., Sass, 2008; Newton et al., 2010; Minaya and Agasisti, 2019). Future research could thus extend the present study by including school VA scores obtained for several student cohorts to investigate whether there are schools with stable VA scores across cohorts/time within (or across) models and the extent to which the stability across cohorts is related to model selection.

While future work might be in creation of theory-driven models rather than a more explorative use of machine learning approaches, we were most interested in comparing standard approaches (both from machine learning and VA models). Our approach thus provides a relevant first step in extending existing research on the estimation of school VA models by investigating all those different approaches that are typically used to estimate school VA scores in particular or to deal with big amounts of data in general. As the multilevel model outperformed any of the standard machine learning approaches used, future research might expand the present study by considering machine learning models with a hierarchical structure that respect the logic underlying VA estimate.

CONCLUSION

The present study investigated different statistical models for the estimation of school VA scores, finding that multilevel models outperformed classical linear and polynomial regressions, as well as a selective sample of different machine learning models. Even though the estimated VA scores from different model types correlated highly across schools, the percentage of disagreement as compared to benchmark classifications based on the multilevel model was substantial. Additionally, real-life implications for individual schools may be consequential depending on the model type used. Based on the present dataset, multilevel models would be recommended for the estimation of school VA scores because these models provide the most accurate predictions of student's achievement. Also, because we observe that VA scores vary depending on specific model choices, we suggest that school VA scores should not be used as the only measure for accountability or high-stakes decisions and that they always be presented with confidence intervals. Optimally, a transparent process for model selection with different model types, combining expertise from multiple disciplines, should be implemented for the estimation of VA scores.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Sensible educational policy data from the Luxembourg School Monitoring Programme "Épreuves Standardisées" (www.epstan.lu) that has been kindly made available for this specific secondary analysis. Requests to access these datasets should be directed to www.epstan.lu.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not provided by the participants' legal guardians/next of kin because the present study is based on existing data. The existing dataset (ÉpStan) has a proper legal basis and has been approved by the national committee for data protection. Appropriate ethical standards were respected (American Psychological Association, 2017). All participating children and their parents or legal guardians were duly informed before the data collection, and had the possibility to opt-out.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The present research was supported by a PRIDE grant (PRIDE/15/10921377) and the ATTRACT grant (ATTRACT/2016/ID/11242114/DIGILEARN) of the Luxembourg National Research Fund (FNR).

ACKNOWLEDGMENTS

We would like to thank the national school monitoring team from the Luxembourg Centre for Educational Testing for providing access to the Épreuves Standardisées database.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02190/full#supplementary-material

REFERENCES

Alpaydin, E., and Bach, F. (2014). *Introduction to Machine Learning*. Cambridge, MA: MIT Press.

American Psychological Association (2017). *Ethical Principles of Psychologists and Code of Conduct*. Available online at: https://www.apa.org/ethics/code/ethics-code-2017.pdf (accessed May 12, 2020).

Amrein-Beardsley, A., Collins, C., Polasky, S. A., and Sloat, E. F. (2013). Value-added model (VAM) research for educational policy: framing the issue. *Educ. Policy Anal. Arch.* 21, 1–14. doi: 10.14507/epaa.v21n4.2013

Amrein-Beardsley, A., and Holloway, J. (2017). Value-added models for teacher evaluation and accountability: commonsense assumptions. *Educ. Policy* 33, 516–542. doi: 10.1177/0895904817719519

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., and Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: the APA publications and communications board task force report. *Am. Psychol.* 73, 3–25. doi: 10.1037/amp0000191

Bacher-Hicks, A., Chin, M. J., Kane, T. J., and Staiger, D. O. (2019). An experimental evaluation of three teacher quality measures: value-added, classroom observations, and student surveys. *Econ. Educ. Rev.* 73:101919. doi: 10.1016/j.econedurev.2019.101919

Baker, R. S. J. D. (2019). Challenges for the future of educational data mining: the baker learning analytics prizes. *J. Educ. Data Min.* 11, 1–17.

Barocas, S., Hardt, M., and Narayanan, A. (2019). Fairness and machine learning. *arxiv* Available online at: https://fairmlbook.org

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2019). *xgboost: Extreme Gradient Boosting (0.90.0.2) [Computer software]*. Available online at: https://CRAN.R-project.org/package=xgboost (accessed March 18, 2020).

Close, K., Amrein-Beardsley, A., and Collins, C. (2020). Putting teacher evaluation systems on the map: an overview of states' teacher evaluation systems post–every student succeeds act. *Educ. Policy Anal. Arch.* 28:58. doi: 10.14507/epaa.28.5252

Cohen, J. (1990). Things I have learned (so far). *Am. Psychol.* 45, 1304–1312. doi: 10.1037/0003-066x.45.12.1304

Couzin-Frankel, J. (2019). Artificial intelligence could revolutionize medical care. But don't trust it to read your x-ray just yet. *Science* doi: 10.1126/science.aay4197

Dalby, D. (1999). *The Linguasphere Register of the World's Languages and Speech Communities*. Hebron: Linguasphere Press.

Dearden, L., Miranda, A., and Rabe-Hesketh, S. (2011). Measuring school value added with administrative data: the problem of missing variables. *Fiscal Stud.* 32, 263–278. doi: 10.1111/j.1475-5890.2011.00136.x

Ehlert, M., Koedel, C., Parsons, E., and Podgursky, M. J. (2014). The sensitivity of value-added estimates to specification adjustments: evidence from school- and teacher-level models in Missouri. *Stat. Publ. Policy* 1, 19–27. doi: 10.1080/2330443X.2013.856152

Eugster, M. J. A., and Leisch, F. (2011). Exploratory analysis of benchmark experiments an interactive approach. *Comput. Stat.* 26, 699–710. doi: 10.1007/s00180-010-0227-z

Everson, K. C. (2017). Value-added modeling and educational accountability: are we answering the real questions? *Rev. Educ. Res.* 87, 35–70. doi: 10.3102/0034654316637199

Ferrão, M. E., and Goldstein, H. (2009). Adjusting for measurement error in the value added model: evidence from Portugal. *Qual. Quant.* 43, 951–963. doi: 10.1007/s11135-008-9171-1

Fischbach, A., Ugen, S., and Martin, R. (2014). *ÉpStan Technical Report*. Luxembourg: University of Luxembourg.

Fitz-Gibbon, C. T. (1997). *The Value Added National Project: Final report: Feasibility Studies for a National System of Value-Added Indicators*. London: School Curriculum and Assessment Authority.

Friedman, J. H., Hastie, T., and Tibshirani, R. (2001). *The Elements of Statistical Learning*, Vol. 1. New York, NY: Springer.

Ganzeboom, H. B. G. (2010). *International Standard Classification of Occupations (ISCO)*. Netherlands: Springer, 3336–3336.

Haertel, G. D., Walberg, H. J., and Weinstein, T. (1983). Psychological models of educational performance: a theoretical synthesis of constructs. *Rev. Educ. Res.* 53, 75–91. doi: 10.3102/00346543053001075

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: estimation using micro data. *Am. Econ. Rev.* 61, 280–288.

Hanushek, E. A. (2019). Testing, accountability, and the american economy. *ANNALS Am. Acad. Polit. Soc. Sci.* 683, 110–128. doi: 10.1177/0002716219841299

Harrell, F. (2018). *Road Map for Choosing Between Statistical Modeling and Machine Learning*. Available online at: https://fharrell.com/post/stat-ml/ (accessed March 19, 2020).

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 1–2. Berlin: Springer.

Hellas, A., Liao, S. N., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., et al. (2018). "Predicting academic performance: a systematic literature review," in *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education – ITiCSE 2018 Companion*, (New York, NY), 175–199.

Hoffmann, D., Hornung, C., Gamo, S., Esch, P., Keller, U., and Fischbach, A. (2018). *Schulische Kompetenzen von Erstklässlern und ihre Entwicklung nach zwei Jahren (Nationaler Bildungsbericht Luxemburg 2018)*. Luxembourg: University of Luxembourg.

Hothorn, T., Leisch, F., Zeileis, A., and Hornik, K. (2005). The design and analysis of benchmark experiments. *J. Comput. Graph. Stat.* 14, 675–699. doi: 10.1198/106186005X59630

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab—an S4 package for Kernel methods in R. *J. Stat. Softw.* 11, 1–20. doi: 10.18637/jss.v011.i09

Kaw, A., Clark, R., Delgado, E., and Abate, N. (2019). Analyzing the use of adaptive learning in a flipped classroom for preclass learning. *Comp. Appl. Eng. Educ.* 27, 663–678. doi: 10.1002/cae.22106

Kim, D. W., Jang, H. Y., Kim, K. W., Shin, Y., and Park, S. H. (2019). Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J. Radiol.* 20, 405–410. doi: 10.3348/kjr.2019.0025

Kuhn, M. (2019). *Caret: Classification and Regression Training (6.0-84) [Computer software]*. Available online at: https://CRAN.R-project.org/package=caret (accessed September 19, 2019).

Kuhn, M., and Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer.

Kurtz, M. D. (2018). Value-added and student growth percentile models: what drives differences in estimated classroom effects? *Stat. Publ. Policy* 5, 1–8. doi: 10.1080/2330443X.2018.1438938

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., et al. (2009). Computational social science. *Science* 323, 721–723.

Leckie, G., and Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: a study of Progress 8 and school accountability in England. *Br. Educ. Res. J.* 45, 518–537. doi: 10.1002/berj.3511

Levy, J., Brunner, M., Keller, U., and Fischbach, A. (2019). Methodological issues in value-added modeling: an international review from 26 countries. *Educ. Assess. Eval. Account.* 31, 257–287. doi: 10.1007/s11092-019-09303-w

Levy, J., Brunner, M., Keller, U., and Fischbach, A. (2020). How sensitive are the evaluations of a school's effectiveness to the selection of covariates in the applied value-added model? Manuscript submitted for publication.

Lopez-Martin, E., Kuosmanen, T., and Gaviria, J. L. (2014). Linear and nonlinear growth models for value-added assessment: an application to Spanish primary and secondary schools' progress in reading comprehension. *Educ. Assess. Eval. Account.* 26, 361–391. doi: 10.1007/s11092-014-9194-1

LUCET (2019). *Épreuves Standardisées (ÉpStan)*. Available online at: https://epstan.lu (accessed March 07, 2019).

LUCET (n. d.). *EpStan. Assessed Competences. Cycle 3.1*. Available online at: https://epstan.lu/en/assessed-competences-31/ (accessed on 21 August 2019).

Martini, S. F., and Ugen, S. (2018). "Language influence on ninth-graders' mathematics performance in a multilingual educational setting," *Paper Presented at Conference on Multilingualism*, Ghent.

Marzano, R. J., and Toth, M. D. (2013). *Teacher Evaluation that Makes a Difference: A New Model for Teacher Growth and Student Achievement*. Alexandria, VA: ASCD.

McCrann, J. T. (2015). *Achievement is not Linear: What Education Policymakers Could Learn From the New York Mets*. New York, NY: Education Week Teacher.

McElreath, R. (2017). *Multilevel Regression as Default. Elements of Evolutionary Anthropology*. Available online at: https://elevanth.org/blog/2017/08/24/multilevel-regression-as-default/ (accessed April 16, 2020).

Minaya, V., and Agasisti, T. (2019). Evaluating the stability of school performance estimates over time. *Fisc. Stud.* 40, 401–425. doi: 10.1111/1475-5890.12201

Ministry of National Education, Children and Youth (2011). *Elementary School. Cycles 1-4. The Levels of Competence*. Luxembourg: Ministry of National Education, Children and Youth.

Molnar, C. (2020). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Available online at: https://christophm.github.io/interpretable-ml-book/ (accessed May 06, 2020).

Moltudal, S., Høydal, K., and Krumsvik, R. J. (2020). Glimpses into real-life introduction of adaptive learning technology: a mixed methods research approach to personalised pupil learning. *Designs Learn.* 12, 13–28. doi: 10.16993/dfl.138

Nagy, G., and Neumann, M. (2010). "Psychometrische aspekte des tests zu den voruniversitären mathematikleistungen in TOSCA-2002 und TOSCA-2006: Unterrichtsvalidität, Rasch-Homogenität und Messäquivalenz," in *Schulleistungen von Abiturienten. Die neu Geordnete Gymnasiale Oberstufe auf dem Prüfstand*, eds U. Trautwein, M. Neumann, G. Nagy, O. Lüdtke, and K. Maaz (Germany: VS Verlag für Sozialwissenschaften), 281–306. doi: 10.1007/978-3-531-92037-5_11

Newton, X., Darling-Hammond, L., Haertel, E., and Thomas, E. (2010). Value-added modeling of teacher effectiveness: an exploration of stability across models and contexts. *Educ. Policy Anal. Arch.* 18, 1–24. doi: 10.14507/epaa.v18n23.2010

Niepel, C., Greiff, S., Keller, U., and Fischbach, A. (2017). "Dimensional comparisons in primary school. A validation of the generalized I/E model," in *Proceedings of the 17th Conference of the EARLI*, Tampere.

OECD and UNESCO Institute for Statistics (2003). *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris: *OECD*.

Paige, M. A. (2020). Moving forward while looking back: lessons learned from teacher evaluation litigation concerning Value-Added Models (VAMs). *Educ. Policy Anal. Arch.* 28:64. doi: 10.14507/epaa.28.5067

Paige, M. A., and Amrein-Beardsley, A. (2020). "Houston, we have a Lawsuit": a cautionary tale for the implementation of value-added models for high-stakes employment decisions. *Educ. Res.* 49, 350–359. doi: 10.3102/0013189X20923046

Papadogiannis, I., Poulopoulos, V., and Wallace, M. (2020). A critical review of data mining for education: what has been done, what has been learnt and what remains to be seen. *Int. J. Educ. Res. Rev.* 353–372. doi: 10.24331/ijere.755047

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., et al. (2020). Examining the mutual relations between language and mathematics: a meta-analysis. *Psychol. Bull.* 146:595. doi: 10.1037/bul0000231

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Geneva: R Foundation for Statistical Computing.

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., and Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: a review. *Appl. Sci.* 10:1042. doi: 10.3390/app10031042

Robitzsch, A., Kiefer, T., and Wu, M. (2019). *TAM: Test Analysis Modules (3.3-10) [Computer software]*. Available online at: https://CRAN.R-project.org/package=TAM (accessed December 02, 2019).

Romero, C., and Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Exp. Syst. Appl.* 33, 135–146. doi: 10.1016/j.eswa.2006.04.005

Romero, C., and Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C* 40, 601–618. doi: 10.1109/TSMCC.2010.2053532

Romero, C., and Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. *WIREs Data Min. Knowl. Discov.* 10:e1355. doi: 10.1002/widm.1355

Sass, T. R. (2008). *The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy*. Available online at: https://eric.ed.gov/?id=ED508273 (accessed February 08, 2017).

Schiltz, F., Sestito, P., Agasisti, T., and De Witte, K. (2018). The added value of more accurate predictions for school rankings. *Econ. Educ. Rev.* 67, 207–215. doi: 10.1016/j.econedurev.2018.10.011

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychol. Assess.* 8:350. doi: 10.1037/1040-3590.8.4.350

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270

Sloat, E., Amrein-Beardsley, A., and Holloway, J. (2018). Different teacher-level effectiveness estimates, different results: inter-model concordance across six generalized value-added models (VAMs). *Educ. Assess. Eval. Account.* 30, 367–397. doi: 10.1007/s11092-018-9283-7

Synced (2018). *2018 in Review: 10 AI Failures. Medium*. Available online at: https://medium.com/syncedreview/2018-in-review-10-ai-failures-c18faadf5983 (accessed April 10, 2020).

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). "Deepface: closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Boston, MA), 1701–1708.

Timmermans, A. C., Doolaard, S., and de Wolf, I. (2011). Conceptual and empirical differences among various value-added models for accountability. *Schl. Effect. Schl. Improv.* 22, 393–413. doi: 10.1080/09243453.2011.590704

Ugen, S., Martin, R., Böhm, B., Reichert, M., Lorphelin, D., and Fischbach, A. (2013). "Einfluss des Sprachhintergrundes auf Schülerkompetenzen," in *PISA 2012*, (Luxemburg: Nationaler Bericht), 100–113.

van der Westhuizen, L., Arens, K., Keller, U., Greiff, S., Fischbach, A., and Niepel, C. (2019). "Dimensional and social comparison effects on domain-specific academic self-concepts and interests with first-and third-grade students," in *Proceedings of the LuxERA Emerging Researchers' Conference*, Esch-sur-Alzette.

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Berlin: Springer.

Wang, M. C., Haertel, G. D., and Walberg, H. J. (1993). Toward a knowledge base for school learning. *Rev. Educ. Res.* 63, 249–294. doi: 10.2307/1170546

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika* 54, 427–450. doi: 10.1007/BF02294627

Wilkinson, L., and Task Force on Statistical Inference (1999). Statistical methods in psychology journals. *Am. Psychol.* 54, 594–604. doi: 10.1037/0003-066x.54.8.594

Wright, M. N., Wager, S., and Probst, P. (2020). *ranger: A Fast Implementation of Random Forests (0.12.1) [Computer software]*. Available online at: https://CRAN.R-project.org/package=ranger (accessed February 19, 2020).

Wu, L. M., Adams, R. J., Wilson, M. R., and Haldane, S. A. (2007). *ACER ConQuest Version 2: Generalised Item Response Modelling Software [Computer Program]*. Camberwell: Australian Council for Educational Research.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 September 2020
doi: 10.3389/fpsyg.2020.562462

Is a Long Essay Always a Good Essay? The Effect of Text Length on Writing Assessment

Johanna Fleckenstein[1]*, *Jennifer Meyer*[1], *Thorben Jansen*[2], *Stefan Keller*[3] and *Olaf Köller*[1]

[1] Department of Educational Research and Educational Psychology, Leibniz Institute for Science and Mathematics Education, Kiel, Germany, [2] Institute for Psychology of Learning and Instruction, Kiel University, Kiel, Germany, [3] School of Education, Institute of Secondary Education, University of Applied Sciences and Arts Northwestern Switzerland, Brugg, Switzerland

The assessment of text quality is a transdisciplinary issue concerning the research areas of educational assessment, language technology, and classroom instruction. Text length has been found to strongly influence human judgment of text quality. The question of whether text length is a construct-relevant aspect of writing competence or a source of judgment bias has been discussed controversially. This paper used both a correlational and an experimental approach to investigate this question. Secondary analyses were performed on a large-scale dataset with highly trained raters, showing an effect of text length beyond language proficiency. Furthermore, an experimental study found that pre-service teachers tended to undervalue text length when compared to professional ratings. The findings are discussed with respect to the role of training and context in writing assessment.

Keywords: text length, writing assessment, text quality, judgment bias, English as a foreign language, human raters, pre-service teachers

OPEN ACCESS

Edited by:
Matthias Stadler, Ludwig Maximilian University of Munich, Germany

Reviewed by:
Hassan Mohebbi, University of Tehran, Iran
Sven Hilbert, University of Regensburg, Germany

Correspondence:
Johanna Fleckenstein
fleckenstein@leibniz-ipn.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *15 May 2020*
Accepted: *31 August 2020*
Published: *25 September 2020*

Citation:
Fleckenstein J, Meyer J, Jansen T, Keller S and Köller O (2020) Is a Long Essay Always a Good Essay? The Effect of Text Length on Writing Assessment. Front. Psychol. 11:562462. doi: 10.3389/fpsyg.2020.562462

INTRODUCTION

Judgments of students' writing are influenced by a variety of text characteristics, including text length. The relationship between such (superficial) aspects of written responses and the assessment of text quality has been a controversial issue in different areas of educational research. Both in the area of educational measurement and of language technology, text length has been shown to strongly influence text ratings by trained human raters as well as computer algorithms used to score texts automatically (Chodorow and Burstein, 2004; Powers, 2005; Kobrin et al., 2011; Guo et al., 2013). In the context of classroom language learning and instruction, studies have found effects of text length on teachers' diagnostic judgments (e.g., grades; Marshall, 1967; Osnes, 1995; Birkel and Birkel, 2002; Pohlmann-Rother et al., 2016). In all these contexts, the underlying question is a similar one: Should text length be considered when judging students' writing – or is it a source of judgment bias? The objective of this paper is to investigate to what degree text length is a construct-relevant aspect of writing competence, or to what extent it erroneously influences judgments.

Powers (2005) recommends both correlational and experimental approaches for establishing the relevance of response length in the evaluation of written responses: "the former for ruling out response length (and various other factors) as causes of response quality (by virtue of their lack of relationship) and the latter for establishing more definitive causal links" (p. 7). This paper draws on data from both recommended approaches: A correlational analysis of a large-scale dataset [MEWS; funded by the German Research Foundation (Grant Nr. CO 1513/12-1) and the Swiss National Science Foundation (Grant Nr. 100019L_162675)] based on expert text quality ratings on the one hand, and an experimental study with untrained pre-service teachers on the other. It thereby incorporates the measurement perspective with the classroom perspective. In the past, (language) assessment research has been conducted within different disciplines that rarely acknowledged each other. While some assessment issues are relevant for standardized testing in large-scale contexts only, others pertain to research on teaching and classroom instruction as well. Even though their assessments may serve different functions (e.g., formative vs. summative or low vs. high stakes), teachers need to be able to assess students' performance accurately, just as well as professional raters in standardized texts. Thus, combining these different disciplinary angles and looking at the issue of text length from a transdisciplinary perspective can be an advantage for all the disciplines involved. Overall, this paper aims to present a comprehensive picture of the role of essay length in human and automated essay scoring, which ultimately amounts to a discussion of the elusive "gold standard" in writing assessment.

THEORETICAL BACKGROUND

Writing assessment is about identifying and evaluating features of a written response that indicate writing quality. Overall, previous research has demonstrated clear and consistent associations between linguistic features on the one hand, and writing quality and development on the other. In a recent literature review, Crossley (2020) showed that higher rated essays typically include more sophisticated lexical items, more complex syntactic features, and greater cohesion. Developing writers also show movements toward using more sophisticated words and more complex syntactic structures. The studies presented by Crossley (2020) provide strong indications that linguistic features in texts can afford important insights into writing quality and development. Whereas linguistic features are generally considered to be construct-relevant when it comes to assessing writing quality, there are other textual features whose relevance to the construct is debatable. The validity of the assessment of students' competences is negatively affected by construct-irrelevant factors that influence judgments (Rezaei and Lovorn, 2010). This holds true for professional raters in the context of large-scale standardized writing assessment as well as for teacher judgments in classroom writing assessment (both formative or summative). Assigning scores to students' written responses is a challenging task as different text-inherent factors influence the accuracy of the raters' or teachers' judgments (e.g., handwriting, spelling: Graham et al., 2011; length, lexical diversity: Wolfe et al., 2016). Depending on the construct to be assessed, the influence of these aspects can be considered judgment bias. One of the most relevant and well-researched text-inherent factors influencing human judgments is text length. Crossley (2020) points out that his review does "not consider text length as a linguistic feature while acknowledging that text length is likely the strongest predictor of writing development and quality." Multiple studies have found a positive relationship between text length and human ratings of text quality, even when controlling for language proficiency (Chenoweth and Hayes, 2001; McCutchen et al., 2008; McNamara et al., 2015). It is still unclear, however, whether the relation between text length and human scores reflects a true relation between text length and text quality (appropriate heuristic assumption) or whether it stems from a bias in human judgments (judgment bias assumption). The former suggests that text length is a construct-relevant factor and that a certain length is needed to effectively develop a point of view on the issue presented in the essay prompt, and this is one of the aspects taken into account in the scoring (Kobrin et al., 2007; Quinlan et al., 2009). The latter claims that text length is either completely or partly irrelevant to the construct of writing proficiency and that the strong effect it has on human judgment can be considered a bias (Powers, 2005). In the context of large-scale writing assessment, prompt-based essay tasks are often used to measure students' writing competence (Guo et al., 2013). These essays are typically scored by professionally trained raters. These human ratings have been shown to be strongly correlated with essay length, even if this criterion is not represented in the assessment rubric (Chodorow and Burstein, 2004; Kobrin et al., 2011). In a review of selected studies addressing the relation between length and quality of constructed responses, Powers (2005) showed that most studies found correlations within the range of $r = 0.50$ to $r = 0.70$. For example, he criticized the SAT essay for encouraging wordiness as longer essays tend to score higher. Kobrin et al. (2007) found the number of words to explain 39% of the variance in the SAT essay score. The authors argue that essay length is one of the aspects taken into account in the scoring as it takes a certain length to develop an argument. Similarly, Deane (2013) argues in favor of regarding writing fluency a construct-relevant factor (also see Shermis, 2014; McNamara et al., 2015). In an analytical rating of text quality, Hachmeister (2019) could showed that longer texts typically contain more cohesive devices, which has a positive impact on ratings of text quality. In the context of writing assessment in primary school, Pohlmann-Rother et al. (2016) found strong correlations between text length and holistic ratings of text quality ($r = 0.62$) as well as the semantic-pragmatic analytical dimension ($r = 0.62$). However, they found no meaningful relationship between text length and language mechanics (i.e., grammatical and orthographical correctness; $r = 0.09$).

Text length may be considered especially construct-relevant when it comes to writing in a foreign language. Because of the constraints of limited language knowledge, writing in a foreign language may be hampered because of the need to focus on language rather than content (Weigle, 2003). Silva (1993), in a review of differences between writing in a first and second language, found that writing in a second language tends to be "more constrained, more difficult, and less effective" (p. 668) than writing in a first language. The necessity of devoting cognitive resources to issues of language may mean that not as much attention can be given to higher order issues such as content or organization (for details of this debate, see Weigle, 2003, p. 36 f.). In that context, the ability of writing longer texts may be legitimately considered as indicative of higher competence in a foreign language, making text length a viable factor of assessment. For example, Ruegg and Sugiyama (2010) showed that the main predictors of the content score in English foreign language essays were first, organization and second, essay length.

The relevance of this issue has further increased as systems of automated essay scoring (AES) have become more widely used in writing assessment. These systems offer a promising way to complement human ratings in judging text quality (Deane, 2013). However, as the automated scoring algorithms are typically modeled after human ratings, they are also affected by human judgment bias. Moreover, it has been criticized that, at this point, automated scoring systems mainly count words when computing writing scores (Perelman, 2014). Chodorow and Burstein (2004), for example, showed that 53% of the variance in human ratings can be explained by automated scoring models that use only the number of words and the number of words squared as predictors. Ben-Simon and Bennett (2007) provided evidence from National Assessment of Educational Progress (NAEP) writing test data that standard, statistically created e-rater models weighed essay length even more strongly than human raters (also see Perelman, 2014).

Bejar (2011) suggests that a possible tendency to reward longer texts could be minimized through the training of raters with responses at each score level that vary in length. However, Barkaoui (2010) and Attali (2016) both compared the holistic scoring of experienced vs. novice raters and – contrary to expectations – found that the correlation between essay length and scores was slightly stronger for the experienced group. Thus, the question of whether professional experience and training counteract or even reinforce the tendency to overvalue text length in scoring remains open.

Compared to the amount of research on the role of essay length in human and automated scoring in large-scale high-stakes contexts, little attention has been paid to the relation of text length and quality in formative or summative assessment by teachers. This is surprising considering the relevance of the issue for teachers' professional competence: In order to assess the quality of students' writing, teachers must either configure various aspects of text quality in a holistic assessment or hold them apart in an analytic assessment. Thus, they need to have a concept of writing quality appropriate for the task and they need to be aware of the construct-relevant and -irrelevant criteria (cf. the lens model; Brunswik, 1955). To our knowledge, only two studies have investigated the effect of text length on holistic teacher judgments, both of which found that longer texts receive higher grades. Birkel and Birkel (2002) found significant main effects of text length (long, medium, short) and spelling errors (many, few) on holistic teacher judgments. Osnes (1995) reported effects of handwriting quality and text length on grades.

Whereas research on the text length effect on classroom writing assessment is scarce, a considerable body of research has investigated how other text characteristics influence teachers' assessment of student texts. It is well-demonstrated, for example, that pre-service and experienced teachers assign lower grades to essays containing mechanical errors (Scannell and Marshall, 1966; Marshall, 1967; Cumming et al., 2002; Rezaei and Lovorn, 2010). Scannell and Marshall (1966) found that pre-service teachers' judgments were affected by errors in punctuation, grammar and spelling, even though they were explicitly instructed to grade on content alone. More recently, Rezaei and Lovorn (2010) showed that high quality essays containing more structural, mechanical, spelling, and grammatical errors were assigned lower scores than texts without errors even in criteria relating solely to content. Teachers failed to distinguish between formal errors and the independent quality of content in a student essay. Similarly, Vögelin et al. (2018, 2019) found that lexical features and spelling influenced not only holistic teacher judgments of students' writing in English as a second or foreign language, but also their assessment of other analytical criteria (e.g., grammar). Even though these studies do not consider text length as a potential source of bias, they do show that construct-irrelevant aspects influence judgments of teachers.

THIS RESEARCH

Against this research background, it remains essential to investigate whether the relation between essay length and text quality represents a true relationship or a bias on the part of the rater or teacher (Wolfe et al., 2016). First, findings of correlational studies can give us an indication of the effect of text length on human ratings above and beyond language proficiency variables. Second, going beyond correlational findings, there is a need for experimental research that examines essay responses on the same topic differing only in length in order to establish causal relationships (Kobrin et al., 2007). The present research brings together both of these approaches.

This paper comprises two studies investigating the role of essay length in foreign language assessment using an interdisciplinary perspective including the fields of foreign language education, computer linguistics, educational research, and psychometrics. Study 1 presents a secondary analysis of a large-scale dataset with N = 2,722 upper secondary school students in Germany and Switzerland who wrote essays in response to "independent writing" prompts of the internet-based Test of English as a Foreign Language (TOEFL iBT). It investigates the question of how several indicators of students' English proficiency (English grade, reading and listening comprehension, self-concept) are related to the length of their essays (word count). It further investigates whether or not essay length accounts for variance in text quality scores (expert ratings) even when controlling for English language proficiency and other variables (e.g., country, gender, cognitive ability). A weak relationship of proficiency and length as well as a large proportion of variance in text quality explained by length beyond proficiency would be in favor of the judgment bias assumption.

Study 2 focused on possible essay length bias in an experimental setting, investigating the effect of essay length on text quality ratings when there was (per design) no relation between essay length and text quality score. Essays from Study 1 were rated by N = 84 untrained pre-service teachers, using the same TOEFL iBT rubric as the expert raters. As text quality scores were held constant within all essay length conditions, any significant effect of essay length would indicate a judgment bias. Both studies are described in more detail in the following sections.

STUDY 1

This study investigates the question of judgment bias assumption vs. appropriate heuristic assumption in a large-scale context with professional human raters. A weak relationship between text length and language proficiency would be indicative of the former assumption, whereas a strong relationship would support the latter. Moreover, if the impact of text length on human ratings was significant and substantial beyond language proficiency, this might indicate a bias on the part of the rater rather than an appropriate heuristic. Thus, Study 1 aims to answer the following research questions:

(1) How is essay length related to language proficiency?
(2) Does text length still account for variance in text quality when English language proficiency is statistically controlled for?

Materials and Methods

Sample and Procedure

The sample consisted of N = 2,722 upper secondary students (11th grade; 58.1% female) in Germany (n = 894) and Switzerland (n = 1828) from the interdisciplinary and international research project *Measuring English Writing at Secondary Level* (MEWS; for an overview see Keller et al., 2020). The target population were students attending the academic track of general education grammar schools (ISCED level 3a) in the German federal state Schleswig-Holstein as well as in seven Swiss cantons (Aargau, Basel Stadt, Basel Land, Luzern, St. Gallen, Schwyz, Zurich). In a repeated-measures design, students were assessed at the beginning (T1: August/September 2016; M_{age} = 17.34; SD_{age} = 0.87) and at the end of the school year (T2: May/June 2017; M_{age} = 18.04; SD_{age} = 0.87). The students completed computer-based tests on writing, reading and listening skills, as well as general cognitive ability. Furthermore, they completed a questionnaire measuring background variables and individual characteristics.

Measures

Writing prompt

All students answered two independent and two integrated essay writing prompts of the internet-based Test of English as a Foreign Language (TOEFL iBT®) that is administered by the Educational Testing Service (ETS) in Princeton. The task instruction was as follows: "In the writing task below you will find a question on a controversial topic. Answer the question in an essay in English. List arguments and counter-arguments, explain them and finally make it clear what your own opinion on the topic is. Your text will be judged on different qualities. These include the presentation of your ideas, the organization of the essay and the linguistic quality and accuracy. You have 30 min to do this. Try to use all of this time as much as possible." This task instruction was followed by the essay prompt. The maximum writing time was 30 min according to the official TOEFL iBT® assessment procedure. The essays were scored by trained human raters on the TOEFL 6-point rating scale at ETS. In addition to two human ratings per essay, ETS also provided scores from their automated essay scoring system (e-rater®; Burstein et al., 2013). For a more detailed description of the scoring procedure and the writing prompts see Rupp et al. (2019) and Keller et al. (2020). For the purpose of this study, we selected the student responses to the TOEFL iBT independent writing prompt "Teachers," which showed good measurement qualities (see Rupp et al., 2019). Taken together, data collections at T1 and T2 yielded N = 2,389 valid written responses to the following prompt: "A teacher's ability to relate well with students is more important than excellent knowledge of the subject being taught."

Text quality and length

The rating of text quality via human and machine scoring was done by ETS. All essays were scored by highly experienced human raters on the operational holistic TOEFL iBT rubric from 0 to 5 (Chodorow and Burstein, 2004). Essays were scored high if they were well-organized and individual ideas were well-developed, if they used specific examples and support to express learners' opinion on the subject, and if the English language was used accurately to express learners' ideas. Essays were assigned a score of 0 if they were written in another language, were generally incomprehensible, or if no text was entered.

Each essay received independent ratings by two trained human raters. If the two ratings showed a deviation of 1, the mean of the two scores was used; if they showed a deviation of 2 or more, a third rater (adjudicator) was consulted. Inter-rater agreement, as measured by quadratic weighted kappa (QWK), was satisfying for the prompt "Teachers" at both time points (QWK = 0.67; Hayes and Hatch, 1999; see Rupp et al., 2019 for further details). The mean text quality score was M = 3.35 (SD = 0.72).

Word count was used to measure the length of the essays. The number of words was calculated by the e-Rater scoring engine. The mean word count was M = 311.19 (SD = 81.91) and the number of words ranged from 41 to 727. We used the number of words rather than other measures of text length (e.g., number of letters) as it is the measure which is most frequently used in the literature: 9 out of 10 studies in the research review by Powers (2005) used word count as the criterion (also see Kobrin et al., 2007, 2011; Crossley and McNamara, 2009; Barkaoui, 2010; Attali, 2016; Wolfe et al., 2016; Wind et al., 2017). This approach ensures that our analyses can be compared with previous research.

English language proficiency and control variables

Proficiency was operationalized by a combination of different variables: English grade, English writing self-concept, reading and listening comprehension in English. The listening and reading skills were measured with a subset of items from the German National Assessment (Köller et al., 2010). The tasks require a detailed understanding of long, complex reading and listening texts including idiomatic expressions and different linguistic registers. The tests consisted of a total of 133 items for reading, and 118 items for listening that were administered in a multi-matrix-design. Each student was assessed with two rotated 15-min blocks per domain. Item parameters were estimated using longitudinal multidimensional two-parameter item response models in M*plus* version 8 (Muthén and Muthén, 1998–2012). Student abilities were estimated using 15 plausible values (PVs) per person. The PV reliabilities were 0.92 (T1) and 0.76 (T2) for reading comprehension, and 0.85 (T1) and 0.72 (T2) for listening comprehension. For a more detailed description of the scaling procedure see Köller et al. (2019).

General cognitive ability was assessed at T1 using the subtests on figural reasoning (N2; 25 items) and on verbal reasoning (V3; 20 items) of the Cognitive Ability Test (KFT 4–12 + R; Heller and Perleth, 2000). For each scale 15 PVs were drawn in a two-dimensional item response model. For the purpose of this study, the two PVs were combined to 15 overall PV scores with a reliability of 0.86.

The English writing self-concept was measured with a scale consisting of five items (e.g., "I have always been good at writing in English"; Eccles and Wigfield, 2002; Trautwein et al., 2012; $\alpha = 0.90$). Furthermore, country (Germany = 0/Switzerland = 1), gender (male = 0/female = 1) and time of measurement (T1 = 0; T2 = 1) were used as control variables.

Statistical Analyses

All analyses were conducted in M*plus* version 8 (Muthén and Muthén, 1998–2012) based on the 15PV data sets using robust maximum likelihood estimation to account for a hierarchical data structure (i.e., students clustered in classes; type = complex). Full-information maximum likelihood was used to estimate missing values in background variables. Due to the use of 15PVs, all analyses were run 15 times and then averaged (see Rubin, 1987).

Confirmatory factor analysis was used to specify a latent proficiency factor. All four proficiency variables showed substantial loadings in a single-factor measurement model (English grade: 0.67; writing self-concept: 0.73; reading comprehension: 0.42; listening comprehension: 0.51). As reading and listening comprehension were measured within the same assessment framework and could thus be expected to share mutual variance beyond the latent factor, their residuals were allowed to correlate. The analyses yielded an acceptable model fit: $\chi^2(1) = 3.65$, $p = 0.06$; CFI = 0.998, RMSEA = 0.031, SRMR = 0.006.

The relationship between text length and other independent variables was explored with correlational analysis. Multiple regression analysis with latent and manifest predictors was used to investigate the relations between text length, proficiency, and text quality.

Results

The correlation of the latent proficiency factor and text length (word count) was moderately positive: $r = 0.36$, $p < 0.01$. This indicates that more proficient students tended to write longer texts. Significant correlations with other variables showed that students tended to write longer texts at T1 ($r = -0.08$, $p < 0.01$), girls wrote longer texts than boys ($r = 0.11$, $p < 0.01$), and higher cognitive ability was associated with longer texts ($r = 0.07$, $p < 0.01$). However, all of these correlations were very weak as a general rule. The association of country and text length was not statistically significant ($r = -0.06$, $p = 0.10$).

Table 1 presents the results of the multiple linear regression of text quality on text length, proficiency and control variables. The analysis showed that proficiency and the covariates alone explained 38 percent of the variance in text quality ratings, with the latent proficiency factor being by far the strongest predictor (Model 1). The effect of text length on the text quality score was equally strong when including the control variables but not proficiency in the model (Model 2). When both the latent proficiency factor and text length were entered into the regression model (Model 3), the coefficient of text length was reduced but remained significant and substantial, explaining an additional 24% of the variance ($\Delta R^2 = 0.24$ from Model 1 to Model 3). Thus, text length had an incremental effect on text quality beyond a latent English language proficiency factor.

Discussion

Study 1 approached the issue of text length by operationalizing the construct of English language proficiency and investigating how it affects the relationship of text length and text quality. This can give us an idea of how text length may influence human judgments even though it is not considered relevant to the construct of writing competence. These secondary analyses of an existing large-scale dataset yielded two central findings: First, text length was only moderately associated with language proficiency. Second, text length strongly influenced writing performance beyond proficiency. Thus, it had an impact on the assigned score that was not captured by the construct of proficiency. These findings could be interpreted in favor of the judgment bias assumption as text length may include both construct-irrelevant and construct-relevant information.

The strengths of this study were the large sample of essays on the same topic and the vast amount of background information that was collected on the student writers (proficiency and control variables). However, there were three major limitations: First, the proficiency construct captured different aspects of English language competence (reading and listening comprehension,

TABLE 1 | Linear regression of text quality on text length, English language proficiency, and control variables: standardized regression coefficients (β) and standard errors (SE).

Predictors/R^2	β (*SE*)		
	Model 1	**Model 2**	**Model 3**
Text length		0.59 (0.02)**	0.41 (0.02)**
English language proficiency	0.65 (0.03)**		0.56 (0.03)**
Country	0.07 (0.02)**	0.14 (0.02)**	0.12 (0.02)**
Gender	0.07 (0.02)**	0.05 (0.02)**	0.02 (0.02)
Cognitive ability	−0.14 (0.03)**	0.14 (0.02)**	−0.08 (0.03)*
Time (T1/T2)	0.03 (0.02)	0.08 (0.02)**	0.06 (0.02)**
R^2	0.38 (0.04)**	0.40 (0.02)**	0.62 (0.02)**

**$p < 0.01$; *$p < 0.05$.

writing self-concept, grade), but that operationalization was not comprehensive. Thus, the additional variance explained by text length may still have been due to other aspects that could not be included in the analyses as they were not in the data. Further research with a similar design (primary or secondary analyses) should use additional variables such as grammar/vocabulary knowledge or writing performance in the first language.

The second limitation was the correlational design, which does not allow a causal investigation of the effect of text length on text quality ratings. Drawing inferences which are causal in nature would require an experimental environment in which, for example, text quality is kept constant for texts of different lengths. For that reason, Study 2 was conducted exactly in such a research design.

Last but not least, the question of transferability of these findings remains open. Going beyond standardized large-scale assessment, interdisciplinary research requires us to look at the issue from different perspectives. Findings pertaining to professional raters may not be transferable to teachers, who are required to assess students' writing in a classroom context. Thus, Study 2 drew on a sample of preservice English teachers and took a closer look at how their ratings were impacted by text length.

STUDY 2

Research Questions

In Study 2, we investigated the judgment bias assumption vs. the appropriate heuristic assumption of preservice teachers. As recommended by Powers (2005), we conducted an experimental study in addition to the correlational design used in Study 1. As text quality scores were held constant within all essay length conditions, any significant effect of essay length would be in favor of the judgment bias assumption. The objective of this study was to answer the following research questions:

(1) How do ratings of pre-service teachers correspond to expert ratings?
(2) Is there an effect of text length on the text quality ratings of preservice English teachers, when there is (per design) no relation between text length and text quality (main effect)?
(3) Does the effect differ for different levels of writing performance (interaction effect)?

Materials and Methods

Participants and Procedure

The experiment was conducted with $N = 84$ pre-service teachers ($M_{Age} = 23$ years; 80% female), currently enrolled in a higher education teacher training program at a university in Northern Germany. They had no prior rating experience of this type of learner texts. The experiment was administered with the Student Inventory ASSET (Jansen et al., 2019), an online tool to assess students' texts within an experimental environment. Participants were asked to rate essays from the MEWS project (see Study 1) on the holistic rubric used by the human raters at ETS (0–5; https://www.ets.org/s/toefl/pdf/toefl_writing_rubrics.pdf). Every participant had to rate 9 out of 45 essays in randomized order, representing all possible combinations of text quality and text length. Before the rating process began, participants were given information about essay writing in the context of the MEWS study (school type; school year; students' average age; instructional text) and they were presented the TOEFL writing rubric as the basis for their judgments. They had 15 min to get an overview of all nine texts before they were asked to rate each text on the rubric. Throughout the rating process, they were allowed to highlight parts of the texts.

The operationalization of text quality and text length as categorical variables as well as the procedure of selecting an appropriate essay sample for the study is explained in the following.

Text Length and Text Quality

The essays used in the experiment were selected on the basis of the following procedure, which took both text quality and text length as independent variables into account. The first independent variable of the essay (overall text quality) was operationalized via scores assigned by two trained human raters from ETS on a holistic six-point scale (0–5; see Study 1 and Appendix A). In order to measure the variable as precisely as possible, we only included essays for which both human raters had assigned the same score, resulting in a sample of $N = 1{,}333$ essays. As a result, three gradations of text quality were considered in the current study: lower quality (score 2), medium quality (score 3) and higher quality (score 4). The corpus included only few texts (10.4%) with the extreme scores of 0, 1, and 5; these were therefore excluded from the essay pool. We thus realized a 3×3 factorial within-subjects design. The second independent variable text length was measured via the word count of the essays, calculated by the e-rater (c) scoring engine. As with text quality, this variable was subdivided in three levels: rather short texts (s), medium-length texts (m), and long texts (l). All available texts were analyzed regarding their word count distribution. Severe outliers were excluded. The remaining $N = 1308$ essays were split in three even groups: the lower (=261 words), middle (262–318 words) and upper third (=319 words). **Table 2** shows the distribution of essays for the resulting combinations of text length and text score.

Selection of Essays

For each text length group (s, m, and l), the mean word count across all three score groups was calculated. Then, the score group

TABLE 2 | Distribution of essays in the sample contingent on text quality and text length groupings.

Text quality	Text length			
	Short (s)	**Medium (m)**	**Long (l)**	**Total**
Low (2)	$n = 147$	$n = 33$	$n = 15$	$n = 195$
Medium (3)	$n = 260$	$n = 299$	$n = 204$	$n = 763$
High (4)	$n = 22$	$n = 110$	$n = 218$	$n = 350$
Total	$n = 429$	$n = 442$	$n = 437$	$N = 1{,}308$

Number of essays excluding text quality scores 0, 1, and 5 as well as severe outliers concerning word count.

(2, 3, or 4) with the smallest number of essays in a text length group was taken as reference (e.g., n = 22 short texts of high quality or n = 15 long texts of low quality). Within each text length group, the five essays being – word count-wise – closest to the mean of the reference were chosen for the study. This was possible with mostly no or only minor deviations. In case of multiple possible matches, the essay was selected at random. This selection procedure resulted in a total sample of 45 essays, with five essays for each combination of score group (2, 3, 4) and length group (s, m, l).

Results

A repeated-measures ANOVA with two independent variables (text quality and text length) was conducted to test the two main effects and their interaction on participants' ratings (see **Table 3**). Essay ratings were treated as a within-subject factor, accounting for dependencies of the ratings nested within raters. The main effect of text quality scores on participants' ratings showed significant differences between the three text quality conditions (*low*, *medium*, *high*) that corresponded to expert ratings; $F(2, 82) = 209.04$, $p < 0.001$, $d = 4.52$. There was also a significant main effect for the three essay length conditions (*short*, *medium*, *long*); $F(2, 82) = 9.14$, $p < 0.001$, $d = 0.94$. Contrary to expectations, essay length was negatively related to participants' ratings, meaning that shorter texts received higher scores than longer texts. The interaction of text quality and text length also had a significant effect; $F(4, 80) = 3.93$, $p < 0.01$, $d = 0.89$. *Post-hoc* tests revealed that texts of low quality were especially impacted by essay length in a negative way (see **Figure 1**).

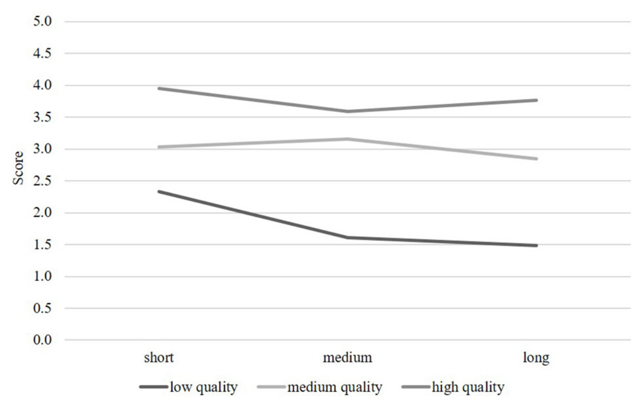

FIGURE 1 | Visualization of the interaction between text length and text quality.

Discussion

The experiment conducted in Study 2 found a very strong significant main effect for text quality, indicating a high correspondence of pre-service teachers' ratings with the expert ratings of text quality. The main effect of text length was also significant, but was qualified by a significant interaction effect text quality x text length, indicating that low quality texts were rated even more negative the longer they were. This negative effect of text length was contrary to expectations: The pre-service teachers generally tended to assign higher scores to shorter texts. Thus, they seemed to value shorter texts over longer texts. However, this was mainly true for texts of low quality.

TABLE 3 | Participants' ratings of text quality: means (M) and standard deviations (SD).

Text quality	Text length			
	Short (s)	**Medium (m)**	**Long (l)**	**Row total**
	***M* (*SD*)**	***M* (*SD*)**	***M* (SD)**	***M* (*SD*)**
Low (2)	2.33 (1.38)^{a}	1.61 (0.92)^{b}	1.49 (1.17)^{b}	1.81 (1.23)
Medium (3)	3.04 (0.96)^{a}	3.15 (1.41)^{a}	2.85 (1.23)^{a}	3.01 (1.22)
High (4)	3.95 (1.10)^{a}	3.58 (1.12)^{b}	3.76 (0.94)^{b}	3.77 (1.06)
Column total	3.11 (1.33)^{a}	2.78 (1.44)^{b}	2.70 (1.46)^{b}	

Different superscript letters within a row indicate significant mean differences ($p < 0.05$).

These findings were surprising against the research background that would suggest that longer texts are typically associated with higher scores of text quality, particularly in the context of second language writing. Therefore, it is even more important to discuss the limitations of the design before interpreting the results: First, the sample included relatively inexperienced pre-service teachers. Further research is needed to show whether these findings are transferable to in-service teachers with reasonable experience in judging students' writing. Moreover, further studies could use assessment rubrics that teachers are more familiar with, such as the CEFR (Council of Europe, 2001; also see Fleckenstein et al., 2020). Second, the selection process of essays may have reduced the ecological validity of the experiment. As there were only few long texts of low quality and few short texts of high quality in the actual sample (see **Table 2**), the selection of texts in the experimental design was – to some degree – artificial. This could also have influenced the frame of reference for the pre-service teachers as the distribution of the nine texts was different from what one would find naturally in an EFL classroom. Third, the most important limitation of this study is the question of the reference norm, a point which applies to studies of writing assessment in general. In our study, writing quality was operationalized using expert ratings, which have been shown to be influenced by text length in many investigations as well as in Study 1. If the expert ratings are biased themselves, the findings of this study may also be interpreted as pre-service teachers (unlike expert raters) not showing a text length bias at all: shorter texts should receive higher scores than longer ones if the quality assigned by the expert raters is held constant. We discuss these issues concerning the reference norm in more detail in the next section.

All three limitations may have affected ratings in a way that could have reinforced a negative effect of text length on text quality ratings. However, as research on the effect of text length on teachers' judgments is scarce, we should consider the possibility that the effect is actually different from the (positive) one typically found for professional human raters. There are a number of reasons to assume differences in the rating processes that are discussed in more detail in the following section. Furthermore, we will discuss what this means in terms of the validity of the gold standard in writing assessment.

GENERAL DISCUSSION

Combining the results of both studies, we have reason to assume that (a) text length induces judgment bias and (b) the effect of text length largely depends on the rater and/or the rating context. More specifically, the findings of the two studies can be summarized as follows: Professional human raters tend to reward longer texts beyond the relationship of text length and proficiency. Compared to this standard, inexperienced EFL teachers tend to undervalue text length, meaning that they sanction longer texts especially when text quality is low. This in turn may be based on an implicit expectation deeply ingrained in the minds of many EFL teachers: that writing in a foreign language is primarily about avoiding mistakes, and that longer texts typically contain more of them than shorter ones (Keller, 2016). Preservice teachers might be particularly afflicted with this view of writing as they would have experienced it as learners up-close and personal, not too long ago. Both findings point toward the judgment bias assumption, but with opposite directions. These seemingly contradictory findings lead to interesting and novel research questions – both in the field of standardized writing assessment and in the field of teachers' diagnostic competence.

Only if we take professional human ratings as reliable benchmark scores can we infer that teachers' ratings are biased (in a negative way). If we consider professional human ratings to be biased themselves (in a positive way), then the preservice teachers' judgments might appear to be unbiased. However, it would be implausible to assume that inexperienced teachers' judgments are less biased than those of highly trained expert raters. Even if professional human ratings are flawed themselves, they are the best possible measure of writing quality, serving as a reference even for NLP tools (Crossley, 2020). It thus makes much more sense to consider the positive impact of text length on professional human ratings – at least to a degree – an appropriate heuristic. This means that teachers' judgments would generally benefit from applying the same heuristic when assessing students' writing, as long as it does not become a bias.

In his literature review, Crossley (2020) sees the nature of the writing task to be among the central limitations when it comes to generalizing findings in the context of writing assessment. Written responses to standardized tests (such as the TOEFL) may produce linguistic features that differ from writing samples produced in the classroom or in other, more authentic writing environments. Moreover, linguistic differences may also occur depending on a writing sample being timed or untimed. Timed samples provide fewer opportunities for planning, revising, and development of ideas as compared to untimed samples, where students are more likely to plan, reflect, and revise their writing. These differences may surface in timed writing in such a way that it would be less cohesive and less complex both lexically and syntactically.

In the present research, such differences may account for the finding that pre-service teachers undervalue text length compared to professional raters. Even though the participants in Study 2 were informed about the context in which the writing samples were collected, they may have underestimated the challenges of a timed writing task in an unfamiliar format. In the context of their own classrooms, students rarely have strict time limitations when working on complex writing tasks. If they do, in an exam consisting of an argumentative essay, for example, it is usually closer to 90 min than to 30 min (at least in the case of the German pre-service teachers who participated in this study). Thus, text length may not be a good indicator of writing quality in the classroom. On the contrary, professional raters may value length as a construct-relevant feature of writing quality in a timed task, for example as an indicator of writing fluency (see Peng et al., 2020).

Furthermore, text length as a criterion of quality cannot be generalized over different text types at random. The genres which are taught in EFL courses, or assessed in EFL exams, differ considerably with respect to expected length. In five paragraph essays, for example, developing an argument requires a certain scope and attention to detail, so that text length is a highly salient feature for overall text quality. The same might not be true for e-mail writing, a genre frequently taught in EFL classrooms (Fleckenstein et al., in press). E-mails are usually expected to be concise and to the point, so that longer texts might seem prolix, or rambling. Such task-specific demands need to be taken into account when it comes to interpreting our findings. The professional raters employed in our study were schooled extensively for rating five-paragraph essays, which included a keen appreciation of text length as a salient criterion of text quality. The same might not be said of classroom teachers, who encounter a much wider range of genres in their everyday teaching and might therefore be less inclined to consider text length as a relevant feature. Further research should consider different writing tasks in order to investigate whether text length is particularly important to the genre of the argumentative essay.

Our results underscore the importance of considering whether or not text length should be taken into account for different contexts of writing assessment. This holds true for classroom assessment, where teachers should make their expectations regarding text length explicit, as well as future studies with professional raters. Crossley (2020) draws attention to the transdisciplinary perspective of the field as a source for complications: "The complications arise from the interdisciplinary nature of this type of research which often combines writing, linguistics, statistics, and computer science fields. With so many fields involved, it is often easy to overlook confounding factors" (p. 428). The present research shows how the answer to one and the same research question – How does text length influence human judgment? – can be very different from different perspectives and within different areas of educational research. Depending on the population (professional raters vs.

pre-service teachers) and the methodology (correlational analysis vs. experimental design), our findings illustrate a broad range of possible investigations and outcomes. Thus, it is a paramount example of why interdisciplinary research in education is not only desirable but imperative. Without an interdisciplinary approach, our view of the text length effect would be uni-dimensional and fragmentary. Only the combination of different perspectives and methods can live up to the demands of a complex issue such as writing assessment, identify research gaps, and challenge research traditions. Further research is needed to investigate the determinants of the strength and the direction of the bias. It is necessary to take a closer look at the rating processes of (untrained) teachers and (trained) raters, respectively, in order to investigate similarities and differences. Research pertaining to judgment heuristics/biases can be relevant for both teacher and rater training. However, the individual concerns and characteristics of the two groups need to be taken into account. This could be done, for example, by directly comparing the two groups in an experimental study. Both in teacher education and in text assessment studies, we should have a vigorous discussion about how appropriate heuristics of expert raters can find their way into the training of novice teachers and inexperienced raters in an effort to reduce judgement bias.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ministry of Education, Science and Cultural Affairs of the German federal state Schleswig-Holstein. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JF analyzed the data and wrote the manuscript. TJ and JM collected the experimental data for Study 2 and supported the data analysis. SK and OK provided the dataset for Study 1. TJ, JM, SK, and OK provided feedback on the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Attali, Y. (2016). A comparison of newly-trained and experienced raters on a standardized writing assessment. *Lang. Test.* 33, 99–115. doi: 10.1177/0265532215582283

Barkaoui, K. (2010). Explaining ESL essay holistic scores: a multilevel modeling approach. *Lang. Test.* 27, 515–535. doi: 10.1177/0265532210368717

Bejar, I. I. (2011). A validity-based approach to quality control and assurance of automated scoring. *Assess. Educ.* 18, 319–341. doi: 10.1080/0969594x.2011.555329

Ben-Simon, A., and Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *J. Technol. Learn. Asses.* 6 [Epub ahead of print].

Birkel, P., and Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychol. Erzieh. Unterr.* 49, 219–224.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* 62, 193–217. doi: 10.1037/h0047470

Burstein, J., Tetreault, J., and Madnani, N. (2013). "The E-rater® automated essay scoring system," in *Handbook of Automated Essay Evaluation*, eds M. D. Shermis and J. Burstein (Abingdon: Routledge), 77–89.

Chenoweth, N. A., and Hayes, J. R. (2001). Fluency in writing: generating text in L1 and L2. *Written Commun.* 18, 80–98. doi: 10.1177/0741088301018001004

Chodorow, M., and Burstein, J. (2004). Beyond essay length: evaluating e-rater® 's performance on toefl® essays. *ETS Res. Rep.* 2004, i–38. doi: 10.1002/j.2333-8504.2004.tb01931.x

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Cambridge, MA: Cambridge University Press.

Crossley, S. (2020). Linguistic features in writing quality and development: an overview. *J. Writ. Res.* 11, 415–443. doi: 10.17239/jowr-2020.11.03.01

Crossley, S. A., and McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *J. Second. Lang. Writ.* 18, 119–135. doi: 10.1016/j.jslw.2009.02.002

Cumming, A., Kantor, R., and Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: a descriptive framework. *Modern Lang. J.* 86, 67–96. doi: 10.1111/1540-4781.00137

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assess. Writ.* 18, 7–24. doi: 10.1016/j.asw.2012.10.002

Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annu. Rev.Psychol.* 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153

Fleckenstein, J., Keller, S., Krüger, M., Tannenbaum, R. J., and Köller, O. (2020). Linking TOEFL iBT® writing scores and validity evidence from a standard setting study. *Assess. Writ.* 43:100420. doi: 10.1016/j.asw.2019.100420

Fleckenstein, J., Meyer, J., Jansen, T., Reble, R., Krüger, M., Raubach, E., et al. (in press). "Was macht Feedback effektiv? Computerbasierte Leistungsrückmeldung anhand eines Rubrics beim Schreiben in der Fremdsprache Englisch," in *Tagungsband Bildung, Schule und Digitalisierung*, eds K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König, and D. Schmeinck (Münster: Waxmann).

Graham, S., Harris, K. R., and Hebert, M. (2011). It is more than just the message: presentation effects in scoring writing. *Focus Except. Child.* 44, 1–12.

Guo, L., Crossley, S. A., and McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: a comparison study. *Assess. Writ.* 18, 218–238. doi: 10.1016/j.asw.2013.05.002

Hachmeister, S. (2019). "Messung von Textqualität in Ereignisberichten," in *Schreibkompetenzen Messen, Beurteilen und Fördern (6. Aufl)*, eds I. Kaplan and I. Petersen (Münster: Waxmann Verlag), 79–99.

Hayes, J. R., and Hatch, J. A. (1999). Issues in measuring reliability: Correlation versus percentage of agreement. *Writt. Commun.* 16, 354–367. doi: 10.1177/0741088399016003004

Heller, K. A., and Perleth, C. (2000). *KFT 4-12+ R Kognitiver Fähigkeitstest für 4. Bis 12. Klassen, Revision*. Göttingen: Beltz Test.

Jansen, T., Vögelin, C., Machts, N., Keller, S. D., and Möller, J. (2019). Das Schülerinventar ASSET zur Beurteilung von Schülerarbeiten im Fach Englisch: Drei experimentelle Studien zu Effekten der Textqualität und der Schülernamen. *Psychologie in Erziehung und Unterricht* 66, 303–315. doi: 10.2378/peu2019.art21d

Keller, S. (2016). Measuring Writing at Secondary Level (MEWS). Eine binationale Studie. *Babylonia* 3, 46–48.

Keller, S. D., Fleckenstein, J., Krüger, M., Köller, O., and Rupp, A. A. (2020). English writing skills of students in upper secondary education: results from an empirical study in Switzerland and Germany. *J. Second Lang. Writ.* 48:100700. doi: 10.1016/j.jslw.2019.100700

Kobrin, J. L., Deng, H., and Shaw, E. J. (2007). Does quantity equal quality? the relationship between length of response and scores on the SAT essay. *J. Appl. Test. Technol.* 8, 1–15. doi: 10.1097/nne.0b013e318276dee0

Kobrin, J. L., Deng, H., and Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assess. Writ.* 16, 154–169. doi: 10.1016/j.asw.2011.01.001

Köller, O., Fleckenstein, J., Meyer, J., Paeske, A. L., Krüger, M., Rupp, A. A., et al. (2019). Schreibkompetenzen im Fach Englisch in der gymnasialen Oberstufe. *Z. Erziehungswiss.* 22, 1281–1312. doi: 10.1007/s11618-019-00910-3

Köller, O., Knigge, M., and Tesch, B. (eds.). (2010). *Sprachliche Kompetenzen im Ländervergleich*. Germany: Waxmann.

Marshall, J. C. (1967). Composition errors and essay examination grades re-examined. *Am. Educ. Res. J.* 4, 375–385. doi: 10.3102/00028312004004375

McCutchen, D., Teske, P., and Bankston, C. (2008). "Writing and cognition: implications of the cognitive architecture for learning to write and writing to learn," in *Handbook of research on Writing: History, Society, School, Individual, Text*, ed. C. Bazerman (Milton Park: Taylor & Francis Group), 451–470.

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., and Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assess. Writ.* 23, 35–59. doi: 10.1016/j.asw.2014.09.002

Muthén, L. K., and Muthén, B. O. (1998–2012). *Mplus user's Guide*. Los Angeles: Muthén & Muthén.

Osnes, J. (1995). "Der Einflus von Handschrift und Fehlern auf die Aufsatzbeurteilung," in *Die Fragwürdigkeit der Zensurengebung (9. Aufl., S)*, ed. K. Ingenkamp (Göttingen: Beltz), 131–147.

Peng, J., Wang, C., and Lu, X. (2020). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Langu. Teach. Res.* 24, 364–381. doi: 10.1177/1362168818783341

Perelman, L. (2014). When "the state of the art" is counting words. *Assess. Writ.* 21, 104–111. doi: 10.1016/j.asw.2014.05.001

Pohlmann-Rother, S., Schoreit, E., and Kürzinger, A. (2016). Schreibkompetenzen von Erstklässlern quantitativ-empirisch erfassen-Herausforderungen und Zugewinn eines analytisch-kriterialen Vorgehens gegenüber einer holistischen Bewertung. *J. Educ. Res. Online* 8, 107–135.

Powers, D. E. (2005). Wordiness": a selective review of its influence, and suggestions for investigating its relevance in tests requiring extended written responses. *ETS Res. Rep.* i–14.

Quinlan, T., Higgins, D., and Wolff, S. (2009). Evaluating the construct-coverage of the e-rater® scoring engine. *ETS Res. Rep.* 2009, i–35. doi: 10.1002/j.2333-8504.2009.tb02158.x

Rezaei, A. R., and Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assess. Writ.* 15, 18–39. doi: 10.1016/j.asw.2010.01.003

Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: comment: a noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. *J. Am. Stat. Assoc.* 82, 543–546. doi: 10.2307/2289460

Ruegg, R., and Sugiyama, Y. (2010). Do analytic measures of content predict scores assigned for content in timed writing? *Melbourne Papers in Language Testing* 15, 70–91.

Rupp, A. A., Casabianca, J. M., Krüger, M., Keller, S., and Köller, O. (2019). Automated essay scoring at scale: a case study in Switzerland and Germany. *ETS Res. Rep. Ser.* 2019, 1–23. doi: 10.1002/ets2.12249

Scannell, D. P., and Marshall, J. C. (1966). The effect of selected composition errors on grades assigned to essay examinations. *Am. Educ. Res. J.* 3, 125–130. doi: 10.3102/00028312003002125

Shermis, M. D. (2014). The challenges of emulating human behavior in writing assessment. *Assess. Writ.* 22, 91–99. doi: 10.1016/j.asw.2014.07.002

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: the ESL research and its implications. *TESOL Q.* 27, 657–77. doi: 10.2307/3587400

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., and Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: a latent interaction modeling study. *J. Educ. Psychol.* 104, 763–777. doi: 10.1037/a0027470

Vögelin, C., Jansen, T., Keller, S. D., Machts, N., and Möller, J. (2019). The influence of lexical features on teacher judgements of ESL argumentative essays. *Assess. Writ.* 39, 50–63. doi: 10.1016/j.asw.2018.12.003

Vögelin, C., Jansen, T., Keller, S. D., and Möller, J. (2018). The impact of vocabulary and spelling on judgments of ESL essays: an analysis of teacher comments. *Lang. Learn. J.* 1–17. doi: 10.1080/09571736.2018.1522662

Weigle, S. C. (2003). *Assessing Writing*. Cambridge: Cambridge University Press.

Wind, S. A., Stager, C., and Patil, Y. J. (2017). Exploring the relationship between textual characteristics and rating quality in rater-mediated writing assessments: an illustration with L1 and L2 writing assessments. *Assess. Writ.* 34, 1–15. doi: 10.1016/j.asw.2017.08.003

Wolfe, E. W., Song, T., and Jiao, H. (2016). Features of difficult-to-score essays. *Assess. Writ.* 27, 1–10. doi: 10.1016/j.asw.2015.06.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 15 October 2020
doi: 10.3389/fpsyg.2020.576250

A Transdisciplinary Approach to Student Learning and Development in University Settings

Nancy Budwig[1] and Achu Johnson Alexander[2]*

[1] Frances L. Hiatt School of Psychology, Clark University, Worcester, MA, United States, [2] Department of Psychology, Anna Maria College, Paxton, MA, United States

This article considers the opportunities and challenges of transdisciplinary research on student learning in university settings. Fifty years ago, at a meeting in France that convened experts in education and psychology as well as higher education leaders, the term transdisciplinarity was coined as issues pertaining to the structure of the university and its impact on teaching and learning were considered. We argue that to move beyond what has already been discussed requires added insights from both the learning sciences and developmental sciences. In this article, these two areas are combined with the perspectives of higher education leaders. First, research is considered from the learning sciences on deep learning in relation to university learning and teaching. This body of work illustrates ways students need to be actively engaged in their learning and simultaneously frames teachers as facilitators of students' constructive efforts rather than disseminators of static knowledge. Second, perspectives from the developmental sciences on processes of development are reviewed, focusing on adolescence and emerging adulthood. Here we highlight the importance of considering developmental systems approaches to aspects of organizing learning at universities in light of extensive research on adolescents and emerging adults. Third, we examine new higher education frameworks that have focused on the importance of student engagement, integration and application of knowledge and the implications of these shifts for organizing higher education learning in more holistic ways, often at the national and transnational levels. In reviewing these three areas, we consider what assumptions are made about the learner, the role of teachers and others in enhancing student learning, and the interaction between learners and contexts where learning takes place. We argue that while progress is being made in undergraduate reform efforts, implementation has been uneven. To deliver on this important work will require further alignment of the sort Jantsch (1972) and Piaget (1972) claimed was central to transdisciplinary approaches, namely aligning these different areas through a systems approach that considers education as a purposeful human activity. This will involve alignment and support from the learning and developmental sciences, as well as local, national and transnational efforts and learning communities to support campus efforts.

Keywords: transdisciplinary, learning, higher education, development, interdisciplinary

OPEN ACCESS

Edited by:
Arthur C. Graesser,
University of Memphis, United States

Reviewed by:
Guadalupe López-Íñiguez,
Sibelius Academy, Finland
Roman Taraban,
Texas Tech University, United States

****Correspondence:***
Nancy Budwig
nbudwig@clarku.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *25 June 2020*
Accepted: *17 September 2020*
Published: *15 October 2020*

Citation:
Budwig N and Alexander AJ (2020) A Transdisciplinary Approach to Student Learning and Development in University Settings. Front. Psychol. 11:576250. doi: 10.3389/fpsyg.2020.576250

INTRODUCTION

Fifty years ago at a meeting in France, Jean Piaget, and other scholars studying human development and knowledge, as well as higher education leaders gathered to speak about the importance of moving beyond the disciplines in considering university teaching and innovation. In fact, the term transdisciplinarity was coined and distinguished from multidisciplinarity and interdisciplinarity at that meeting (Apostel, 1972)[1]. Fifty years later, scholars and practitioners still are discussing the importance of a transdisciplinary approach to teaching and learning. Many of the challenges discussed at the original conference on teaching and learning ring as true today, and the question can be raised how to move forward to build on the original thinking, using the vast amount of research accumulated since that time in the learning and developmental sciences to guide this work. We will argue that progress can be made if these separate treatments of teaching and learning at universities are considered in a unified way.

Interdisciplinary and transdisciplinary approaches to knowledge have developed considerably over the last 50 years. While claims for similar problems identified at that meeting still exist (Bok, 2013), there is an increasing trend for openness to discuss new views of student learning, who our students are, and the goals of university education as they relate to societal needs and students' professional and civic lives (Davidson, 2017; Wieman, 2017; Klemenèiè, 2019; Kamp, 2020). Furthermore, national and transnational involvement at the level of considering quality frameworks at the general level as well as within the disciplines help to mitigate some of the challenges of university silos and disciplinary limitations. We believe that simultaneous analysis of key foci of learning science and developmental science approaches when explicitly considered and aligned with current frameworks for innovation and advanced knowledge as they relate to organizing university structures and curricula is needed. Weaving together the sort of transdisciplinary approach Piaget and others imagined 50 years ago, holds promise to augment student learning and development, but also highlight the value of higher education in new and important ways.

[1]Piaget (1972, p. 136–138) used "interdisciplinarity to designate... cooperation among various disciplines lead to actual interactions, to a certain reciprocity of exchanges resulting in mutual enrichment. " This was distinguished from the term he considered a higher stage- "transdisciplinarity", ... (would) place these relationships within a total system without any firm boundaries between disciplines." These two terms contrast with multidisciplinary approaches which simply juxtapose different disciplinary contributions. Thirty years after the original conference where these terms were coined, a further conference took place and a new consensus emerged on how to define transdisciplinarity. Klein elaborated the systems approach of Piaget by including practitioners outside a given discipline. "The core idea of transdisciplinarity is different academic disciplines working jointly with practitioners to solve a real-world problem. It can be applied in a great variety of fields" (Klein et al., 2001: 4). This is the approach that we will adopt to transdisciplinarity in this article.

THREE PERSPECTIVES ON STUDENT LEARNING AND DEVELOPMENT IN HIGHER EDUCATION

Learning Sciences: The Importance of Deep Learning in University Settings

The organization of teaching and learning in higher education has often been described as students passively absorbing material presented by an expert, drawing on processes of memorization, learning material in ways unrelated to what they already know, and often as disconnected from other learning within and between courses. We know from discoveries by learning scientists that these traditional views of learning and the pedagogies supporting them do not work in educational settings, and yet the vast majority of students experience this passive method of delivery in university classrooms. Furthermore, the 21st century needs citizenry and workforce able not only to master knowledge, but also create knowledge. For the last two decades, the Organization for Economic Co-operation and Development (OECD) and other organizations have stressed the importance of restructuring educational institutions based on theory and research from the learning sciences (Bransford et al., 2000). Graesser et al. (2008) produced an early and particularly rich list of 25 principles in an attempt to scale current learning research into various settings – whether K-12 schools, colleges, and lifelong learning. These principles suggest the importance of having students ask deep questions, highlight the assistance students need in self-regulating their learning, and advocate for anchoring learning in real world contexts important to the student. We also know from this body of work that students bring to their learning, not only a sense of agency but also their current understandings of topical areas. Learning is gradual and involves students revising their own intuitive understandings and change conceptual frameworks in light of new knowledge (Vosniadou, 2013, 2019).

In this section, we will examine learning science research with an eye toward learning in college classrooms with a particular focus on the cognitive underpinnings of learning[2]. To illustrate this point, we will focus on what learning scientists have called deep learning, looking into research on inquiry, the organization of knowledge, and metacognition to illustrate how learning scientists have focused on teaching and learning. Though we have accumulated a lot of evidence on how people learn, far too little of it has made its way into rethinking teaching at the college level (Budwig, 2013; Wieman, 2019). We will review some findings from this literature not to provide a thorough review (which is beyond the scope of this paper) but to consider how this body of work sheds light on the role of the student in learning, as well as the role of teachers in guiding learning.

[2]Elsewhere we have discussed social aspects of learning and its relation to college learning (see Budwig, 2013, 2015).

Inquiry-Based Learning

Inquiry has been described as central to human learning in both formal and informal settings (Bransford et al., 2000). Student questioning actively engages the learner, as does students' consideration of multiple solutions found in open-ended problem solving, which are both fundamental to student success (Darling-Hammond et al., 2019). When given the chance for exploration, students learn to frame interesting questions. While student-centered, inquiry has been noted to best be achieved when teachers provide guidance such as setting broad goals and encouraging students to focus on subgoals (Collins and Kapur, 2014). In fact, a meta-analysis of research on problem-based learning reveals many instances of students not learning when left completely on their own in formal learning settings (Alfieri et al., 2011). Thus, teachers play a critical role in selecting interesting problems of inquiry and providing high quality facilitation in order to produce learning outcomes (Walker and Leary, 2009; Lu et al., 2014).

The ability to inquire is something most college professors expect by the time students enter their classrooms and yet college students vastly differ in prior experience with practicing this capacity in formal learning settings. The tendency to approach formal learning contexts with an inclination to inquire often depends on the kind of schools students have attended prior to attending college (Kritt, 2018). Most college students have extensive practice with what Bloom (1956), has called "knowledge verbs," that is, students have extensive practice with how to list, define, tell, and label information, but fall short in the capacity to inquire. These capacities are central to college learning, and yet many students arrive at and finish college insufficiently prepared to formulate appropriate questions and hypotheses, recognize assumptions and formulate premises, analyze, synthesis, and evaluate information, and formulate logical conclusions (see Eng, 2017). This lack of readiness has profound effects on students' capacities for lifelong learning and professional engagement and has been noted in employer surveys in several recent studies (Archer and Davison, 2008, National Association of Colleges and Employers, 2017; Adecco, 2019).

Organizing and Generating Knowledge

Learning scientists have helped us understand learning and teaching by also contributing to an understanding of the importance of examining how students organize knowledge. Students must actively construct new knowledge building of their earlier novice conceptions (Piaget, 1978; Darling-Hammond et al., 2019). This implies that when teachers design learning environments, consideration must be made of what existing knowledge learners bring to the process of acquiring new information (Bransford et al., 2000; Sawyer, 2006). Novices (including most students) need significant help in developing the rich and meaningful knowledge structures central to high quality learning. In contrast to experts, novices have less complex and connected knowledge structures, making it difficult for them to process information in coherent chunks as experts do (Sawyer, 2006; Collins and Kapur, 2014). Learning science work has highlighted that in addition to the questions different disciplines engage in, each discipline has distinct ways of knowing. For example, it is important for students to not only know critical findings in science classes, but also that they deeply understand the ways scientists come to that knowledge, for instance, students need to grasp how scientists use models and representations. Following up on this, learning scientists have studied how students come to understand this. This body of work has highlighted the importance of focusing on learning principles guiding authentic experiences, including in the disciplines (Greeno and Engeström, 2014). The main point here is that the organization of knowledge is something students need to figure out, and research has suggested that in optimal teaching situations, the teacher scaffolds learning of both the content of new knowledge as well as practices engaged in by experts. This helps students increasingly and gradually acquire the capacity to engage in the practices of experts in the discipline.

Metacognition and Self-Regulated Learning (SRL)

The ability to inquire and organize knowledge is dependent on a third aspect of deep learning identified by learning scientists, namely metacognition and self-regulated learning. Metacognition put simply is thinking about one's own thinking and involves a conscious attempt to regulate one's own learning (Bransford et al., 2000). Examples of metacognition and self-regulated learning include thinking about ways individuals successfully learn, the necessary sequence of learning something, what one knows already and more importantly, what one does not know. Self-explanation and having the opportunity to explain your learning both to yourself and others has been noted to aid learning (Chi et al., 1994). It also has been helpful for learners to employ metacognitive strategies involved in reflecting on what one has learned, how what is learned relates to other knowledge, and ways to apply what is learned in different contexts. To this extent, metacognition can occur before, during and after a learning event and has been noted to enhance deeper understanding of the content learned (Zimmerman, 2000; Winne and Azevedo, 2014; Darling-Hammond et al., 2019).

A central question has been whether metacognition comes naturally or must be taught to learners. Evidence of metacognition has been found in preschoolers (Wertsch et al., 1980; Flavell et al., 1993) long before they enter formal schooling, when interacting on complex problems in the context of everyday interactions with others, though much research has highlighted that the breadth and depth of metacognitive awareness is something that develops well through adolescence. It has been noted that even many college students struggle with reflective practices involved with metacognition (Schraw and Moshman, 1995). It would seem as students begin college, opportunities to engage in metacognition would be extremely useful, since students are given much more autonomy for guiding their own learning on our campuses.

It has been shown that metacognition is learned in context as one engages in authentic problems (Palincsar and Brown, 1984; Bransford et al., 2000). Across age ranges, what holds constant is that metacognitive learning typically involves scaffolding or guidance with more experienced others (often experts) modeling or guiding how one draws on metacognitive strategies in the context of solving authentic problems in context. Central to the process of transferring agency for learning from expert to learner has been the use of specific symbolic tools which themselves come to scaffold the procedural steps guiding the learner to actively pull relevant information from complex settings through a series of prompts. Sometimes these tools involve the use of multimedia (Mayer, 2014) and other times, guidance is provided more directly through tools that provide classrooms with powerful mechanisms to guide reflection, often matching the kind of disciplinary practices engaged in by experts (Bielaczyc et al., 2013). These tools scaffold interactions and support learners by suggesting steps for practice and reflection as groups work together on improving one another's ideas in classroom settings. Such tools have been used in elementary or secondary school classrooms in ways that help shift the classroom culture from a typical 20th century focus on dissemination of knowledge, to more active models of learning. The tools, employed in teacher-student dialogues, peer dialogues, as well as by learners themselves would seem to be useful in college contexts by helping to scale reflective practice in different disciplines by encouraging learners to engage in authentic inquiry, as well as integration and application of knowledge they are learning.

The Relation Between Students, Teachers, and Context in Learning Science Views of Deep Learning

While learning science research varies on a number of points, the views of deep learning described above share a similar perspective on the relation between learners and teachers characterizing their relationship as intricately linked and mutually influential. That is, deep learning involves an agentive learner who can actively draw upon their environments to examine, synthesize and build new knowledge. At the same time, research on the science of learning reviewed here has emphasized ways in which teachers and other experts, as well as mediational means and tools they employ support student learning. To this extent, learning scientists are both student centered and focused on the specific ways learning environments support student learning. More specifically, across all three areas (inquiry, organization of knowledge, and metacognition) while students actively engage in learning, it is a core aspect of learning science research to consider the specific and carefully sequenced ways teachers guide learning and gradually transfer increasing responsibility over to their students.

Learning scientists who have studied teacher knowledge (Shulman, 1986; Fishman et al., 2014) highlight the detrimental impact on learning when teachers have superficial pedagogical knowledge or content knowledge. Teachers may lack expert knowledge of the discipline or lack a solid understanding of the science of learning. To this extent, there is an important difference between instructors formally trained in the science of learning and formal experts who find themselves helping novices learn. For instance, instructors must consider whether students have appropriate prior knowledge and if so, whether and how students can activate that prior knowledge in order to learn new material. As experts, many college faculty underestimate this need (if they consider it at all) and thus do not spend time explicitly considering strategies to help students engage their prior learnings or revise inaccurate knowledge. Prior study of higher education teachers has revealed the positive effect pedagogical training can have on teachers (Postareff et al., 2007), as well as examples of how to improve classroom practices in light of learning science research (Ambrose et al., 2010). Challenges and barriers to faculty pedagogic training in this area has also been reported and is important to consider (Mälkki and Lindblom-Ylänne, 2012; Brownell and Tanner, 2017).

In sum, learning scientists have highlighted the intricate relation between student agency and how students draw on support from their environments. Especially for disciplinary learning in formal settings, students' knowledge is built up gradually based on an assumption of constructivist effort on the part of the learner, and with simultaneous guidance by a knowledgeable expert who catalyzes student learning through carefully designing environments suitable for learning. Important for scaling efforts, learning science scholarship has also highlighted tools, such as guiding questions and protocols, can assist learners with more minimal intervention on the part of individual instructors, particularly relevant in larger classroom designs.

Developmental Science Perspectives: Processes and Stages of Development Matter to Student Learning in University Settings

Developmental scientists examine behavioral and psychological aspects of human development. Recently there has been growing agreement that human development is best viewed from a systems perspective, as a process, with the organism viewed as inherently active (Witherington and Boom, 2019). In this section, we examine core features of developmental systems approaches, and then consider their application to stages of development relevant to college-attending students[3]. We use this developmental framework to examine identity formation and self-authorship during the adolescent and emerging adulthood years. Similar to our argument presented above, we will argue that college instructors rarely get any formal training about human development, and yet as we will argue, such knowledge is imperative to helping students learn. At the conclusion of this section, we will consider how developmental scientists view the role of the individual and environment in the complex process of human development and more specifically the relationship between student and teachers in our consideration of teaching and learning from a developmental lens.

Features of Developmental Systems Approaches

Systems theories provide a framework for understanding human functioning and development. The central claim relevant here is that development consists of multiple, interrelated processes that both affect and contribute to the dynamic organization of human systems (von Bertalanffy, 1972; Raeff, 2016). Rather than focusing on developmental outcomes or things that humans can and cannot do at particular ages, developmental

[3] We focus here on traditional age college students, typically 18–25 year olds, recognizing that adult learners, while not the majority, are a growing group attending college.

systems approaches emphasize *developmental processes* involved in human functioning. Within a system, the developmental processes function as a whole (Raeff, 2016; Witherington and Boom, 2019). This appreciation of human organisms as functioning wholes, also presupposes constructivist accounts assuming individuals "are active agents in their own learning and development" (Amsel and Smetana, 2011, p. 4). To this extent, development does not stem directly from biological or environmental factors; rather individual and context are viewed as mutually influencing one another, as organisms actively engage in meaning construction (Overton, 2015; Lerner, 2016; Witherington and Lickliter, 2016).

Adolescence and Emerging Adulthood: A Holistic Examination of the Learner

While it seems common sense to assume that different stages of the life cycle are made up of distinct characteristics and abilities, theory and research stemming from developmental systems approaches have cautioned about developmental stage theories and milestones. What is central when looking at particular stages is the importance of *processes of development* and not simply outcomes, and to recognize that the developmental phases differ due to the ways individual, socio-historical, and cultural systems interact over time. With these caveats in mind, we turn to consider age-related developmental theories that are relevant to learning and teaching in university settings.

Those studying adolescence from a developmental systems approach, argue that what is distinctive during adolescence, is the nature of "adolescent *coordinating activities*" (Amsel and Smetana, 2011, p. 7, italics in original). Central here are ways in which organisms cognitively repackage what was present in prior organizational states. Many students who are still adolescents find the expectations of critical thinking and evaluation of contrasting points of view expected in college learning to be difficult. Instead, they readily accept ideas passed on by experts without critically evaluating them (Hodge et al., 2009). Furthermore, over the course of college students' understanding of disciplines such as psychology and physics becomes more scientific with each year of majoring in that discipline (Amsel, 2018). Through ongoing attempts to make sense of their world, adolescents, as active agents, have opportunities, but also vulnerabilities if these co-ordinations are unsuccessful (Amsel and Smetana, 2011). For example, these vulnerabilities may show up with regard to academic underachievement (Crosnoe, 2011; Kuhn and Holman, 2011), as well as other areas such as vulnerabilities related to well-being, risk taking, and the like.

Following adolescence, Arnett (2000) argues for a distinctive stage in the lifespan that broadly represents the experiences of 18-29-year-olds (narrowly representing 18–25 year olds) as they transition into adulthood. Known as "emerging adulthood," this stage is said to result from several demographic changes, one of which he notes is the increasing rise in the number of individuals of this age attending college. Arnett argues that of five features demarcating emerging adulthood, identity explorations is one of the most central as emerging adults explore a variety of areas including education, work, and love. The central point here is that developmental scientists not only examine ongoing processes of development, but also have identified core milestones and areas of interest that are in the foreground at particular junctures in the life cycle that influence and guide learning and development.

While adolescence has been noted to be a time of enhanced cognitive achievement for students, Arum and Roksa's (2011, 2014) analysis of college-attending students suggests that college seniors spend only a minor amount of their time engaged in academics compared to time they spend socializing. Furthermore, they claim to have found only modest gains in critical thinking in emerging adults while in college. From a developmental lens, the question can be raised as to whether cognitive development has occurred in college and whether findings from the CLA test, which views cognition in isolation from every day and social settings in which it is embedded, is an appropriate way to test cognitive advances in this age group. Developmental scientists have focused on cognitive advances as part of larger processes and a developing system that includes areas like identity formation and self-authorship. The idea that during both high school and college students are pre-occupied with social relationships and questions of identity is hardly surprising to developmental scientists familiar with Erikson's theory of development or Arnett's (2000) portrayal of emerging adulthood. According to Erikson (1963) psychosocial stage theory of development, the fifth stage, which occurs during adolescence, is a time when teenagers explore questions of who they are and explore different roles and activities as they work to construct a sense of self.

Identity Formation and Self-Authorship in College-Attending Emerging Adults

A central claim we make here is that being in college helps emerging adults to engage in a period of identity exploration. It is an incubating period to try out multiple courses, majors, jobs, friends, and romantic partners before making more enduring choices (Arnett, 2016). College ideally provides a venue for individuals to explore and make long-term commitments in career, relationships, and worldviews (Arnett, 2016; Baxter Magolda and Taylor, 2016). Most importantly, college offers a fertile ground for exploring and developing skills and capacities that are necessary for making adult choices and decisions, central to this being the search for self. College attending emerging adults simultaneously engage with learning in new ways, dialoging with multiple others whose perspectives enlarge their worldview and offers the opportunity to practice skills that sets them on a path for lifelong learning as adults (Tanner et al., 2009).

A core aspect of identity exploration involves the search for a sense of self (Schwartz et al., 2016). College, if structured appropriately, can provide a space for students to engage their increasing cognitive capacities for abstract thinking toward this search as reflected in their consideration of multiple possibilities of who they are and what professions they can join in the future. Three such explorations where college provides a platform include taking on increased autonomy, developing cognitive acumen (e.g., critical inquiry, integration, and reflection), and finding identity-based work. Inherent in the aforementioned explorations of college-going emerging adults – autonomy taking, cognitive acumen, identity-based work – is the struggle to make meaning and write the initial drafts of their life stories (McAdams, 2013, 2016; Baxter Magolda and Taylor, 2016). These young people face questions of "Who am I," "How do I relate to others," and "What do I want myself to be" as they search for meaning in life and consider different worldviews, often through discussions with peers, faculty, advisors, and other staff who develop relations with students. Classroom interactions around intellectual and ethical issues offer students the opportunity to learn and select from multiple possibilities, which in turn broadens capacities for constructing a coherent life story (life authorship). These classroom and other academic experiences also can assist in developing internalized meaning structures (self-authorship). Both of these – life and self-authorship – are considered as important tasks during emerging adulthood (McAdams, 2013, 2016; Baxter Magolda and Taylor, 2016; Schwartz et al., 2016), with self-authorship central to autonomy taking (Baxter Magolda, 2001).

Self-authorship theory draws upon a constructivist framework by suggesting that young adults construct meaning in and through their interactions with the world. At this stage of development, emerging adults are passionate about compelling social issues such as social justice and equity. Self-authorship as a developmental process offers emerging adults the opportunity to connect the interactions between individuals, contexts and environments. On the pathway to self-authorship, individuals begin by blindly following and accepting formulas and knowledge presented by others, especially instructors. Over time, they come to realize the need to develop more autonomous values and beliefs in order to begin to robustly "author" one's life. This involves arriving at a "comprehensive system of belief" (Baxter Magolda, 2001, p. 155) to guide life decisions (Baxter Magolda, 2001, p. 155). Baxter Magolda argues that learning environments are central to self-authorship especially when teachers give agency to students and downplay their own authority, situate learning as a process that is relevant to students' experiences, and provide examples of teachers' modeling thinking and learning processes, while also encouraging significant space for student reflection (Baxter Magolda, 2008).

The Role of Students, Teachers, and Context in Dynamic Theories of Development

Because scholars of human development have emphasized the importance of examining human systems, rather than isolated developments (cognitive, social) of human functioning, student development must be looked at holistically and not simply in terms of the cognitive structures and processes that students bring to the classroom. Students enter the classroom not just with a mind, but fully embodied to engage with their surroundings, including seeing the classroom as a social activity. For instance, adolescents have significant challenges in coordinating budding knowledge systems and understanding the difference between facts and theories. In addition, adolescent and emerging adult students' preoccupation with social relations, identity and work are all factors relevant to understanding student learning as it is being considered in university settings. In particular, this age group is particularly interested in weaving their academics with issues of social concern, identity, and work.

Teachers also need to take into account that their students are continuing to develop such that the cognitive abilities of first year students differ significantly from seniors often in the same class. Furthermore, college access has changed such that there exists tremendous variation in individual differences in learning in a given class as well (Gagné, 2005). Universities have long overlooked this, grouping students together without considering these differences that can be productively used to augment teaching and learning. Those entering college are continuing to coordinate prior systems of development in new ways, use their burgeoning ability to reflect in increasingly abstract ways, and all of this is centrally linked to their exploration of identity, which is far more developed by senior year of college. While developmental scientists acknowledge the role of others in students' development, compared to the other perspectives in this article, their work focuses more on the individual's own construction of knowledge and efforts to engage in meaning making.

Higher Education Perspectives: Reimagining the Undergraduate Degree and Learning Outcomes Within the Disciplines

Higher education leadership is a third group that has played a significant role in considering the learning experience of students at colleges and universities. After a flurry of activity in the 1960s and 1970s (Apostel, 1972; Levine, 1980), new issues emerged as universities have been noted to serve a much broader set of regional and governmental needs (Davies et al., 2001), and with a broader range of students attending college, many underprepared for the curriculum offered (Baum and McPherson, 2019). Neoliberalism and models of higher education that are said to treat universities more like corporate organizations have become more the norm (Taylor, 2017). Such trends not only encourage specialization and compartmentalization, but have posed challenges to developing a view of higher education in terms of lifelong learning, student agency, and education has shifted from a public good to a private good.

Fresh discussions about curricular models have been increasing during the first decades of the 21st century, with some at the level of institutional planning, while global initiatives have brought together individuals from around the globe (Elkana et al., 2010; Elkana and Klopper, 2016). The most enduring reform efforts have been tied to multi-institution and governmental platforms, with the most ambitious scalability effort in the first decade of the 21st century known as the Bologna Process. The primary goal of the Bologna Process has been to bring more cooperation between countries within the European Union with the aim to increase mobility and increase recognition for a coordinated European higher education system (Zahavi and Friedman, 2019). In the United States, national attempts to reimagine a vision for higher education built on the enduring aims of a liberal education but simultaneously connecting that vision more clearly to the complex challenges of our world (Schneider, 2008) also gained significant momentum. Each of these approaches had a different relationship to the disciplines where reimaging and reform also has taken place. We turn to consider these efforts and then provide a discussion of their mutual implications for considering the relation between students, teachers, and contexts in university settings.

Visions for Undergraduate Education: New Learning Outcomes for the 21st Century[4]

The Bologna process noted above has been said to be one of the most ambitious credible attempts to scale for accountability in higher education (Adelman, 2008). Working at three levels (transnational, national, and disciplinary), a central feature of this work has been the establishment of a quality framework. Prior to this, degrees were awarded without much attention to the quality of learning. By 2003, a set of core competencies were proposed. Known as the Dublin Descriptors these included: Knowledge and understanding; Applying knowledge and understanding; Making judgments; Communication; and Lifelong learning skills. Adopted in 2005 as the Qualifications Framework of the European Higher Education Area, work shifted to implementing the common set of learning outcomes as a mechanism of improving the quality of an undergraduate degree in a transparent and holistic way across European countries. Notable here within the European context, was the decision to focus on quality assurance through what has been called the Tuning project, where tuning takes place at the levels of individual disciplines. This work has not been without its challenges that also will be discuss below (Reichert and Tauch, 2005; Kehm, 2010).

During the same period Americans also have aimed to address quality issues. National associations have been active in this arena, most notably the Association of American Colleges and Universities (AAC&U) through its Liberal Education and America's Promise (LEAP) initiative. Although for many liberal education has been associated with small residential colleges, the definition of the difference between liberal arts colleges and liberal education has been the subject of recent clarification. Liberal arts colleges are typically small and residential, while the modern notion of liberal education extends beyond particular features, or the kind of students who chose to attend those schools. Liberal learning has long been unified as an approach that promotes breadth and depth of knowledge, intellectual skills of inquiry and analysis, and personal and social responsibility. In the recent 15 years, a fourth learning outcome promoting the integration and application of knowledge has been added. Schneider (2018) argues that the aim is for all university students to experience liberal learning and not just students who either attend particular kinds of institutions, or who major in particular disciplines (such as the arts and humanities). According to Lynn Pasquerella, President of the Association of American Colleges and Universities (AAC&U, 2020, p. 2) "AAC&U remains steadfast in our conviction that a liberal education offers the best preparation for work, citizenship, and life."

One striking example of the more integrated and applied academic experiences encouraged to be at the heart of a liberal education involves student participation in what Schneider refers to as a signature work project (Schneider, 2015). Such an experience involves an extended project (at least 6 weeks of work) that reflects "cumulative and integrative learning across general and specialized studies" (Schneider, 2015, p. 6). In addition, the project should connect to a significant problem that has no clear answer and require significant student agency to solve in a way that is meaningful to the student and society, often as part of a capstone experience (Peden, 2015; Schneider, 2015). Central to the aspiration of signature work is the idea that all college students, and not just the very best, would actively engage in an integrative and applied project before leaving college. The call for applied and project based work for all college students, one that activates the agency and imagination of students can be found transnationally (Elkana and Klopper, 2016; Kamp, 2020).

Reimagining the Disciplines

As noted above, The Tuning Project locates reform efforts within the disciplines, leaving disciplines to rethinking teaching and learning. Wieman (2019, p. 65) speaking about STEM fields argues: "The acquisition of basic information is now of limited value, while complex reasoning and decision-making skills that can be broadly applied have high value in many aspects of modern society." More specifically, expertise in a discipline involves a set of cultural practices often not explicitly discussed. Until recently, teaching in the disciplines has been taken as a solitary activity left to individual tastes and styles. Current discussions in various disciplinary groups have begun to stress the importance of helping students think like an expert in the discipline, using the tools and complex reasoning that experts in a discipline employ (Wieman, 2019).

Historians have also become more explicit about the learning outcomes for the discipline. According to the American Historical Association's Tuning Project (2016) learning history is more than dates, and rather involves "a deliberative stance toward the past; the sophisticated use of information, evidence, and argumentation; and the ability to identify and explain continuity and change over time. Its professional ethics and standards demand peer review, citation, and acceptance of the provisional nature of knowledge." With direct traces to the Quality Framework Tuning Project discussed above, the American Historical Association has provided faculty with tools and resources to engage in forward moving conversations in order to reinvigorate the classrooms for students in learner-centered ways around an explicit set of disciplinary learning outcomes that are tied to enhancing student agency.

While other disciplinary groups have similarly adopted new learning outcomes, one particularly forward looking attempt is that provided for engineering education developed by Kamp (2020). Not only are quality frameworks with student learning outcomes outlined, but one finds explicit discussion of working between the gap of broad vision and on the ground implementation of educational reform. Kamp outlines both the mindsets and competencies needed by engineers in the 21st century, and highlights the important role that students must play in their own educational process, recognizing that students bring to their learning a very different approach than

[4]While our focus here is on Europe and the United States, important work in this area is going on elsewhere [see for instance, Godwin (2015); Al-Hendawi and Albertine (2019)].

students of the past. According to Kamp (2020), navigating engineering education must be viewed as a lifelong process, with individuals knowing how to continuously relearn given new contexts and developments.

Combined work going on in the disciplines points to the importance of linking both disciplinary knowledge and expertise with core outcomes that go beyond any single disciplinary field. This body of work highlights enormous new responsibilities and roles for teachers as they transition to more learner-centered strategies, which makes this work rewarding but challenging.

The Structural Changes Necessary to Deliver on This New Vision of Student Learning in Higher Education

Those in the higher education literature have been highly attuned to the processes and structures necessary for implementing the new vision of student learning. identifying at least three difficulties. Already by 1970, members of the OECD conference on interdisciplinarity noted that the siloed nature of higher education, with its focus on disciplinary units would make more integrative and applied models of learning difficult to implement. Higher education institutional structures are set up around disciplinary knowledge and practice and these structures, as well as the extensive disciplinary training experienced by faculty in such departments constrains interdisciplinary transformations. The Tuning Project and some of the work going on through disciplinary societies linked to overarching learning outcomes in part has worked to address this challenge. A second issue considered by universities as they have attempted to incorporate a curricular framework supporting student-centered learning is that institutions vary tremendously in their mission and goals and as such each campus working on such an implementation will look different – one size does not fit all, again making institutional change difficult. As Davies et al. (2001) have noted, leadership matters and there has been some conflicting messages about the importance of equity and learning along side what has been called an "arms race" of elitism, especially tied to research excellence that are in tension (Kehm, 2010). A third challenge identified is that the student-centered learning and quality delivery of ambitious outcomes require extensive time and effort on the part of faculty. Looking at the implementation of quality standards as part of the Bologna Process, Reichert and Tauch (2005) have noted the importance of campuses finding their own ways in. Similarly, AAC&U has organized cohorts of schools under grant funded initiatives such as *Faculty Leadership for Integrative Liberal Learning* and the *LEAP Challenge: Building Capstone and Signature Work*, facilitating and supporting institutions as they created and scaled the kind of integrative and applied learning experiences. In addition, a set of resources has been created through reporting out regularly on findings for other schools to draw upon (Ferren and Paris, 2015; Budwig and Jessen-Marshall, 2018).

The Relation Between Students, Teachers, and Contexts in University Settings

Higher education assumes the importance of students' constructive efforts and the importance of faculty leadership for designing environments where learning can flourish. Higher education also has given significant attention to the role of university systems and processes to support the teaching and learning efforts at universities. A growing trend in higher education is to emphasize the importance of shifting from what teachers do (e.g., teaching) to what students learn (Wright, 2011). While endorsing this view, Wright (2011) argues that it is important to recognize that leaving students with more responsibility and agency for their own learning is not an easy pivot for higher education to make (Wright, 2011). While there is recognition of the importance of student-centered learning and student agency in the construction of knowledge, the bulk of the discussion by higher education leaders has focused on the guidance received not only by individual faculty and staff, but also through intentional institutional design.

Higher education discussions primarily have focused on the need for organizational supports and structures to aid in assuring the necessary dynamic between student agency and engagement and faculty support and guidance. University leaders have recognized the lack of training faculty bring with them regarding teaching and learning in general, and for integrative and applied learning in particular. As universities have begun to strategically emphasize student-centered approaches to learning, teaching and learning centers have been built up on campuses as cross-disciplinary spaces to support and nurture quality teaching (Hutchings et al., 2011). At the same time, as one turns to more holistic approaches to student learning, and the importance of student agency and lifelong learning has led to consideration of the role others can play in student learning highlighting the need for consideration of more complex institutional structures and non-academic supports needed. Recognizing the limited feasibility of charging faculty with sole responsibility for student learning there is need for coordination when students are expected to integrate their learning and apply it to problems that often involve participation beyond university gates.

DISCUSSION: OPPORTUNITIES OF AND CHALLENGES FOR A TRANSDISCIPLINARY APPROACH TO LEARNING AND DEVELOPMENT IN UNIVERSITY SETTINGS

As noted in the introduction to this article, 50 years ago higher education leaders and professors came together to discuss teaching and learning in higher education. At the original seminar, and subsequent publication from this important meeting, the terms multi-, inter-, and *trans*-disciplinarity were coined and debated (Apostel, 1972). Piaget argued for the need to situate the discussion in the context of epistemological views of knowledge. Building off of Piaget's structural approach, Jantsch (1972) argued for a view of knowledge more strongly linked to practice: "A systems approach (that)... would consider education and its motivation, above all, as . . . a purposeful human activity" (p. 99). The seminar and subsequent publication tied the problems and necessary solutions to stronger examination of institutional structures and reorganizations to simulate further work in this area. While there is no doubt that issues of interdisciplinarity and transdisciplinarity have led to a significant amount of discussion, the question raised here is whether it has impacted work on transdisciplinary approaches to learning and development in university settings.

The research topic guiding the papers in this issue of *Frontiers* starts from the assumption that in the area of research on learning and teaching, there has been significant effort to engage in interdisciplinary and transdisciplinary research in the fields of education, psychology, and learning sciences. Nevertheless, it is argued that an obstacle has been in breaking out of those silos to integrate those findings. Building on this claim, this article has brought together three perspectives never before considered side by side. The findings of each area are not new, but a careful examination highlights important insights when linking distinct areas of thought. The lens from the learning and developmental sciences, when placed next to higher education practitioners bring unique vantage points to our understanding of student learning, the role of teaching, and their interaction within larger university systems of which they are a part.

The opportunity of piecing together the distinct areas is precisely what Piaget (1972) described, namely the attempt to understand the development of knowledge within a systems framework, with each part contributing a level of analysis. Piaget's (1972) abstract view of knowledge did not take up on the situatedness of learning and the important role real world problem solving plays for college students. Theory and research from the learning and developmental sciences offer fresh perspective. Across all three areas (learning sciences, developmental sciences, and higher education), there were important areas of agreement in discussions of learning and teaching in university settings. All groups ideally take a constructivist approach to learning and development, sharing the belief that students must be viewed as agents of their learning and development. Furthermore, there is agreement across all three groups that more experienced others should focus less on teaching and disseminating set knowledge and practices to newcomers, and rather aim to be guides helping scaffold student engagement. Furthermore, all three groups recognize the importance of larger ecosystems in learning and teaching, with learning science research focusing on classroom design and disciplinary guidelines, developmental sciences examining student background, interaction with peers and others, and higher education focusing on disciplinary and university contexts and broader quality goals of modern higher education to assure citizenry and workforce readiness. We conclude by suggesting that if the views of students as learners, the importance of considering students' civic and professional identity formation, and university and disciplinary learning outcomes by national, transnational and disciplinary groups are aligned this sort of systems approach could mitigate some of the challenges that have been discussed. Work to date has shown that systems of support at the disciplinary, campus, national, and transnational levels and especially learning communities assist in bringing about necessary undergraduate reforms.

While there was shared agreement across the different areas, each area studies learning and teaching with different amounts of focus. Furthermore, the joint consideration of all three areas provides added insights into future avenues of work. For instance, the learning science focus on deep learning in general (and inquiry, integration, and metacognition in particular), illustrates how delineating what it means to know impacts the design of learning environments in university settings to support student learning. Learning scientists also have revealed the guiding role of teachers and others in creating particular kinds of contexts where learning happens. From the work of developmental scientists, we have a much deeper appreciation of what individual learners bring with them to the classroom at distinct periods of the lifespan. Particularly salient for issues of learning and teaching in university settings is the understanding of the ways cognitive and social development interact with learning and teaching. For example, developmental science has highlighted adolescents' and emerging adults' growing ability to entertain the multiple perspectives often presupposed by university teachers, and their nascent abilities to self-author their developmental trajectories influencing and influenced by their interest in making society a better place. The higher education work has highlighted how important consideration of the university eco-system and its ties to national and transnational efforts is to progress in learning and teaching in university settings. This work highlights structural and other aspects of learning and teaching, such as the extensive focus on disciplines, the power of disciplines and autonomy of individual faculty to teach as they see fit, and the lack of faculty training in teaching and learning as examples. Ongoing learning communities are needed to cultivate and support faculty efforts.

A systems approach, pulling together these disparate levels of analysis from the learning sciences, developmental sciences, and higher education, provides a powerful way forward and work against neoliberal fragmentation. **Figure 1** depicts the nested relationship between these different dimensions and their nested connections.

This work though is not without its challenges, challenges that were articulated 50 years ago, and which serve the neoliberal university well. Nevertheless, these issues can be freshly addressed by the transdisciplinary areas described in this article. **Table 1** provides an overview of the goals of reform efforts, main attributes, opportunities, and challenges of each of these fields.

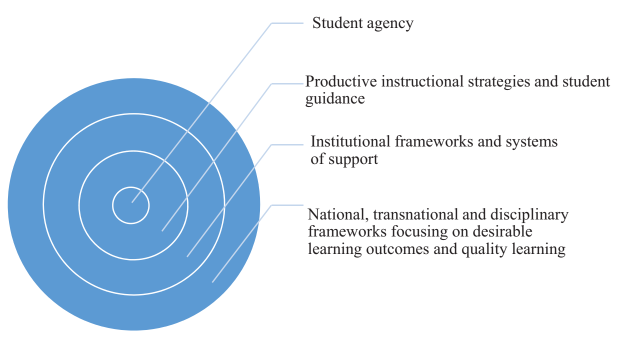

FIGURE 1 | A transdisciplinary approach to student learning.

TABLE 1 | Fields, main attributes, opportunities, and challenges.

Field	Main attributes	Opportunities	Challenges
Learning sciences	Interaction between students as inquirers and productive instructional strategies for deep learning	Extended knowledge about how students learn Emphasis on lifelong learning as a way to approach the complex and changing world and learners' role in problem-setting and problem-solving Centers for excellence in teaching and learning	Most faculty focus on "teaching" and lack explicit training in student learning and productive teaching practices Primary and secondary education has not set all students up equally for higher education learning
Developmental sciences	Recognition of the connection between students' construction of knowledge and development of agency and identity formation Recognition of emerging adults searching for meaning and purpose in work and civic life	Extended knowledge about emerging adulthood Importance of holistic approach to development Recognition of the importance of viewing learning as a process and deeply connected to identity formation	Instructors lack explicit knowledge about student development Academic support services have such background knowledge on student development but universities often lack mechanisms to coordinate academic units with student facing support services at the university
Higher education	Leadership matters: disciplinary societies, national and transnational organizations, and individual universities set agenda, frameworks, and outcomes for student learning	University leaders and academic organizations explicitly address these matters rather than leaving them to individual faculty preferences and styles Alignment across disciplinary, university, national and transnational goals New thinking about cultures of learning and role of students, faculty and staff in that work	Learning as a public good is often in tension with "arms race" approach to university rankings, which often focuses on grants and research Reward structures Organizational design based on silos, and university procedures and policies are not well aligned with modern understanding of student learning and development

Purpose: A transdisciplinary approach to learning aims to enhance students' ability to be lifelong learners, creating systems of support to bring about curricular reform.

The most major challenge is how to work with the training and reorganization of universities to allow integration across these levels of analysis to happen. At the undergraduate level, significant work is already underway to rethink curricular structures (e.g., the work reviewed above to create overarching learning outcomes associated with an undergraduate education). More directly linking this work to what is currently known from the learning and developmental sciences would provide fresh answers absent in the work identifying this problem 50 years ago. A transdisciplinary approach would also require significant changes to doctoral training, assuring that the next generation of faculty receive training in modern day learning and developmental science and are prepared for their roles as teachers. In addition, more work is needed to examine current university structures and rewards based on a transdisciplinary approach to learning and teaching to be sure our institutions are ready to support optimal learning and value excellence and success in student learning. Enhancement of opportunities for students and faculty to work collaboratively with other units on campus, and for universities to build partnerships beyond the campus gates have been highlighted as important as well. While these challenges are significant, the progress made in the last decades in the learning and developmental sciences show promise for new answers to questions that have been identified and stubbornly resistant to change. This issue of *Frontiers* symbolically represents one extremely important change necessary to move this dialogue forward.

It seems clear that higher education is moving closer to a vision of higher education that entails a common agenda- one that values broadening access, considers quality enhancements, and views higher education as a public good. More difficult has been figuring out how best to implement this common agenda and we have seen different approaches in the United States and within European countries. There are some commonalities suggesting best practices for sustained change. For instance, individual, institutional, and regional efforts have worked best when implementation involves active participation of individuals who not only adopt but adapt a broad agenda in contextually relevant ways. Consistent support rather than sanctions have aided implementation efforts, whether the sort of learning communities formed through the AAC&U learning institutes or collectives of institutions working on common problems, or the networks of support formed in both Europe and the United States as communities of learners similarly work together to construct resources and guidance. A close examination of these efforts shows that macro-level implementation efforts at the national or transnational level are working better than the microlevel change taking place on individual campuses (Sabatier, 2005; Pálvölgyi, 2017; Budwig and Jessen-Marshall, 2018). This highlights the importance of reviewing local efforts, and whether the support systems and guidance are in place to promote this work. Interestingly, this theory of higher education change involves precisely the sort of principles and frameworks advocated for student learning and that was suggested by Piaget and Janatsch, namely a systems approach that considers education and its motivation in transdisciplinary ways as a purposeful human activity, not only for students but for faculty, staff, and administrators guiding that change. The next logical step would entail a workshop like that 50 years ago with explicit discussion of the benefits of the learning and developmental sciences, alongside what we are learning from higher education reform efforts.

The work ahead is complex but can be justified in that it is exceedingly important at this juncture. Education holds the possibility to positively change society in transformative ways. As noted 50 years ago by Piaget (1972, p. 103): "If education is accepted as being essentially education for the self-renewal of society, it becomes an important, or even the most important agent of innovation." Viewing education as a public good, and innovation as intricately tied to education makes building a transdisciplinarity exemplar around the topic of teaching and learning at the university level a particularly important area of scholarship to work on. To enact the levels of analysis identified in this paper and address the challenges, it will be necessary to not only consider contributions from the learning sciences, developmental inquiry, and efforts of higher education leadership side by side, but also how best to align them in ways that assure not only a compelling vision, but also successful implementation.

AUTHOR CONTRIBUTIONS

NB contributed the original conception of the study, wrote the sections of the manuscript, and focused on learning sciences and higher education. The authors worked collaboratively on the section on developmental science with AA playing a major role in writing the section on emerging adulthood and self-authorship. Both authors contributed to manuscript revision and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Michael Bamberg for comments on a previous version of this manuscript. We also are grateful for the support of the Frances L. Hiatt School of Psychology and the Provost's Office at Clark University.

REFERENCES

AAC&U (2020). *What Liberal Education Looks Like: What it is, Who it's for, & Where it Happens*. Washington, DC: AAC&U.

Adecco (2019). *Skills Gap in the American Workforce*. Zürich: The Adecco Group.

Adelman, C. (2008). *Learning Accountability from Bologna: A Higher Education Policy Primer*. Washington, DC: Institute for Higher Education Policy.

Alfieri, L., Brooks, P. J., Aldrich, N. J., and Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *J. Educ. Psychol.* 103, 1–18. doi: 10.1037/a0021017

Al-Hendawi, M., and Albertine, S. (2019). "General education in the Middle East and North Africa: deep roots, contemporary context," in *Tradition Shaping Change: General Education in The Middle East and North Africa*, eds M. Al-Hendawi, A. Ahmed, and S. Albertine (Washington, DC: Association of American Colleges and Universities), 1–11.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., and Norman, M. K. (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco, CA: Jossey-Bass.

American Historical Association Tuning Project (2016). *AHA History Tuning Project: 2016 History Discipline Core*. Washington, DC: American Historical Association.

Amsel, E., and Smetana, J. G. (eds) (2011). *Adolescent Vulnerabilities and Opportunities: Developmental and Constructivist Perspectives*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139042819

Amsel, E. (2018). Scaffolding students' transformative experiences in higher education. *Paper Presented at Association of American Colleges and Universities Annual Meeting*, Washington, DC.

Apostel, L. (ed.) (1972). *Interdisciplinarity: Problems of Teaching and Research in Universities*. Paris: Organization for Economic Cooperation and Development.

Archer, W., and Davison, J. (2008). *Graduate Employability*. London: The Council for Industry and Higher Education.

Arnett, J. J. (2000). Emerging adulthood: a theory of development from the late teens through the twenties. *Am. Psychol.* 55, 469–480. doi: 10.1037/0003-066X.55.5.469

Arnett, J. J. (2016). College students as emerging adults: the developmental implications of the college context. *Emerg. Adulthood* 4, 219–222. doi: 10.1177/2167696815587422

Arum, R., and Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. Chicago, IL: University of Chicago Press.

Arum, R., and Roksa, J. (2014). *Aspiring Adults Adrift*. Chicago, IL: University of Chicago Press.

Baum, S., and McPherson, M. (2019). Improving teaching: strengthening the college learning experience. *Daedalus* 148, 5–13. doi: 10.1162/daed_e_01757

Baxter Magolda, M. (2001). *Making Their Own Way: Narratives for Transforming Higher Education to Promote Self-Development*. Sterling, VA: Stylus.

Baxter Magolda, M. (2008). Three elements of self-authorship. *J. Coll. Stud. Dev.* 49, 269–284. doi: 10.1353/csd.0.0016

Baxter Magolda, M., and Taylor, K. (2016). "Developing self-authorship in college to navigate emerging adulthood," in *The Oxford Handbook of Emerging Adults*, ed. J. J. Arnett (Oxford: Oxford University Press), 299–315.

Bielaczyc, K., Kapur, M., and Collins, A. (2013). "Cultivating a community of learners in k-12 classrooms," in *International Handbook of Collaborative Learning*, eds C. E. Hmelo-Silver, A. M. O'Donnell, C. Chan, and C. A. Chinn (London: Taylor and Francis, Inc), 233–249.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. Philadelphia, PA: David McKay Co Inc.

Bok, D. (2013). *Higher Education in America*. Princeton, NJ: Princeton University Press.

Bransford, J., Brown, A., and Cocking, R. (2000). *How People Learn: Brain, Mind, Experience, and School (Expanded Edition)*. Washington, DC: National Academy Press.

Brownell, S., and Tanner, K. (2017). Barriers to faculty pedagogic change: lack of training, time, incentives and tensions with professional identity. *CBE Life Sci. Educ.* 11, 339–346. doi: 10.1187/cbe.12-09-0163

Budwig, N. (2013). The learning sciences and liberal education. *Change Magaz. High. Learn.* 45, 40–48. doi: 10.1080/00091383.2013.787271

Budwig, N. (2015). Concepts and tools from the learning sciences for linking research, teaching and practice around sustainability issues. *Sustain. Sci.* 16, 99–104. doi: 10.1016/j.cosust.2015.08.003

Budwig, N., and Jessen-Marshall, A. (2018). Making the case for capstones and signature work. *Peer Rev.* 20, 4–7.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., and LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cogn. Sci.* 18, 439–477. doi: 10.1016/0364-0213(94)90016-7

Collins, A., and Kapur, M. (2014). "Cognitive apprenticeship," in *The Cambridge Handbook of the Learning Sciences*, 2nd Edn, ed. R. Keith Sawyer (Cambridge: Cambridge University Press), 109–127.

Crosnoe, R. (2011). "Schools, peers, and the big picture of adolescent development," in *Adolescent Vulnerabilities and Opportunities: Developmental and Constructivist Perspectives*, eds E. Amsel and J. Smetana (Cambridge: Cambridge University Press), 182–204. doi: 10.1017/cbo9781139042819.011

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., and Osher, D. (2019). Implications for educational practice of the science of learning and development. *Appl. Dev. Sci.* 24, 97–140. doi: 10.1080/10888691.2018.1537791

Davidson, C. (2017). *The New Education: How to Revolutionize the University to Prepare Students for a World in Flux*. New York, NY: Basic Books.

Davies, J., Hides, M. T., and Casey, S. (2001). Leadership in higher education. *Total Qual. Manage.* 12, 1025–1030. doi: 10.1080/09544120120096197

Elkana, Y., and Klopper, H. (2016). *The University in the 21st Century: Teaching the New Enlightenment at the Dawn of the Digital Age*. Budapest: Central European University Press.

Elkana, Y., Laubichler, M., and Wilkins, A. (2010). A call to reshape university curricula. *Nature* 467:788. doi: 10.1038/467788c

Eng, N. (2017). *Teaching College*. New York, NY: CUNY Academic Works.

Erikson, E. H. (1963). *Childhood and Society*, 2nd Edn. New York, NY: Norton.

Ferren, A., and Paris, D. (2015). *Faculty Leadership for Integrative Liberal Learning*. Washington, DC: AAC&U.

Fishman, B. J., Davis, E. A., and Chan, C. K. K. (2014). "A learning sciences perspective on teacher learning research," in *The Cambridge Handbook of the Learning Sciences*, 2nd Edn, ed. R. Keith Sawyer (Cambridge: Cambridge University Press), 707–725. doi: 10.1017/cbo9781139519526.042

Flavell, J. H., Miller, P. H., and Miller, S. A. (1993). *Cognitive Development*, 3rd Edn. Upper Saddle River, NJ: Prentice Hall.

Gagné, F. (2005). From noncompetence to exceptional talent: exploring the range of academic achievement within and between grade levels. *Gift. Child Q.* 49, 139–153. doi: 10.1177/001698620504900204

Godwin, K. A. (2015). The worldwide emergence of liberal education. *Int. High. Educ.* 79, 2–4. doi: 10.6017/ihe.2015.79.5835

Graesser, A. C., Halpern, D. F., and Hakel, M. (2008). *25 Principles of Learning, Task Force on Lifelong Learning at Work and at Home*. Washington, DC: Association for Psychological Science.

Greeno, J. G., and Engeström, Y. (2014). "Learning in activity," in *The Cambridge Handbook of the Learning Sciences*, ed. R. K. Sawyer (Cambridge: Cambridge University Press), 128–147. doi: 10.1017/CBO9781139519526.009

Hodge, D. C., Baxter Magolda, M., and Haynes, C. (2009). Engaged learning: Enabling self-authorship and effective practice. *Liberal Education* 95, 16–23.

Hutchings, P., Huber, M. T., and Ciccone, A. (2011). *The Scholarship of Teaching and Learning Reconsidered: Institutional Integration and Impact*. San Francisco, CA: Jossey—Bass.

Jantsch, E. (1972). "Towards interdisciplinarity and transdisciplinarity in education and innovation," in *Interdisciplinarity: Problems of Teaching and Research in Universities*, ed. L. Apostel (Paris: Organization for Economic Cooperation and Development), 97–121.

Kamp, A. (2020). *Navigating the Landscape of Higher Engineering Education: Coping with Decades of Accelerating Change Ahead*. Delft: 4TU.

Kehm, B. (2010). Quality in European higher education: the influence of the Bologna process. *Change Magaz. High. Learn.* 42, 40–46. doi: 10.1080/00091381003704677

Klein, J. J., Grossenbacher-Mansuy, W., Bill, A., Scholz, R., and Welti, M. (eds) (2001). *Transdisciplinarity: Joint Problem Solving Among Science, Technology, and Society*. Basel: Birkhäuser Verlag.

Klemenèiè, M. (2019). "Higher education expansion and students," in *International Encyclopedia of Higher Education Systems and Institutions*, eds J. C. Shin and P. Teixeira (Berlin: Springer).

Kritt, D. (ed.) (2018). *Constructivist Education in an Age of Accountability*. London: Palgrave Press.

Kuhn, D., and Holman, A. (2011). "What are the cognitive skills adolescents need for life in the twenty-first century?," in *Adolescent Vulnerabilities and Opportunities: Developmental and Constructivist Perspectives*, eds E. Amsel and J. Smetana (Cambridge: Cambridge University Press), 65–85. doi: 10.1017/cbo9781139042819.006

Lerner, R. (2016). Complexity embraced and complexity reduced: a tale of two approaches to human development: commentary on witherington and lickliter. *Hum. Dev.* 59, 242–249. doi: 10.2307/26765142

Levine, A. (1980). *Why Innovation Fails*. Albany, NY: State University of New York Press.

Lu, J., Bridges, S., and Hmelo-Silver, C. E. (2014). "Problem-based learning," in *The Cambridge Handbook of the Learning Sciences*, 2nd Edn, ed. R. Keith Sawyer (Cambridge: Cambridge University Press), 298–318.

Mälkki, K., and Lindblom-Ylänne, S. (2012). From reflection to action? Barriers and bridges between higher education teachers' thoughts and actions. *Stud. High. Educ.* 33–50. doi: 10.1080/03075079.2010.492500

Mayer, R. E. (ed.) (2014). *The Cambridge Handbook of Multimedia Learning*, 2nd Edn. Cambridge: Cambridge University Press, doi: 10.1017/CBO9781139547369

McAdams, D. P. (2013). Life authorship: a psychological challenge for emerging adulthood, as illustrated in two notable case studies. *Emerg. Adulthood* 1, 151–158. doi: 10.1177/2167696813481774

McAdams, D. P. (2016). "Life authorship in emerging adulthood," in *The Oxford Handbook of Emerging Adults*, ed. J. J. Arnett (New York, NY: Oxford University Press), 438–448.

National Association of Colleges and Employers (2017). *Job Outlook 2018: College Hiring to Increase by 4 Percent*. Bethlehem, PA: National Association of Colleges and Employers.

Overton, W. F. (2015). "Processes, relations, and relational-developmental-systems," in *Handbook of Child Psychology and Developmental Science: Theory and Method*, eds W. Overton, P. C. M. Molenaar, and R. M. Lerner (Hoboken, NJ: John Wiley & Sons Inc), 9–62.

Palincsar, A., and Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cogn. Instruct.* 1, 117–175. doi: 10.1207/s1532690xci0102_1

Pálvölgyi, K. (2017). Implementation through innovation: a literature-based analysis of the tuning project. *High. Learn. Res. Commun.* 7. doi: 10.18870/hlrc.v7i2.380

Peden, W. (2015). Signature work: a survey of current practices. *Lib. Educ.* 101, 22–29.

Piaget, J. (1972). "The epistemology of interdisciplinary relationships," in *Interdisciplinarity: Problems of Teaching and Research in Universities*, ed. L. Apostel (Paris: Organization for Economic Co-operation and Development), 127–139.

Piaget, J. (1978). *Success and Understanding*. Cambridge, MA: Harvard University Press.

Postareff, L., Lindblom-Ylänne, S., and Nevgi, A. (2007). The effect of pedagogical training on teaching in higher education. *Teach. Teach. Educ.* 23, 557–571. doi: 10.1016/j.tate.2006.11.013

Raeff, C. (2016). *Exploring the Dynamics of Human Development: An Integrative Approach*. Oxford: Oxford University Press.

Reichert, S., and Tauch, C. (2005). *Trends IV: European Universities Implementing Bologna*. Brussels: European University Association.

Sabatier, P. (2005). "From policy implementation to policy change: A personal odyssey," in *Reform and Change in Higher Education: Analysing Policy Implementation*, eds A. Gornitzka, M. Kogan, and A. Amaral (Dordrecht: Springer), 17–35. doi: 10.1007/1-4020-3411-3

Sawyer, R. K. (ed.) (2006). *The Cambridge Handbook of the Learning Sciences*. Cambridge: Cambridge University Press.

Schneider, C. G. (2008). *Liberal Education takes a new turn", NEA 2008 Almanac of Higher Education*. Washington, DC: The National. Education Association, 29–40.

Schneider, C. G. (2015). The LEAP challenge: transforming students, essential for liberal education. *Lib. Educ.* 101, 6–15.

Schneider, C. G. (2018). From cloistered to connected: the practical turn in American liberal education. *Change Magaz. High. Learn.* 50, 46–51. doi: 10.1080/00091383.2018.1509589

Schraw, G., and Moshman, D. (1995). Metacognitive theories. *Educ. Psychol. Rev.* 7, 351–371. doi: 10.1007/BF02212307

Schwartz, S. J., Zamboanga, B. L., Luyckx, K., Meca, A., and Ritchie, R. (2016). "Identity in emerging adulthood: reviewing the field and looking forward," in *The Oxford Handbook of Emerging Adulthood*, ed. J. J. Arnett (Oxford: Oxford University Press), 401–420.

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educ. Res.* 15, 4–14. doi: 10.3102/0013189x015002004

Tanner, J. L., Arnett, J. J., and Leis, J. A. (2009). "Emerging adulthood: learning and development during the first stage of adulthood," in *Handbook of Research on Adult Learning and Development*, eds M. C. Smith and N. DeFrates-Densch (Abingdon: Routledge), 34–67.

Taylor, A. (2017). Perspectives on the university as a business: the corporate management structure, neoliberalism and higher education. *J. Crit. Educ. Policy Stud.* 15, 108–135.

von Bertalanffy, L. (1972). The history and status of general systems theory. *Acad. Manage. J.* 15, 407–426. doi: 10.5465/255139

Vosniadou, S. (2013). "Conceptual change in learning and instruction: the framework theory approach," in *The International Handbook of Conceptual Change*, 2nd Edn, ed. S. Vosniadou (Abingdon: Routledge), 11–30.

Vosniadou, S. (2019). The development of students' understanding of science. *Front. Educ.* 4:32. doi: 10.3389/feduc.2019.00032

Walker, A., and Leary, H. (2009). A problem based learning meta analysis: differences across problem types, implementation types, disciplines, and assessment levels. *Interdiscip. J. Probl. Based Learn.* 3, 12–43. doi: 10.7771/1541-5015.1061

Wertsch, J. V., McNamee, G. D., McLane, J. B., and Budwig, N. A. (1980). The adult–child dyad as a problem-solving system. *Child Dev.* 51, 1215–1221. doi: 10.2307/1129563

Wieman, C. E. (2017). *Improving How Universities Teach Science*. Cambridge, MA: Harvard University Press.

Wieman, C. E. (2019). Expertise in university teaching & the implications for teaching effectiveness, evaluation & training. *Daedalus* 148, 47–78. doi: 10.1162/daed_a_01760

Winne, P. H., and Azevedo, R. (2014). "Metacognition," in *The Cambridge Handbook of the Learning Sciences*, 2nd Edn, ed. R. Keith Sawyer (Cambridge: Cambridge University Press), 63–87.

Witherington, D., and Boom, J. (2019). Conceptualizing the dynamics of development in the 21st century: process, (inter)action, and complexity. *Hum. Dev.* 63, 147–152. doi: 10.1159/000504097

Witherington, D., and Lickliter, R. (2016). Integrating development and evolution in psychological science: evolutionary developmental psychology, developmental systems, and explanatory pluralism. *Hum. Dev.* 59, 200–234. doi: 10.2307/26765140

Wright, G. B. (2011). Student-centered learning in higher education. *Int. J. Teach. Learn. High. Educ.* 23, 92–97.

Zahavi, H., and Friedman, Y. (2019). The Bologna process: an international higher education regime. *Eur. J. High. Educ.* 9, 23–39. doi: 10.1080/21568235.2018.1561314

Zimmerman, B. (2000). "Attaining self-regulation: A social cognitive perspective," in *Handbook of Self-Regulation*, eds M. Boekaerts, P. Pintrich, and M. Zeidner (San Diego, CA: Academic Press), 13–39. doi: 10.1016/b978-012109890-2/50031-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 20 October 2020
doi: 10.3389/fpsyg.2020.562665

Check for updates

Diagnostic Activities and Diagnostic Practices in Medical Education and Teacher Education: An Interdisciplinary Comparison

Elisabeth Bauer[1], Frank Fischer[1], Jan Kiesewetter[2], David Williamson Shaffer[3], Martin R. Fischer[2], Jan M. Zottmann[2] and Michael Sailer[1]*

[1] Education and Educational Psychology, Department Psychology, LMU University of Munich, Munich, Germany, [2] Institute for Medical Education, University Hospital, LMU University of Munich, Munich, Germany, [3] Epistemic Analytics Lab, Department of Educational Psychology, University of Wisconsin Madison, Madison, WI, United States

OPEN ACCESS

Edited by:
Bernhard Ertl,
Munich University of the Federal Armed Forces, Germany

Reviewed by:
Tom Rosman,
Leibniz Institute for Psychology Information and Documentation (ZPID), Germany
Mohamed Taha Mohamed,
British University in Egypt, Egypt

****Correspondence:***
Elisabeth Bauer
elisabeth.bauer@psy.lmu.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *15 May 2020*
Accepted: *23 September 2020*
Published: *20 October 2020*

Citation:
Bauer E, Fischer F, Kiesewetter J, Shaffer DW, Fischer MR, Zottmann JM and Sailer M (2020) Diagnostic Activities and Diagnostic Practices in Medical Education and Teacher Education: An Interdisciplinary Comparison. Front. Psychol. 11:562665. doi: 10.3389/fpsyg.2020.562665

In this article, we investigate diagnostic activities and diagnostic practices in medical education and teacher education. Previous studies have tended to focus on comparing knowledge between disciplines, but such an approach is complicated due to the content specificity of knowledge. We compared 142 learners from medical education and 122 learners from teacher education who were asked to (a) diagnose eight simulated cases from their respective discipline in a simulation-based learning environment and (b) write a justificatory report for each simulated case. We coded all justificatory reports regarding four diagnostic activities: *generating hypotheses*, *generating evidence*, *evaluating evidence*, and *drawing conclusions.* Moreover, using the method of Epistemic Network Analysis, we operationalized diagnostic practices as the relative frequencies of co-occurring diagnostic activities. We found significant differences between learners from medical education and teacher education with respect to both their diagnostic activities and diagnostic practices. Learners from medical education put relatively more emphasis on generating hypotheses and drawing conclusions, therefore applying a more hypothesis-driven approach. By contrast, learners in teacher education had a stronger focus on generating and evaluating evidence, indicating a more data-driven approach. The results may be explained by different epistemic ideals and standards taught in higher education. Further research on the issue of epistemic ideals and standards in diagnosing is needed. Moreover, we recommend that educators think beyond individuals' knowledge and implement measures to systematically teach and increase the awareness of disciplinary standards.

Keywords: diagnostic activities, diagnostic practices, medical education, teacher education, interdisciplinary research

INTRODUCTION

Interdisciplinary research involves various challenges, for example, the comparability of specific variables. In this article, we refer to a framework of diagnostic activities (Fischer et al., 2014; Heitzmann et al., 2019) that was applied to compare learners' diagnostic assessments within two disciplines (i.e., medical education and teacher education). We aim to investigate diagnostic activities in these disciplines and explore their conceptual integration into diagnostic practices. Hereby, we also seek to facilitate future interdisciplinary research on diagnostic practices and the learning of diagnostic activities.

Facilitating diagnostic skills in higher education is an important objective in many disciplines (e.g., Chernikova et al., 2020). This is certainly the case in medical education, which focuses on training future physicians in the assessment of patient symptomology. Similarly, future teachers' professional challenges include diagnosing students' performance, progress, learning difficulties such as behavioral and learning disorders, or other learning prerequisites (Reinke et al., 2011). Independent of the discipline, we broadly define diagnosing as "a process of goal-oriented collection and integration of case-specific information to reduce uncertainty in order to make medical or educational decisions" (Heitzmann et al., 2019, p.4).

Professional knowledge is a crucial prerequisite for diagnosing (Blömeke et al., 2015). There are numerous models conceptualizing professional knowledge (e.g., Shulman, 1987; Kopp et al., 2009; Charlin et al., 2012), e.g., in terms of content like biological knowledge in medicine (Charlin et al., 2012) and pedagogical knowledge in teaching (Shulman, 1987). Research has even suggested that professional knowledge in diagnostic reasoning may not only be discipline-specific but case-specific, since abstract types of e.g., strategic knowledge (Kopp et al., 2009) do not seem to transfer well across cases (e.g., Wimmers et al., 2007; Schwartz and Elstein, 2009). A recently proposed interdisciplinary perspective on professional diagnostic knowledge integrated conceptualizations in medical education and teacher education into an interdisciplinary model with the two dimensions of content-related facets and abstract types of knowledge (Förtsch et al., 2018). The model acknowledges that the issue of content-specificity also affects abstractions like types of professional knowledge, and thus emphasizes limited comparability of professional diagnostic knowledge across disciplines.

We argue nonetheless that interdisciplinary research in diagnosing may still benefit from a more abstracted level of observation, namely: diagnostic practices. We build on the idea of epistemic practices, which are defined as "the specific ways members of a community propose, justify, evaluate, and legitimize knowledge claims within a disciplinary framework" (Kelly, 2008, p. 99). Epistemic practices involve community-specific or discipline-specific epistemic aims (e.g., that a claim is justified), epistemic ideals (standards and criteria to assess the achievement of aims, e.g., that the evidence supports the claim and disconfirms competing claims), and processes that are considered reliable (e.g., disconfirming competing claims; Duncan and Chinn, 2016). Transferring the idea of epistemic practices into the context of diagnosing, we define diagnostic practices as systematic approaches that are applied to collect and integrate case-specific information to reduce uncertainty, and to make and communicate informed and justifiable decisions in a professional situation (Kelly, 2008; Heitzmann et al., 2019). We assume that diagnostic practices within disciplines may involve specificities concerning their epistemic aims, ideals and processes (Duncan and Chinn, 2016), e.g., the standards for justifying a diagnosis. Therefore, comparing diagnostic practices across disciplines may improve our understanding and facilitate future research.

To conceptualize diagnostic practices across different disciplines, we refer to underlying diagnostic activities such as *generating hypotheses*, *generating evidence*, *evaluating evidence*, and *drawing conclusions* (Fischer et al., 2014; Heitzmann et al., 2019; see **Supplementary Material** section "Supplementary Illustration of the Framework of Diagnostic Activities" for further details). The activities framework has been investigated in different disciplines, e.g., social work education (Ghanem et al., 2018), teacher education (Csanadi et al., 2018), and medical education (Lenzer et al., 2017). We assume, that although concrete hypotheses, evidence, and conclusions are specific, the epistemic purpose of these diagnostic activities is conceptually transferable across disciplines (Hetmanek et al., 2018): Although different hypotheses are appropriate for different diagnostic cases, the activity of generating hypotheses holds the purpose of identifying potential explanations, which may require further investigation. Thus, in investigating diagnostic activities, the case-specific content may be less important compared to characteristics concerning the structure of cases (e.g., the form and amount of potentially available evidence).

As a starting point in investigating diagnostic practices, we can interpret and integrate disciplinary conceptualizations used in previous research in terms of diagnostic activities: In medical education, research has focused in particular on process characteristics of diagnostic reasoning (e.g., Coderre et al., 2003; Norman, 2005; Mamede and Schmidt, 2017). Several studies found that medical students conform to a diagnostic practice, which was characterized as hypothesis-driven approach: Students generated different hypotheses and evaluated evidence accordingly to draw conclusions about their initial hypotheses (e.g., Coderre et al., 2010; Kiesewetter et al., 2013). The hypothesis-driven approach reflects an epistemic ideal of differential diagnosing, which is considered a reliable process in medicine and is thereby taught in medical education (see Duncan and Chinn, 2016). However, research has also found that some medical students exhibit a data-driven approach instead, which focuses on generating and evaluating evidence without considering specific hypotheses or integrating evidence into conclusions (e.g., Gräsel and Mandl, 1993; Norman et al., 2007; Kiesewetter et al., 2013).

In teacher education, research has mostly conceptualized diagnostic practices in terms of professional vision (Goodwin, 1994). Two subcomponents of professional vision have been distinguished: noticing, which includes identifying problems and generating hypotheses, and reasoning, which comprises describing, explaining, and predicting (e.g., Seidel and Stürmer, 2014). *Describing* refers to reporting generated evidence. *Explaining* means evaluating evidence in reference to professional knowledge. Therefore, describing and explaining both focus on evidence and seldom involve generating hypotheses or drawing conclusions, both of which point to *predicting* consequences of observations. Research indicates that expert teachers integrate *describing*, *explaining*, and *predicting* into their diagnostic practice (Seidel and Prenzel, 2007).

However, *describing* seems to be a prevailing aspect, while the use of *predicting* is more variant (Stürmer et al., 2016).

Given that work surrounding diagnostic assessment has primarily emerged from the disciplines of medical education and teacher education, we aimed to compare and integrate these two theoretical approaches with respect to diagnostic activities and diagnostic practices. Specifically, we operationalized diagnostic practices as the co-occurrence of diagnostic activities, which we investigated via the use of Epistemic Network Analysis (ENA) (Shaffer, 2017). The research questions are as following:

RQ1: To what extent do learners' *diagnostic activities* differ between medical education and teacher education?

RQ2: To what extent do learners' *diagnostic practices* differ between medical education and teacher education?

METHOD

Participants

A total of 142 medical students and 122 pre-service teachers participated in two matched data collections. Medical students were in their 5th to 11th semester ($M = 8.15$; SD = 1.82). Their mean age was $M = 24.41$ (SD = 2.89). A total of 102 were women and 40 were men. Pre-service teachers were in their 1st to 13th semester ($M = 4.55$; SD = 3.40), were on average $M = 22.96$ years old (SD = 4.10), and were mostly women (106 women; 15 men; 1 non-binary). Since half of the sample in teacher education was in their 1st to 4th semester, we defined a subsample of students in teacher education in the 5th or a higher semester for additional subsample analyses (see **Supplementary Material** section "Supplementary Subsample Analyses").

Materials

We developed simulation-based learning environments for medical education and teacher education, using the authoring tool CASUS (Hege et al., 2017). Both learning environments included eight cases with a parallel structure: The cases began with an initial problem concerning a virtual patient or student. Next, learners could freely choose to access several informational sources in any sequence. Learners solved two tasks in each of the eight cases: First, they provided a diagnosis of the virtual patient or virtual student's problem; second, they had to write a justificatory report, after being prompted, to justify their diagnosis by indicating how they approached and processed the case information.

The medical education cases presented virtual patients with symptoms of fever and back pain. Medical students were asked to take over the role of a general practitioner. After reading the initial problem statement, where the patient revealed his or her reason for seeing a physician, learners accessed the patient's history and had the option to access the results of different examinations and tests, e.g., physical examination, laboratory, X-ray, ECG.

In the teacher education cases, we asked pre-service teachers to take over the role of a teacher who was encountering a student with some initial performance-related or behavioral problems that might even be clinically relevant, e.g., ADHD or dyslexia. We chose these topics because they are relevant for teachers and at the same time entail structural similarities to medical cases. After reading the initial problem, the learners could access informational sources such as reports of observations from inside and outside of the classroom as well as transcripts of conversations with the student, the parents, and other teachers. Moreover, participants could explore samples of the student's written exercises and school certificates.

For further details on the learning environment and the cases used, see **Supplementary Material** sections "Supplementary Case Materials for Medical Education" and "Supplementary Case Materials for Teacher Education."

Procedure

The data collection was computer-based and took place in a laboratory setting. We introduced participants to the aims, scope, and procedure of the study and familiarized them with the materials. Next, participants entered the simulation-based learning environment that was designed for their field of study. After giving informed consent to participate in the study, they had to answer a knowledge pretest that took up to 35 min. Afterward, they entered the learning phase, consisting of the eight simulated cases of their respective discipline. Time on task for all cases was $M = 45.1$ min (SD = 12.2) in medical education and $M = 51.8$ min (SD = 16.5) in teacher education. After four cases, participants took a break of 10 min before continuing with the second part of the learning phase and solving cases five to eight. Subsequently, they had to answer a knowledge posttest, which again took up to 35 min. Finally, participants received monetary compensation.

Data Sources and Instruments

For this paper, we analyzed only the text data from the justificatory reports that all learners wrote for the eight simulated cases. Participants wrote the justificatory reports in an empty text field, right after indicating their diagnosis for each case. There was no template or additional support apart from the standardized prompt to justify the diagnosis by indicating how they approached the case and how they processed the case information. The overall data set used in this paper consisted of 1,136 justificatory reports written by the 142 medical students (average number of words per report $M = 57.4$; SD = 32.6) and 976 justificatory reports written by the 122 pre-service teachers (average number of words per report $M = 89.6$; SD = 53.2).

Diagnostic Activities

We coded the two sets of justificatory reports on four diagnostic activities: *generating hypotheses*, *generating evidence*, *evaluating evidence*, and *drawing conclusions*. **Table 1** presents definitions and examples of the four codes. We developed a coding scheme applicable for medical education and teacher education. Coding and segmentation were done simultaneously to account for overlap in the activities as well. In both disciplines, the raters were first to second year doctoral students and student assistants (minimum 6th semester) from the respective fields. All raters were blind to this study's research questions. Raters did four

TABLE 1 | Definitions, examples, and inter-rater reliabilities (IRRs indicated as Krippendorff's α_U) for the four codes: *generating hypotheses*, *generating evidence*, *evaluating evidence*, and *drawing conclusions*.

		Medical education		Teacher education	
Code	**Definition**	**Example**	**IRR**	**Example**	**IRR**
Generating hypotheses	Explicit collection of different potential diagnoses or pointing to one diagnosis involving expressed insecurity, e.g., using conjunctive mood.	I believe this is a case of nerve entrapment.	0.60	The initial information makes me think of impaired vision, a reading disorder, or emotional problems as potential explanations for Annika's issues.	0.43
Generating evidence	Explicit description of accessing informational sources, e.g., tests, interviews, or observations.	Subsequently, I looked at the MRI and X-ray.	0.65	I observed Anna's school-related behavior and achievement.	0.56
Evaluating evidence	Explicit listing and/or interpretation of separate case information.	Among other results, the patient has an increased CRP and leukocytosis.	0.75	Markus behaves aggressively and gets offended very easily.	0.75
Drawing conclusions	Explicit conclusion or rejection of at least one diagnosis.	The patient clearly has tonsillitis involving a fever.	0.65	Consequently, I rejected the diagnosis of ADHD.	0.49

rounds of joined coding training, starting with 20 reports and increasing the number in every round of training. To evaluate inter-rater reliability (IRR), five raters in medical education and four in teacher education coded 150 reports for the respective project (13% of the data set in medical education; 15% in teacher education). The overall IRR for the simultaneous segmentation and coding was Krippendorff's $\alpha_U = 0.67$ in medical education and $\alpha_U = 0.65$ in teacher education (see **Table 1**), which we consider as satisfactory. For the analyses, we calculated the share of diagnostic activities within medical education and teacher education, respectively, as the percentages of the different diagnostic activities relative to the overall amount.

Diagnostic Practices

We operationalized diagnostic practices as the co-occurrences of diagnostic activities in the justificatory reports, using the method of ENA (Shaffer, 2017). The ENA algorithm analyzes co-occurring diagnostic activities within a moving window of two sentences (Siebert-Evenstone et al., 2017). Therefore, subsequent to the coding, we determined presence or absence of the four diagnostic activities per sentence. We accumulated the co-occurrences and created one network graph per discipline. In the network graphs, the colored edges refer to co-occurrences between diagnostic activities, with thickness indicating their relative frequencies. Relative frequencies of co-occurring activities allowed us to draw inferences about the general diagnostic practices of each discipline. Additionally, a comparison graph (i.e., showing only the difference between both graphs), allowed us to isolate the differences between the two disciplines' diagnostic practices.

We also centered the networks and created one centroid per learner as well as per discipline. The centroids' position is relative to the co-occurrences between diagnostic activities in the respective network. On the level of single learners, the representation of centroids can be used to depict the learners' distribution within the network space, which can be interpreted as an indicator of interindividual heterogeneity in diagnostic practices. On the level of disciplines, we can consider centroids as group means. ENA enables statistical testing of the group differences in overall diagnostic practices between learners in medical education and teacher education. To facilitate the testing of the group differences, we used the option of means rotation, which aligns the two disciplines' group means on the X-axis, thus depicting systematic variance on only one dimension.

Statistical Analyses

To address RQ1, the extent to which diagnostic activities differ between learners from medical education and teacher education, we calculated t tests for independent samples, one test per diagnostic activity, using Bonferroni-adjusted alpha levels of $\alpha = 0.0125$ per test ($\alpha = 0.05/4$). To statistically test RQ2, differences in diagnostic practices between learners from medical and teacher education, we used an independent-samples t test as well, comparing the two group means from the two disciplines' ENA networks at an alpha level of $\alpha = 0.05$. If Levene's test indicated unequal variances, we adjusted the degrees of freedom accordingly.

RESULTS

Comparing the two disciplines, there was a significant difference regarding the number of semesters studied (medical education $M = 8.15$; SD = 1.82; teacher education $M = 4.55$; SD = 3.40), $t(173) = 10.35$, $p < 0.001$, Cohen's $d = 2.75$. Therefore, we analyzed the relation with the percentages of diagnostic activities within the disciplines. There was no significant correlation found between number of semesters studied and the percentages of the different diagnostic activities (for details see **Supplementary Material** section "Supplementary Results of a Correlation Between Semesters Studied and Number of Diagnostic Activities"). However, to ensure that the number of semesters studied did not bias the results, we performed the following analyses not only with the full sample as reported in the following sections, but a second time, comparing learners from medical education to the specified subsample of learners from teacher education in their 5th or a higher semester (see **Supplementary Material** section "Supplementary Subsample Analyses").

Diagnostic Activities in Medical Education and Teacher Education (RQ1)

In both disciplines, *evaluating evidence* was clearly the most prominent activity found in the justificatory reports with a share of more than half of the diagnostic activities found in the reports (medical education $M = 60.96\%$; SD = 10.24%; teacher education $M = 66.08\%$; SD = 17.02%). The difference in the relative frequencies for *evaluating evidence* was significant with a small effect size [$t(192) = 2.91$, $p = 0.004$, Cohen's $d = 0.37$]. We found that in medical education, the share for *generating hypotheses* was about twice as high ($M = 16.26\%$; SD = 7.96%) as in teacher education ($M = 8.37\%$; SD = 6.41%). This difference was significant with a large effect size [$t(261) = 8.92$, $p < 0.001$, Cohen's $d = 1.08$]. By contrast, the share for *generating evidence* was about twice as high in teacher education ($M = 13.74\%$; SD = 14.81%) as in medical education ($M = 6.79\%$; SD = 8.26%), and this was also significantly different with a medium-sized effect [$t(183) = 4.60$, $p < 0.001$, Cohen's $d = 0.59$]. In medical education, we also found a significantly higher share for *drawing conclusions* ($M = 15.99\%$; SD = 6.39%) than in teacher education ($M = 11.82\%$; SD = 6.83%), with a medium effect size [$t(262) = 5.13$, $p < 0.001$, Cohen's $d = 0.63$].

Comparing medical education with the specified subsample from teacher education (see section "Participants"), the results show the same results pattern (for detailed results see **Supplementary Material** section "Supplementary Subsample Analyses"). However, there was no significant difference in the relative frequencies for *evaluating evidence* [medical education $M = 60.96\%$; SD = 10.24%; teacher education $M = 65.40\%$; SD = 18.00%; $t(77) = 1.81$, $p = 0.075$, Cohen's $d = 0.34$].

Diagnostic Practices in Medical Education and Teacher Education (RQ2)

In **Figure 1**, we present the diagnostic practices of learners from medical education (**Figure 1A**) and teacher education (**Figure 1C**) as network graphs. The colored edges and their thickness reflect the relative frequencies of co-occurrences of diagnostic activities. The overall network across all learners from medical education (**Figure 1A**) showed some similarities to the overall network across all learners from teacher education (**Figure 1C**): First, in both disciplines, we found that the relative frequencies of co-occurrences were in accordance with the relative frequencies of the individual diagnostic activities (see the results for RQ1). In both network graphs, the three relatively most frequent co-occurrences were the ones including *evaluating evidence*. This is why we found *evaluating evidence* near the center of the disciplines' overall networks. However, by looking at its temporal context indicated by co-occurrences with other diagnostic activities, we can draw inferences about the purpose of *evaluating evidence* within the respective context. When it co-occurs with *drawing conclusions* or *generating hypotheses*, *evaluating evidence* serves the purpose of *explaining*; whereas when co-occurring with *generating evidence*, *evaluating evidence* may rather *describe* the evidence (see **Table 2** for examples). To compare learners from medical education and teacher education, the comparison graph (**Figure 1B**) shows the difference between the two disciplines' overall networks, therefore indicating only the differences in co-occurrences. In medical education, there was a relatively higher frequency of *evaluating evidence* co-occurring with *generating hypotheses*, pointing to a rather hypothesis-driven approach that puts more emphasis on *explaining* evidence; whereas learners in teacher education exhibited a relatively higher frequency of co-occurrences between *evaluating evidence* and *generating evidence*, indicating a tendency toward *describing* evidence or a data-driven approach.

In addition to the disciplines' overall networks, **Figure 2** presents the distribution of single learners across the two disciplines' overall networks. The colored points represent the networks' centroids on the level of single learners from medical education (**Figure 2A**) and teacher education (**Figure 2C**). In teacher education, single learners' centroids (red colored points) are more scattered across the network space, compared to the positioning of the single learners' centroids in medical education (blue colored points). This indicates that the diagnostic practices

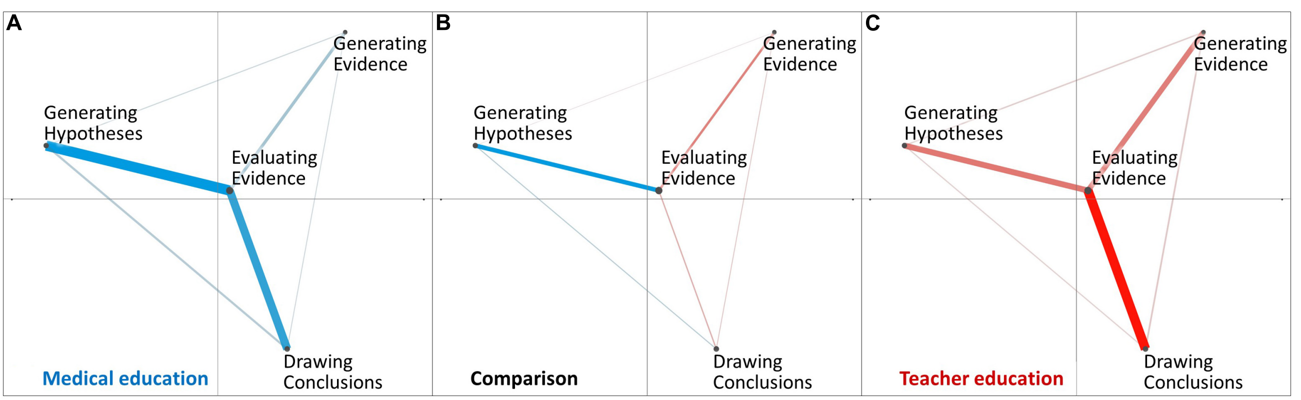

FIGURE 1 | ENA networks from medical education **(A)**, and teacher education **(C)**. The comparison network **(B)** depicts only the differences between the other two networks.

TABLE 2 | Examples of *evaluating evidence*, co-occurring with *generating evidence*, *generating hypotheses*, or *drawing conclusions* in a temporal context of one to two sentences in the disciplines of medical education and teacher education.

Case	Text	Generating hypotheses	Generating evidence	Evaluative evidence	Drawing conclusions
Section a: Examples of *evaluating evidence* co-occurring with *drawing conclusions* or *generating hypotheses* in the discipline of medical education					
2	Due to his age and the sudden symptomatology in only his lumbar spine, I would diagnose a rheumatic disease.	0	0	1	1
7	Upon physical examination, she mostly indicated pain in the upper abdomen, which highlights the region of the liver, gall bladder, and eventually the biliary tract and pancreatic duct.	0	0	1	0
	Laboratory results indicated increased liver values, which is why I believe the patient has hepatitis.	1	0	1	0
Section b: Examples of *evaluating evidence* co-occurring with *drawing conclusions* or *generating hypotheses* in the discipline of teacher education					
8	The characteristic writing, confusion of characters, deficits in stringing together syllables, as well as deficits in syllabification and slow reading speed, combined with an otherwise good school performance, clearly indicate dyslexia.	0	0	1	1
6	Thomas might have eventually developed ADHD and therefore low concentration.	1	0	0	0
	This assumption is backed by the fact that his performance in all subjects decreased and that he does not fully answer all questions on exams.	0	0	1	0
Section c: Examples of *evaluating evidence* co-occurring with *generating evidence* in the discipline of medical education					
7	First, I examined all the available information, before focusing on the most relevant points.	0	1	0	0
	They mostly seemed to be related to the liver.	0	0	1	0
8	Even after being treated by the general practitioner, the patient still had a fever and symptoms of a systemic infection.	0	0	1	0
	This is why, considering the anamnesis regarding previous travels, I decided to administer an HIV test.	0	1	1	0
Section d: Examples of *evaluating evidence* co-occurring with *generating evidence* in the discipline of teacher education					
6	I examined the teacher's report and the available documents.	0	1	0	0
	It seems that Thomas' symptoms have only been observable recently and that he has repeatedly complained about small font sizes.	0	0	1	0
5	Initially, I collected information from observations, conversations, the annual report, and recent school exams.	0	1	0	0
2	My attention was caught by the mother's description of her reading behavior at home, especially in terms of reading aloud.	0	0	1	0

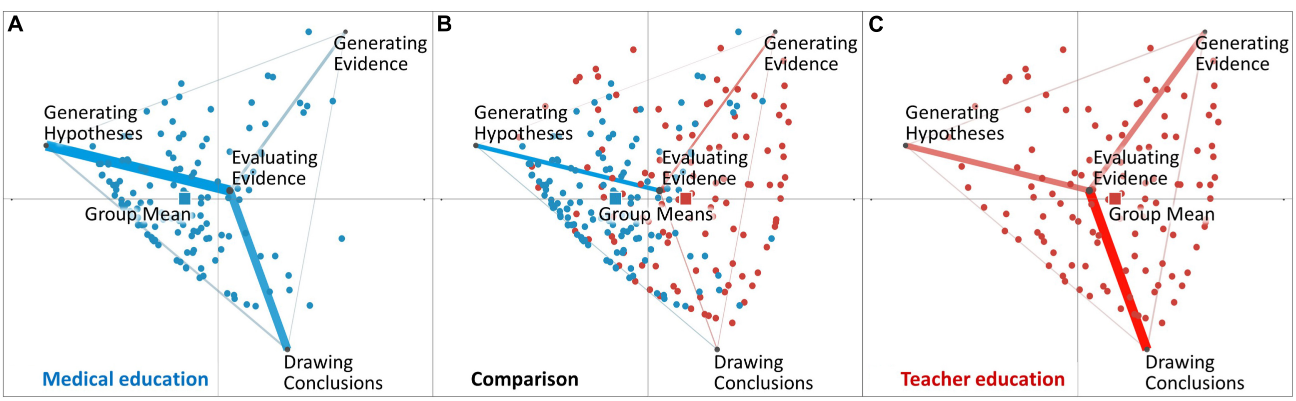

FIGURE 2 | Distributions of learners within medical education **(A)**, and teacher education **(C)**. The figures also contain group means (squares) across the learners within the two disciplines. The comparison graph **(B)** depicts both distributions and the differences between the other two networks.

of learners from medical education are more homogeneous compared with the diagnostic practices of learners from teacher education.

Figure 2 presents centroids on the group level, representing the means of all learners within the two disciplines of medical education and teacher education as indicated by the colored squares. The positioning of the group mean of learners from medical education ($M = -0.36$, SD = 0.63, $N = 142$) was statistically significantly different from the positioning of the group mean of learners from teacher education [$M = 0.42$, SD = 0.74, $N = 122$; $t(240.48) = -9.16$, $p < 0.01$, Cohen's $d = 1.14$]. This result indicates a significant difference in diagnostic practices between teacher education and medical education. Repeating these analyses, comparing students from medical education with the specified subsample from teacher education, revealed basically the same result (for details see **Supplementary Material** section "Supplementary Subsample Analyses").

DISCUSSION

In analyzing learners' reports of their diagnostic activities in medical education and teacher education, we found that future physicians and future teachers put the most focus toward *evaluating evidence*. Moreover, learners from teacher education focused more on *generating evidence*, whereas learners from medical education put more focus toward *generating hypotheses* and *drawing conclusions*. These results support the notion that the relative emphasis on each diagnostic activity differs between these disciplines.

The disciplinary differences in the use of diagnostic activities is also reflected by overall diagnostic practices. Because the overall network across all learners from medical education was similar to the network across all learners from teacher education, this similarity suggests that the overall diagnostic practices are similar. Still, there were significant disciplinary differences in the relative frequencies of the co-occurrences of diagnostic activities. In general, we found that learners from medical education showed a more explanation-driven or hypothesis-driven approach (see Coderre et al., 2010; Kiesewetter et al., 2013; Seidel and Stürmer, 2014), whereas learners from teacher education showed a more description-driven or data-driven approach (see Gräsel and Mandl, 1993; Norman et al., 2007; Kiesewetter et al., 2013; Seidel and Stürmer, 2014). Furthermore, learners from teacher education showed greater variability in their diagnostic practices than learners from medical education.

We interpret the results relating to epistemic ideals as the "criteria or standards used to evaluate epistemic products" (Duncan and Chinn, 2016, p. 158). In the context of medical education, differential diagnosing is considered as ideal for ensuring a reliable process. Differential diagnosing essentially refers to a hypothesis-driven approach of generating and testing hypotheses (see Fischer et al., 2014), which is what we observed in learners from medical education. This diagnostic standard is put into practice on different levels (e.g., in guidelines and university curricula), and is systematically taught to future physicians in their medical programs. In teacher education, we are not aware of a widespread use of such specific standards for diagnosing in general and particularly regarding the topic of students' behavioral and performance-related disorders. Research in teacher education was referred to as a rather "young" field (Grossman and McDonald, 2008) and thus, the evolvement of standards for diagnosing might be less advanced than in medical education. In comparison with medical students, pre-service teachers also seem to show greater variability in their diagnostic practices, which may support the notion of lower standardization in diagnostic practices or at least in educating pre-service teachers to apply diagnostic practices. However, there might be some implicit ideals that enhance pre-service teachers' tendency to embrace a data-driven approach in their diagnostic practices. First, as a reaction to findings of teachers' biases in diagnostic tasks (e.g., Südkamp et al., 2012), some teacher education programs have subsequently taught the concept of professional vision (Goodwin, 1994) to pre-service teachers, emphasizing the need to focus on *describing* observations before *explaining* them (e.g., Seidel and Stürmer, 2014). This development may complement other implicit values (see Duncan and Chinn, 2016) in teaching, such as to avoid being judgmental toward students (Aalberts et al., 2012). Therefore, the findings may reflect disciplinary differences in epistemic ideals implemented in higher education and diagnostic practices, respectively.

Limitations

One limitation of the study involves the inter-rater reliabilities for *generating hypotheses* and *drawing conclusions*, which were relatively low in the teacher education data. This could limit the conclusions that can be drawn about the variability in diagnostic practices of teacher education learners in particular.

Another limitation may be the learners' study progress: In the full sample, learners from medical education had completed significantly more semesters than learners from teacher education. However, the number of semesters did not correlate with the proportion of the different diagnostic activities. The subsample analyses, which compared students from medical education with students from teacher education in their 5th or a higher semester revealed the same patterns of results as the analyses of the full sample. Hence, it seems unlikely that the *a priori* difference in the number of semesters would lead to substantial bias in our results.

Furthermore, we acknowledge that although we argue for the interdisciplinary comparability of the diagnostic activities' epistemic purpose, this conceptualization may still not fully eliminate the issues associated with comparing disciplinary diagnostic practices. Yet, we think that diagnostic activities and diagnostic practices are more advantageous in terms of interdisciplinary comparability than other investigated approaches, e.g., professional diagnostic knowledge.

The choice of clinical topics in both disciplines served the purpose of having similarly structured problems. Nevertheless, in teacher education there are other than clinical areas where diagnosing is relevant (e.g., assessing a student's level of skill). Thus, our choice might limit the generalizability of the findings to other areas of assessment in teacher education. However, if we consider diagnostic practices as discipline-specific approaches, it is reasonable to assume that the findings may replicate in other areas of teachers' diagnostic assessments, which could be investigated in further research.

Finally, similar to verbal protocols, assessing reported activities raises the question of validity, concerning the degree to which the reports effectively represent actually performed activities. Therefore, further research might additionally complement reported diagnostic activities with behavioral data like user-logs.

CONCLUSION

In this article, we have argued that interdisciplinary research on diagnostic assessments benefits from comparisons drawn at the level of diagnostic activities (Fischer et al., 2014) and diagnostic practices (Kelly, 2008; Heitzmann et al., 2019) as comparing professional diagnostic knowledge has been found to be difficult due to its content specificity. In an interdisciplinary comparison of justifications by learners from teacher education and medical education, we found significant differences in their diagnostic activities and diagnostic practices. We found a more hypothesis-driven approach in justifications of learners from medical education, who put relatively more emphasis on generating hypotheses and drawing conclusions. Learners from teacher education instead seemed to apply a more data-driven approach, with a stronger focus on generating and evaluating evidence. The results may allude to different epistemic ideals and diagnostic standards (see Duncan and Chinn, 2016) taught in higher education and thereby put into diagnostic practices.

Diagnostic activities can provide a useful and interdisciplinary framework to analyze diagnostic practices across disciplines. For future interdisciplinary research, we recommend considering matched study designs, as implemented in our project, to maximize interdisciplinary comparability. Additionally, from a practically oriented viewpoint, we recommend that educators from both the medical education and teacher education fields reflect further on their standards in diagnosing and their underlying epistemic ideals to further increase the awareness of practitioners and systematization in teaching. Finally, we encourage researchers to further investigate the potential relation between epistemic ideals and diagnostic practices in terms of interdisciplinary differences, commonalities, and their continuing evolvement.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors. Requests to access the data should be directed to elisabeth.bauer@psy.lmu.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Medical Faculty of LMU Munich (no. 17-249). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EB, FF, MF, JK, MS, and JZ developed the study concept and contributed to the study design. EB and MS performed the data analysis. EB, FF, and MS interpreted the data. EB drafted the manuscript. FF, MF, JK, MS, DS, and JZ provided critical revisions. All authors approved the final version of the manuscript for submission.

FUNDING

This research was funded by a grant of the German Federal Ministry of Research and Education (16DHL1040 und 16DHL1039) and the Elite Network of Bavaria (K-GS-2012-209). Parts of this work were funded by the National Science Foundation (DRL-1661036, DRL-1713110), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.562665/full#supplementary-material

REFERENCES

Aalberts, J., Koster, E., and Boschhuizen, R. (2012). From prejudice to reasonable judgement: integrating (moral) value discussions in university courses. *J. Moral Educ.* 41, 437–455. doi: 10.1080/03057240.2012.677600

Artelt, C., and Rausch, T. (2014). "Accuracy of teacher judgments: when and for what reasons?," in *Teachers' Professional Development*, eds S. Krolak-Schwerdt, S. Glock, and M. Böhmer (Leiden: Brill | Sense), 27–43. doi: 10.1007/978-94-6209-536-6_3

Blömeke, S., Gustafsson, J.-E., and Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Z. Psychol.* 223, 3–13. doi: 10.1027/2151-2604/a000194

Charlin, B., Lubarsky, S., Millette, B., Crevier, F., Audétat, M. C., Charbonneau, A., et al. (2012). Clinical reasoning processes: unravelling complexity through graphical representation. *Med. Educ.* 46, 454–463. doi: 10.1111/j.1365-2923.2012.04242.x

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., and Fischer, F. (2020). Facilitating diagnostic competences in higher education—a meta-analysis in medical and teacher education. *Educ. Psychol. Rev.* 32, 157–196. doi: 10.1007/s10648-019-09492-2

Coderre, S., Mandin, H., Harasym, P. H., and Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Med. Educ.* 37, 695–703. doi: 10.1046/j.1365-2923.2003.01577.x

Coderre, S., Wright, B., and McLaughlin, K. (2010). To think is good: querying an initial hypothesis reduces diagnostic error in medical students. *Acad. Med.* 85, 1125–1129. doi: 10.1097/acm.0b013e3181e1b229

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., and Fischer, F. (2018). When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *Int. J. Comput. Support. Collab. Learn.* 13, 419–438. doi: 10.1007/s11412-018-9292-z

Duncan, R. G., and Chinn, C. A. (2016). New directions for research on argumentation: insights from the AIR framework for epistemic cognition. *Z. Padagog. Psychol.* 30, 155–161. doi: 10.1024/1010-0652/a000178

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Front. Learn. Res.* 2, 28–45. doi: 10.14786/flr.v2i3.96

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., et al. (2018). Systematizing professional knowledge of medical doctors and teachers: development of an interdisciplinary framework in the context of diagnostic competences. *Educ. Sci.* 8:207. doi: 10.3390/educsci8040207

Ghanem, C., Kollar, I., Fischer, F., Lawson, T. R., and Pankofer, S. (2018). How do social work novices and experts solve professional problems? A micro-analysis of epistemic activities and the use of evidence. *Eur. J. Soc. Work.* 21, 3–19. doi: 10.1080/13691457.2016.1255931

Goodwin, C. (1994). Professional vision. *Am. Anthropol.* 96, 606–633.

Gräsel, C., and Mandl, H. (1993). Förderung des Erwerbs diagnostischer Strategien in fallbasierten Lernumgebungen. *Unterrichtswissenschaft* 21, 355–369.

Grossman, P., and McDonald, M. (2008). Back to the future: directions for research in teaching and teacher education. *Am. Educ. Res. J.* 45, 184–205. doi: 10.3102/0002831207312906

Hege, I., Kononowicz, A. A., and Adler, M. (2017). A clinical reasoning tool for virtual patients: design-based research study. *JMIR Med. Educ.* 3:e21. doi: 10.2196/mededu.8100

Heitzmann, N., Seidel, T., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., et al. (2019). Facilitating diagnostic competences in simulations in higher education. *Front. Learn. Res.* 7, 1–24. doi: 10.14786/flr.v7i4.384

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., et al. (2018). Teachers' assessment competence: integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teach. Teach. Educ.* 76, 181–193. doi: 10.1016/j.tate.2017.12.001

Hetmanek, A., Engelmann, K., Opitz, A., and Fischer, F. (2018). "Beyond intelligence and domain knowledge," in *Scientific Reasoning and Argumentation: The Roles of Domain-Specific and Domain-General Knowledge*, eds F. Fischer, C. A. Chinn, K. Engelmann, and J. Osborne (New York: Routledge). doi: 10.1016/j.engappai.2013.11.002

Kelly, G. (2008). "Inquiry, activity and epistemic practice," in *Teaching Scientific Inquiry*, eds R. A. Duschl and R. E. Grandy (Leiden: Brill | Sense), 99–117. doi: 10.1163/9789460911453_009

Kiesewetter, J., Ebersbach, R., Görlitz, A., Holzer, M., Fischer, M. R., and Schmidmaier, R. (2013). Cognitive problem solving patterns of medical students correlate with success in diagnostic case solutions. *PLoS One* 8:e71486. doi: 10.1371/journal.pone.0071486

Kopp, V., Stark, R., Kühne-Eversmann, L., and Fischer, M. R. (2009). Do worked examples foster medical students' diagnostic knowledge of hyperthyroidism? *Med. Educ.* 43, 1210–1217. doi: 10.1111/j.1365-2923.2009.03531.x

Lenzer, B., Ghanem, C., Weidenbusch, M., Fischer, M. R., and Zottmann, J. (2017). Scientific reasoning in medical education: a novel approach for the analysis of epistemic activities in clinical case discussions. *Paper Presented at the Conference of the Association for Medical Education in Europe (AMEE)*, Helsinki, Finland.

Mamede, S., and Schmidt, H. G. (2017). Reflection in medical diagnosis: a literature review. *Health Prof. Educ.* 3, 15–25. doi: 10.1016/j.hpe.2017.01.003

Norman, G. (2005). Research in clinical reasoning: past history and current trends. *Med. Educ.* 39, 418–427. doi: 10.1111/j.1365-2929.2005.02127.x

Norman, G., Young, M., and Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Med. Educ.* 41, 1140–1145.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., and Goel, N. (2011). Supporting children's mental health in schools: teacher perceptions of needs, roles, and barriers. *Sch. Psychol. Q.* 26, 1–13. doi: 10.1037/a0022714

Schwartz, A., and Elstein, A. S. (2009). "Clinical problem solving and diagnostic decision making: a selective review of the cognitive research literature," in *The Evidence Base of Clinical Diagnosis*, eds J. A. Knottnerus and F. Buntinx (Hoboken, NJ: Wiley), 237–255. doi: 10.1002/9781444300574.ch12

Seidel, T., and Prenzel, M. (2007). How teachers perceive lessons-assessing educational competencies by means of videos. *Zeitschrift für Erziehungswissenschaft* 10, 201–216.

Seidel, T., and Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *Am. Educ. Res. J.* 51, 739–771. doi: 10.3102/0002831214531321

Shaffer, D. W. (2017). *Quantitative Ethnography*. Madison: Cathcart Press.

Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harv. Educ. Rev.* 57, 1–23. doi: 10.17763/haer.57.1.j463w79r56455411

Siebert-Evenstone, A. L., Irgens, G. A., Collier, W., Swiecki, Z., Ruis, A. R., and Shaffer, D. W. (2017). In search of conversational grain size: modelling semantic structure using moving stanza windows. *J. Learn. Anal.* 4, 123–139.

Stürmer, K., Seidel, T., and Holzberger, D. (2016). Intra-individual differences in developing professional vision: preservice teachers' changes in the course of an innovative teacher education program. *Instr. Sci.* 44, 293–309. doi: 10.1007/s11251-016-9373-1

Südkamp, A., Kaiser, J., and Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: a meta-analysis. *J. Educ. Psychol.* 104:743. doi: 10.1037/a0027627

Wimmers, P. F., Splinter, T. A., Hancock, G. R., and Schmidt, H. G. (2007). Clinical competence: general ability or case-specific? *Adv. Health Sci. Educ.* 12, 299–314. doi: 10.1007/s10459-006-9002-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

ORIGINAL RESEARCH
published: 08 December 2020
doi: 10.3389/feduc.2020.544194

Research in Education Draws Widely From the Social Sciences and Humanities

Kristine Lund[1,2]*, *Heisawn Jeong*[3], *Sebastian Grauwin*[2,4] *and Pablo Jensen*[2,4]

[1] Laboratoire ICAR, ENS de Lyon, CNRS, Université Lumière Lyon 2, Lyon, France, [2] Université de Lyon, Lyon, France, [3] Department of Psychology, Hallym University, Chuncheon-si, South Korea, [4] Laboratoire de Physique, Inst. Systèmes Complexes, ENS de Lyon, CNRS, Lyon, France

It is well-known that education related research is carried out within different disciplines and frameworks, but how is it specifically connected through citations to the larger social sciences and humanities? And how can this knowledge be mobilized to improve dialogue between researchers in different communities, given the benefits of integrating different frameworks and methods? We used different scientometric methods to show where exactly research in education connects to social sciences and humanities. This multidisciplinary context provokes a set of integration challenges for research in education. We propose how our work can supplement an existing model in order to give a framework for meeting these challenges with the goal of achieving broader education-related collective knowledge advancement.

Keywords: education research, social sciences and humanities (SSH), bibliometics, discipline (s) (of knowledge), multidisciplinarity, interdisciplinarity, transdisciplinarity, integration challenges

OPEN ACCESS

Edited by:
Arthur C. Graesser,
University of Memphis, United States

Reviewed by:
Chih Ming Chu,
National Ilan University, Taiwan
Susan R. Goldman,
University of Illinois at Chicago, United States

****Correspondence:***
Kristine Lund
kristine.lund@ens-lyon.fr

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *19 March 2020*
Accepted: *28 October 2020*
Published: *08 December 2020*

Citation:
Lund K, Jeong H, Grauwin S and Jensen P (2020) Research in Education Draws Widely From the Social Sciences and Humanities. Front. Educ. 5:544194. doi: 10.3389/feduc.2020.544194

INTRODUCTION

It is commonly argued whether diversity in a research field is a strength or a curse. Diversity of traditions is a strength when complex questions are answered with a broader perspective and when methodological complementarity can be discovered. It's a curse when participating researchers have difficulty communicating due to no shared vision or common understanding of the literature, thus leading to dispersion of discourse and an unstable epistemic community. The challenge is thus to build on the strengths while addressing the difficulties of such diversity. In this paper, given that educational research is carried out in close relationship to other research fields, we examine the nature of this diversity through bibliometric analyses, and ask four questions. The first two questions empirically examine the nature of educational research using a set of bibliometric methods whereas the second two questions ask what consequences such a nature implies, as well as what can be done, given these consequences.

1. What is the place of educational research in the larger context of social science research and how has it changed since the early 2000's?
2. How does educational research compare to the disciplinary composition of other fields?
3. What can such a disciplinary composition tell us about the challenges facing educational research?
4. And finally, what kind of model for broader education-related collective knowledge advancement can meet these challenges?

In what follows we motivate each of these research questions by giving preliminary definitions, explaining what we want to show, and describing why the second two questions are related to the first two. Finally, we illustrate why answering these questions are important for research in education.

The Connection of Research in Education to Research in Social Sciences and Humanities

Let's unpack the first research question. We are interested in understanding how research in education relates to social science research in terms of how closely connected the references of different publications are. When two publications cite another publication (i.e., a reference), we can surmise shared topics, but also shared theories and methods, depending on the nature of the shared reference. Publications that share references can be grouped into clusters that are either loosely or tightly connected around core references and relations between different clusters can also be evaluated in terms of which references connect clusters to each other. In this case, the notion of "place" refers first to sets of references in particular topic-based clusters that connect research in education to other research in education and second, to sets of references in particular topic-based clusters that connect research in education to research in the social sciences and humanities. This notion of "place" will have changed over some 20 years, between 2000–2004 (Education corpus 1) and 2015–2018 (Education corpus 2), the time periods studied for this article[1].

Is Research in Education Undisciplined?

The second research question addresses the disciplinary composition of a field of inquiry. The report on Interdisciplinarity Problems of Teaching and Research in Universities Apostel and Centre for Educational Research Innovation (1972) defines discipline as referring to "the tools, methods, procedures, exempla, concepts, and theories that account coherently for a set of objects or subjects" (Klein, 1990, p. 104); Miller (1982) argues that "within each discipline there are rational, accidental and arbitrary factors responsible for the peculiar combination of subject matter, techniques of investigation, orienting thought models, principles of analysis, methods of explanation and aesthetic standards." But a discipline is a fuzzy notion.

The level of integration of disciplinary approaches has survived as an indicator that distinguishes between the forms of so-called "non-disciplinary" research: multidisciplinarity, interdisciplinarity and transdisciplinarity. In (van den Besselaar and Heimeriks, 2001), neither theoretical perspectives nor actual results from different participating disciplines are integrated during multidisciplinarity. Rather, "the subject under study is approached from different angles, using different disciplinary perspectives (op. cit., p. 706)." Choi and Pak (2006) hold a similar view, arguing that multidisciplinarity draws on knowledge from different disciplines, but each researcher group stays within its own boundaries.

On the other hand, interdisciplinary research integrates contributing disciplines by creating its own theoretical, conceptual and methodological identity or in other words, "analyzes, synthesizes and harmonizes links between disciplines into a coordinated and coherent whole (Choi and Pak, 2006, p. 351)." The view emphasizing the integration of disciplinary perspectives as a marker of interdisciplinarity is a popular one [e.g., Birnbaum (1981), Cotterell (1979), Hanisch and Vollman (1983), Hausman (1979), Klein (1990, 1996), Kocklemans (1979), Epton et al. (1983), Hermeren (1986)]. But this notion of the necessity of some kind of integration for research to be labeled as interdisciplinarity is contested. Are participants in interdisciplinary projects purposefully taking an integrative stance? Lattuca (2003) argues that integrating presupposes a compatible framework in which such integration can take place — in other words, regardless of the disciplines concerned, interdisciplinary inquiry would naturally take the form of the scientific method found in the natural and physical sciences. This implies that each discipline's way of thinking about concepts, constructs, methods and theories are necessarily compatible if they can be integrated on an a priori basis into an agreed-upon general method of scientific inquiry. However, the measurement within a method is affected by the vantage point from which the phenomena in question are measured (Longino, 2013) and so if the general method of scientific inquiry or the levels of analysis are not compatible, then integration will be difficult. In addition, as Lattuca (op. cit.) also argues, perhaps some interdisciplinary projects attempt to redefine knowledge such as some scholarship in women's studies, ethnic studies, cultural studies and literary studies (Klein, 1996). So, while such redefinition might include integration of disciplinary perspectives, it may also include dismantling disciplinary perspectives rather than integrating them.

According to Gibbons et al. (1994), transdisciplinarity takes interdisciplinarity a step further. Whereas, interdisciplinary approaches explicitly formulate uniform discipline-transcending vocabulary or propose common methodologies, a transdisciplinary approach takes a common theoretical understanding and succeeds in integrating it into both participating disciplinary epistemologies. If enough researchers join in this effort, one could begin to refer to a new transdisciplinary field with a homogenized theory or set of models (e.g., social psychology or psycholinguistics). In other work, transdisciplinarity takes on a broader meaning. A transdisciplinary orientation works to overcome the disconnection between knowledge production on the one hand, and the demand for knowledge to contribute to the solution of

[1] We examine in detail the relation of research in education in the period 2000-2004 to research in the social sciences and humanities in the year 2000, but we cannot do a similar comparison for 2015-2018 and a more recent year in the social sciences and humanities corpus. For example, the 2000 corpus has around 120,000 publications and in 2018, it would have 600,000. The extraction alone for 2018 would be a challenge and the analysis would unfortunately require more computer power than we have available. In addition, given the large difference in the corpora, we could expect some surprises regarding the Louvain algorithm (cf. section Bibliographic Coupling and Construction of Network Clusters), thus requiring additional analyses and interpretations. We plan to address the challenge of examining the relation of research in education to research in human and social sciences with more recent data in a future publication. But we will describe the Education corpus for both 2000-2004 (Education corpus 1) and 2015-2018 (Education corpus 2) and this allows us to extrapolate on the recent relation between education aon the one hand nd social sciences and humanities on the other.

persistent, complex, societal problems on the other hand (Jäger, 2007). For Hall et al. (2012), a hallmark of transdisciplinary research is its focus on advancing progress toward practical solutions to social problems - for example, translating research findings into practice and policy applications. This latter view of transdisciplinarity is at the heart of action-research in education in that researchers, teachers, and policy-makers need to work together to address how to organize education for the developing child, adolescent, and young adult, but also for life-long learners. Such organizational attempts are carried out in particular socio-cultural contexts and will vary across countries, and involve a variety of stakeholders, often illustrating differences in underlying value systems. Is learning a pleasure or is it equated with effort and pain? Why do teachers get high salaries in some countries and low salaries in others? Disentangling the forces that enter into tension in such complex problem solving is a major challenge for policy making around teaching and learning.

Our position is that research questions may be investigated through a set of theoretical and methodological lenses that stem from different disciplines but also from different frameworks within the same discipline, which is to say that even within a discipline, there can be epistemological diversity. These lenses are heavily influenced by the vantage point from which the phenomenon under study is measured (Longino, 2013). In discussing our results, we will assimilate the notion of "Scopus subject area" to the notion of discipline (see below).

Challenges for Research in Education, Given Its Broad Scope

The third research question asks what the challenges are that research in education faces, given its well-known diversity in assumptions, theories, objectives, and methods. No matter the research domain, there are general challenges to be met that occur in multi, inter, and transdisciplinary research. However, there are also specific challenges having to do with the nature of the precise frontiers between research in education and research in the other social sciences and humanities. Both are useful in defining the challenges that face researchers in education.

Meeting the Challenges of the Diversity of Research in Education

The fourth research question is designed to investigate a way for the academic communities involved to move forward with broader education-related collective knowledge advancement, given the challenges of diversity. Such advancement can be guided with a model that mobilizes multiple theories that come from different disciplines, according to analytic need and questions posed. Each theory allows for observations that are oriented in particular ways and methods of interpretation that are based on specific assumptions, but there needs to be a way for them to be considered in concert. In addition, there should be a way of exploring how different fields may be posing similar questions that could be combined for a broader view of the phenomenon under study.

In what follows, we present the method, including database selection and extraction, the notion of bibliographic coupling and how it constructs the network clusters in education and social sciences research. The extent to which two articles are related by virtue of them both referencing the same research is the basis for our analysis. The nature of the shared reference (e.g., theory, method, etc.) determines the nature of the relation. Next, we present the results that allow us to discuss the above research questions. We end with a conclusion, a presentation of the limitation of our work, and ideas about future perspectives.

MATERIALS AND METHODS

Database Selection and Extraction

There are a number of electronic databases that index academic publications. They all provide wide coverage of academic publications and are practical to use. We chose to use Scopus. It has a similar overall coverage as the Web of Science, but a slightly wider coverage of non-English journals. That said, it remains unfortunate that so much research in other languages is not part of these large databases. As automatic translation progresses, hopefully the scope of the research world can be widened. Publication sources in Scopus are classified under four broad subject categories (Life Sciences, Physical Sciences, Health Sciences and Social Sciences & Humanities). These are further divided into 27 subject areas. The Scopus web interface does not provide a subject category corresponding to educational research. In order to select a corpus of publications in educational research, we relied on a list of educational research publication sources provided by AERES (AERES, 2014), the French national agency for the evaluation of research and higher education, which was in turn based on the European Educational Research Quality Indicator[2].

Figure 1A shows the evolution of the number of documents and corresponding number of journals in the AERES Education journal list being indexed by Scopus, since 1980. Scopus is known to focus mainly on recent articles (published after 1995), which explains the discontinuity observed in the figure. The increase in the number of indexed documents from 1996 to 2014, notably after 2005, is mainly due to the evolution of the number of journals coming into existence and being indexed. In this paper, we chose to compare two periods. The first is from 2000 to 2004 in order to get a sense of the research clusters before the steady increase of publications from 2005 (cf. **Figure 1A**). Choosing a time period before this initial increase also gave us a more manageable task. The second period is 2015–2018. **Figure 1B** shows the evolution of the number of documents within the Scopus "Social Sciences" broad category during the same period of 1995 to 2015. Similar to the "Education" list, we observe here a steady increase of the number of documents being indexed. This is also due to the evolution of journals coming into existence and being indexed by Scopus. In this paper, we chose to study the corpus of "Social Sciences" documents published in 2000. This snapshot allowed us to study the position of research in education

[2]EERQI project: http://www.eerqi.eu.

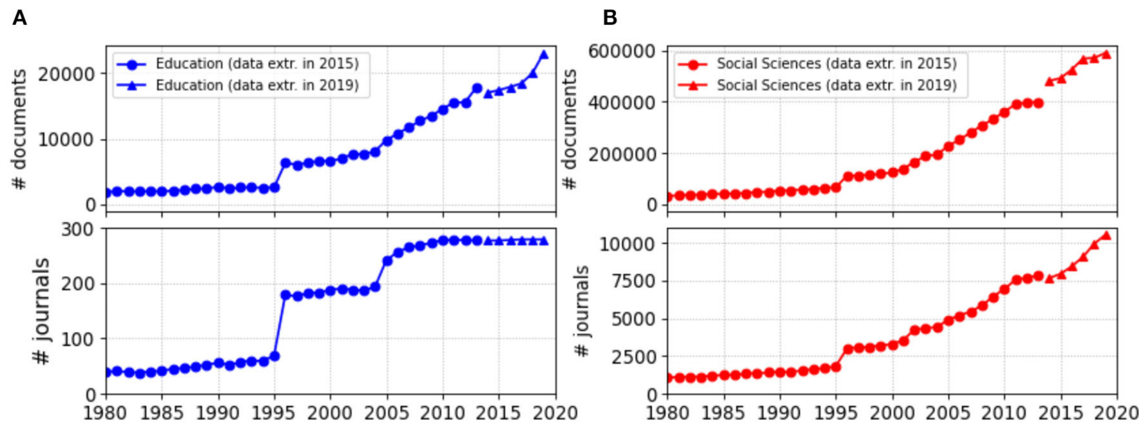

FIGURE 1 | (A) Evolution of the number of documents and corresponding number of journals in the AERES Education journal list being indexed by Scopus, since 1980 and **(B)** evolution of the number of documents within the Scopus "Social Sciences and Humanities" broad category.

within the whole field of social sciences while keeping the number of bibliographic records to extract from Scopus reasonable.

For the three corpora studied in this paper (on the one hand, for the two periods of the Education corpus and on the other hand, for the Social Sciences and Humanities corpus[3], we saved the full records of the documents: authors, names of publication source, years of publication, publication titles, keywords (given by the authors and/or Scopus), and the lists of references included in the publications. From the references lists, we additionally extracted authors, year of publication, and title of the reference. References were not always formatted consistently throughout all the records in Scopus (e.g., use of different abbreviations and/or inclusion of subtitles) and/or included missing information (e.g., final publication year of pre-print articles). All in all, around 2% of the references were wrongly formatted in the Education corpora and it was 6% in the Social Sciences.

In total, the extracted Education corpus 1 contains 36,715 bibliographic records from 2000 to 2004, the extracted Education corpus 2 contains 75,037 bibliographic records from 2015 to 2018 and the Social Sciences and Humanities corpus 122,936 records from 2000 (5.1% of which also belong to Education corpus 1). Publications in the Education corpora (as follows % for Education corpus 1, % for Education corpus 2) are mainly written in English (94/92.5%) by authors from the United States (44/35%) followed by the United Kingdom (14/12%), and Australia (5/7%). Publications in the Social Sciences and Humanities corpus are also written mainly in English (92%) by authors from the United States (33%) followed by the United Kingdom (14%), and Germany (4%). **Tables 1**, **2** list the publication sources that contribute the most articles to Education corpus 1 & 2. It illustrates that Scopus indexes peer-reviewed journals as well as professional magazines (such as Phi Delta Kappan).

The reader should note that these are the major *contributing* publication sources so these are the journals in which appear the most publications in our corpus. The list of education journals was much broader in scope.

There are a number of observations to make regarding how these contributing publication sources have changed between the periods 2000–2004 and 2015–2018. The Journal of Chemical Education still leads and Medical Education and Medical Teacher (Teaching) remain strong. But most strikingly, research involving computers has gained much ground. In the first period, there were no journals that dealt with education and computers by name (but see Educational Technology and Society) whereas in the latter period the second most contributing journal is Computers in Human Behavior and Computers & Education is in the top five. Research in Developmental Disabilities has also gained in visibility; it was not previously present and it is now the third most contributing journal. Child Development has subsequently dropped and specific journals on psychology have also disappeared (Teaching of Psychology, Journal of Educational Psychology). The professional magazine Phi Delta Kappan, focused on discussions of research, policy, and practice in K-12 education remains a steady contributor. Finally, perhaps research on more foundational knowledge (The Reading Teacher, International Journal of Science Education, Biochemistry and Molecular Biology Education) have given way to transversal concerns (learning and individual differences), as well as methodological orientations (Quality & Quantity).

Bibliographic Coupling

Scientometrics is a well-established field that applies mathematical methods to academic publications in order to understand science organization and evolution (see Mingers and Leydesdorff, 2015 for a review). It can be used simply to describe a field and its boundaries or it can be used as a first step in understanding the nature of research in a particular domain. It capitalizes on the citations by scientists to detect linkages between different articles, leading to the bottom-up building of

[3]Education corpus 1 is referred to as Educmap 2000-2004, Education corpus 2 is referred to as Educmap 2015-2018 and the Social Sciences and Humanities corpus is referred to as Sociomap 2000 in the on-line visualizations at http://sebastian-grauwin.com/XYZ_EDUCMAP/).

TABLE 1 | Major contributing publication sources in Education corpus 1 and Social Sciences and Humanities corpus.

Journal sources	Documents (N)	Documents (%)
Journal of Chemical Education	2,352	6.41
Medical Education	1,162	3.16
Educational Leadership	727	1.98
Child Development	609	1.66
Phi Delta Kappan	607	1.65
Medical Teacher	590	1.61
International Journal of Engineering Education	492	1.34
The Reading Teacher	397	1.08
Teaching of Psychology	388	1.06
International Journal of Science Education	373	1.02
Biochemistry and Molecular Biology Education	371	1.01
IEEE Transaction on Education	356	0.97
Science Education	355	0.97
Educational Technology and Society	343	0.93
Journal of Educational Psychology	337	0.92

TABLE 2 | Major contributing publication sources in Education corpus 2.

Journal sources	Documents (N)	Documents (%)
Journal of Chemical Education	1,855	2.43
Computers in Human Behavior	1,639	2.14
Research in Developmental Disabilities	1,590	2.08
Medical Teaching	1,458	1.91
Computers & Education	1,252	1.64
Medical Education	1,053	1.38
Quality & Quantity	861	1.13
Phi Delta Kappan	760	0.99
Child Development	729	0.95
International Journal of Engineering Education	704	0.92
Educational Leadership	696	0.91
Journal of Youth and Adolescence	677	0.89
Teaching and Teacher Education	645	0.84
Learning and Individual Differences	622	0.81
Science Education	606	0.79

homogeneous scientific subfields by clustering methods to be detailed below. Two of us have been working on scientometrics for several years, building a variety of tools to map scientific institutions (Grauwin and Jensen, 2011), the field of complex systems (Grauwin et al., 2012), or assess the interdisciplinarity of several hundred French laboratories Jensen and Lutkouskaya (2014).

Here, we use the well-known bibliographic coupling approach (BC; Kessler, 1963) to create a network, using articles as nodes and their common references as links. As we have argued elsewhere (Grauwin et al., 2012), BC achieves a faithful representation of the fields, giving equal weight to all published articles, regardless of whether and how often they are cited. Moreover, the links are established on the basis of the author's own decisions (to include or not to include a given reference) rather than retrospectively from other scientists' citations, as in the popular co-citation approach (Small, 1973). Thus, bibliographic coupling can be used to analyze the research clusters as they are built by researchers themselves.

Bibliographic Coupling and Construction of Network Clusters

In order to determine how different articles are linked through common references, we systematically compared the reference lists of two publications and identified shared references. Articles are linked if they share at least two references, leading to a network of articles connected to each other. The resulting network is schematically represented in **Figure 2A** where nodes represent individual articles. The thicker the link, the more references are shared between the articles. The links are weighted by Kessler's (1963) cosine similarity

$$\omega_{ij} = \frac{\left|R_i \cap R_j\right|}{\sqrt{\left|R_i\right| \left|R_j\right|}}$$

where R_i is the set of references of article i. By definition, the cosine similarity is equal to zero when two articles do not share any reference and is equal to 1 when their sets of references are identical.

Clusters are then detected using modularity maximization (Newman and Girvan, 2004) and the fast Louvain Algorithm (Blondel et al., 2008). Modularity quantifies the possibility to split a network into clusters in such a way that the links between nodes (i.e., articles) are dense inside clusters but not between them. There are many techniques available for clustering the nodes of a graph into relevant "communities" (for a review, see Fortunato, 2010). Thanks to its conceptual simplicity and easiness of computation, modularity is by far the most popular, even if its results should be interpreted with care (Good et al., 2010). In previous work on similar bibliometrics networks (Grauwin and Jensen, 2011; Grauwin et al., 2012), we have shown that the clusters obtained by modularity maximization do represent the scientific structure of research in a meaningful way.

An example of the resulting cluster membership of each node is represented in different colors in **Figure 2B**. Note that articles belonging to the same cluster (e.g., node 1 and 5 or node 12 and 15) are not always linked directly. Note also that articles belonging to different clusters may share links as well. The Louvain method detects clusters of nodes so that the number of "external" connections is as small as possible. In the networks studied in this paper, more than 70% of the links of an article are with articles belonging to the same cluster.

Clusters titles (**Figure 2C** shows placeholder names) were initially generated automatically as a function of frequent and significant title and keywords of their articles. This sometimes led to duplicate labels for the clusters, given shared overlap in research focus in educational research in general (e.g., two clusters with a "child" label). The final labels of the education network were determined after checking them against the

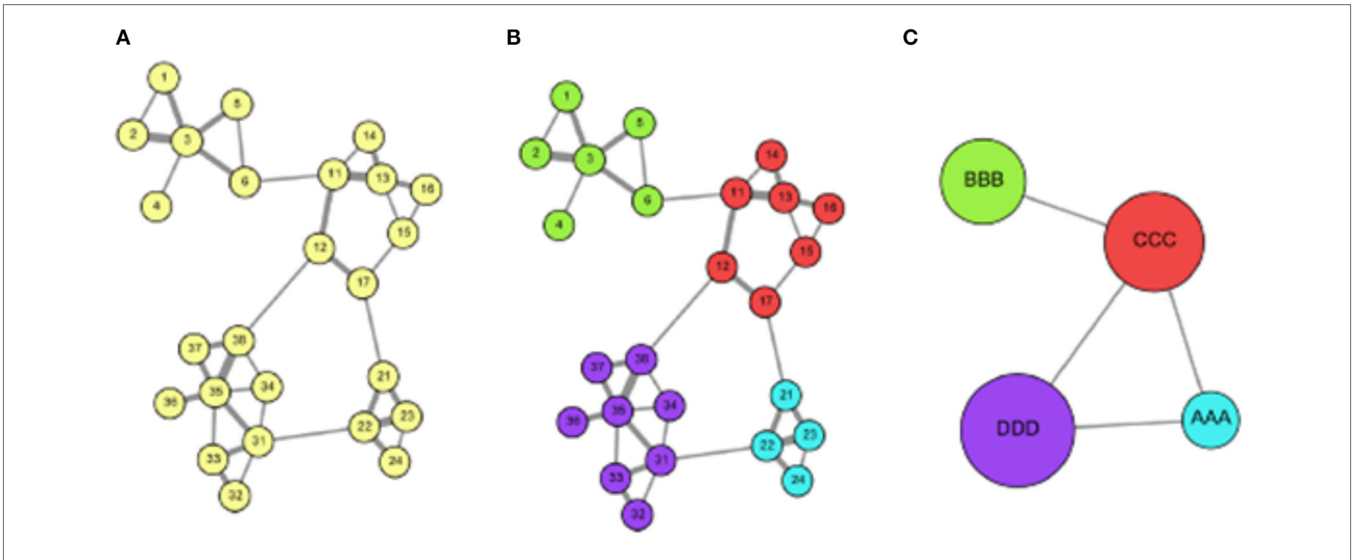

FIGURE 2 | Cluster detection. **(A)** Nodes are individual papers. **(B)** Colored nodes represent emerging themes of clusters of papers. **(C)** Finalized clusters of papers with placeholder title names.

references and other aspects of the clusters so that the labels would uniquely represent the clusters. Automatically generated labels were kept for social sciences and humanities clusters as they were distinct enough due to the diversity of research in social sciences and humanities research at large.

Cluster Cohesiveness as an Indicator of Cluster Focus

In **Figure 2**, the thickness of an edge between two clusters is proportional to the extent to which they share the same references, that is, the average weight between articles of both clusters. Note that this link arises mostly from shared references that are not within the clusters' core references. Summing up the weight of all edges of the cluster pairs, one obtains the weighted degree of a cluster, a "centrality" measure in the language of network theory, which indicates the overall connectivity of the cluster (see **Tables 3**, **4** in the results section). Clusters with a high W value share more references with other clusters than clusters with low W values, which use more specialized references. The layout algorithm tends to position highly connected clusters at the center of the map and less connected clusters on the periphery of the map[4].

It follows then that clusters can vary in size and cohesiveness, that is, they can vary in the extent to which articles are connected to one another within the same cluster. In cohesive clusters, articles are highly connected to each other around a set of core references. In less cohesive clusters, the connections are not as strong and may not be homogeneously distributed so that further breakdown into sub-clusters can be done if needed. The relevance of such a sub-partition can be measured by an internal modularity measure Q varying between−1 and 1, where values above 0.4 are often considered as an indication of a relevant sub-partition.

RESULTS

In reporting these results, we are working from the assumption that these clusters correspond to the different areas of educational research, given that articles in the same cluster share at least two references with another paper of that cluster. We first show through cluster cohesiveness how clusters in Education corpus 1 and Education corpus 2 focus either specifically on education or can be broken down into other clusters. In other words, one cluster will focus on education, but others will focus on other subjects that are aligned with the social sciences and/or humanities disciplinary focus of the cluster in question (cf. section Education clusters focus either clearly on education or in addition to other topics). We also point out how these clusters have evolved between the periods of 2000–2004 and 2015–2018, yet maintain their interdisciplinary footprint. Then, we show how a similar phenomenon plays out in the Social Sciences and Humanities corpus from 2000 where clusters that focus on education appear in different Scopus subject area regions of the map (section Overall Structure of Educational Research Within Social Sciences and Humanities). These two sections answer our first research question by illustrating the links that research in education has to research in social sciences and humanities, first from the bottom-up perspective of the cohesiveness and focus of clusters labeled as education with our bibliographic method, and then from the top-down perspective of in which Scopus subject area

[4]The overall spatial structure of such clusters presented in **Figure 4** is obtained thanks to a force-directed algorithm, used to draw the network in an aesthetically pleasing way. This algorithm simulates a physical system by assigning 3 types of forces among the nodes and links of the network: (1) a gravity force, that attracts every node to the center of the graph, (2) a spring-like attractive force, that attracts linked nodes toward each other, (3) a repulsive force, that tends to separate all pairs of nodes. The final layout results from an equilibrium between all these forces.

TABLE 3 | Education corpus 1 cluster sizes (N), Normalized Weighted Degree (W) and Internal Modularity (Q_i).

Clusters	*N*	*W* (%)	Q_i
Science of learning	1,883	11.2	0.43
Motivation	1,514	10.1	0.47
Science education	1,370	10.1	0.48
Math education	685	8.5	0.36
Teacher training	799	8.2	0.64
Cognitive studies of learning	790	6.8	0.59
Evaluation & assessment	733	5.7	0.59
Educational equality	1,800	5.4	0.67
Reading education	1,140	5.3	0.43
Measurement	648	5.3	0.61
Higher education	1,207	4.3	0.68
Cooperative learning	554	4.2	0.79
Child behavioral development	1,534	3.4	0.71
Language teaching methods	675	2.8	0.57
Sociology of education	1,715	2.5	0.71
Child cognitive development	415	2.5	0.64
Civic education	417	2.1	0.79
Developmental disabilities	667	1.5	0.82

Clusters are ranked from highest to lowest by W. Sum of all W equals 100.

TABLE 4 | Education corpus 1 cluster sizes (N), Normalized Weighted Degree (W) and Internal Modularity (Q_i).

Clusters	*N*	*W* (%)	Q_i
Adolescent behavior	4,847	13.85	0.48
Navigation behavior	440	13.27	0.45
Science education	4,692	12.41	0.52
Interactive learning environment	1,932	10.13	0.49
Self efficacy	4,612	9.75	0.62
Problem based learning	3,378	9.75	0.60
Medical school	4,894	6.48	0.55
Social networking (online)	3,950	5.86	0.50
Pathophysiology	7,498	5.38	0.65
English languages	1,241	3.89	0.71
Medical education	2,364	3.27	0.78
Bullying	1,332	3.2	0.68
Education policy	12,887	2.48	0.69
Biochemistry	936	0.28	0.83
School health service	729	0	0.78

Clusters are ranked from highest to lowest by W. Sum of all W equals 100.

space such clusters are located within the larger social sciences and humanities.

In section Educational Research in Social Science Research at Large: A Deep Multidisciplinary Interface, we address our second research question and report on how education research compares to the disciplinary composition of other fields, illustrating its deep multidisciplinary interface. In section The Challenges that Face Researchers in Education we evoke the challenges of such multidisciplinarity and we propose a framework for meeting them.

Education Clusters Focus Either Clearly on Education or in Addition to Other Topics

Table 3 shows the cluster sizes (N), the normalized weighted degree (W) and the internal modularity (Qi) of each education cluster. The first clusters in **Table 3** are highly cohesive and clustered around a set of core references. These core references reveal an education focus, once they are analyzed. The last clusters in **Table 3** are less cohesive and clustered around multiple sets of core references. This means they have multiple foci having to do with multiple subject areas, one of which, at the very least, is education related. Clusters are designated here by their most significant keywords among the 20 most frequent ones (cf the online interactive map http://sebastian-grauwin.com/XYZ_EDUCMAP/BCclusters.html).

A closer look at the top five highly connected, cohesive and bottom five loosely connected clusters in **Table 1** allow us to delve more deeply into the first research question: what is the place of educational research in the larger context of social science research? If we explore the top five highly connected, cohesive clusters from **Table 3** (i.e., "Science of Learning," "Motivation," "Science Education," "Math Education," and "Teacher Training"), they appear to focus on teaching and learning directly. This can be in terms of theoretical approaches to education (e.g., socio-constructivist view), of teacher and student attitudes and identity, or practices in the classroom. The focus can also be more specific, on math and science teaching and learning practices. These clusters tend to be organized around references that connect the whole network (e.g., Lund et al., 2015, 2017). On the other hand, the bottom five loosely connected clusters (i.e., "Language Teaching Methods," "Sociology of Education," "Child Cognitive Development," "Civic Education," and "Developmental Disabilities") deal only partially with teaching and learning practices. Their main focus is on different fields (i.e., linguistics, sociology, psychology) and education is only one object of inquiry, one example of application among others.

Some specific examples may help to understand why some clusters concentrate only peripherally on education. In the sub topic cluster "Developmental Disabilities," most sub topic clusters are tied to the medical or psychology domain, as seen for example by their high use of the DSM manual (i.e., the Diagnostic and Statistical Manual of Mental Disorders). The sole strong connection to education flows through a psychology sub topic cluster, as a possible application of their more general work on disabilities. A similar phenomenon can be seen for the "Child Cognitive Development" cluster. Only a single sub-cluster is strongly connected to education, with articles focused specifically on the impact of child cognitive development on education. Similarly, education is but one example of a political phenomenon for the "Sociology of Education" sub topic cluster, which deals with general sociological matters as globalization, or colonialism. Take as an example the title of one of its most connected articles: "Re-thinking trust in a performative culture: the case of education" (Avis, 2003). Finally, despite what its name might imply, the "Language Teaching Methods" sub topic cluster does not solely focus on education either. Some research focuses on language policy and the politics of development or more specifically linguistic topics such as syntactic complexity in synchronous and asynchronous communication. Therefore, whereas the highly connected, cohesive sub topic clusters are more focused specifically on learning and teaching *per se*, at times in specific topic areas that make sense to the sub topic cluster (e.g., the science education sub topic cluster will focus on physics learning), the loosely connected, less cohesive sub topic clusters are recognized as focusing more on disciplines in their own right (e.g., the sociology of education sub topic cluster deals with general sociology in a variety of ways), in which there are links to many different foci of research, one or more of which happen to be education related.

This phenomenon of cohesiveness is repeated at the level of the social sciences and humanities corpus (see next section). In other words, the articles which deal with education in this larger corpus network are loosely connected together, yet span different sub topic clusters that each belong to a set of different Scopus subject areas. This characteristic will be further explored in section Educational Research in Social Science Research at Large: A Deep Multidisciplinary Interface.

Before we turn to examining the overall structure of educational research within social sciences and humanities, let's review the top five highly connected, cohesive and bottom five loosely connected clusters in **Table 4**. This will allow us to compare the Education corpus clusters from 2000 to 2004 to the clusters from 2015 to 2018 and give evidence for how the place of educational research in the larger context of social science research may have evolved. First, it's striking how much the landscape has changed. Science education is the only cluster remaining from the period 2000–2004. It is still a highly connected and cohesive cluster, and research here has continued to have this characteristic for almost 20 years. This aside, **Table 4** is not as clear cut as **Table 3** where the highly connected, and cohesive clusters were focused on learning and teaching. Here, apart from the third cluster ("Science Education,") "Adolescent Behavior," "Navigation Behavior," "Interactive Learning Environment," and "Self Efficacy" focus rather on behavioral and contextual aspects of learning. In addition, the sources referenced are quite multidisciplinary, connecting also to psychology for Adolescent Behavior, and Navigation Behavior. This latter also connects to the field of education sciences through sources that are highly influenced by psychology (Educational Psychologist, Learning and Instruction, Cognition and Instruction, and Instructional Science). Unsurprisingly, the "Interactive Learning Environment" cluster's most frequent references' sources include journals from computer science, psychology, and education. And looking to the loosely connected clusters ("Bullying," "Education Policy," "Biochemistry," and "School Health Service"), only "Biochemistry" stands out as focused on teaching and learning of that subject, whereas the others connect to other disciplines that treat the topic in question from different perspectives (e.g., "School Health Service" is examined by pediatricians, but also through public health and behavioral psychology; "Bullying" is the focus of child development, but also studied through the lens of child psychology and psychiatry, through disability research, and also by taking into account aggression and adolescent health).

In sum, the latter period of research in education (2015–2018) exhibits a more multidisciplinarity approach for a given topic, whether or not the clusters are highly connected and cohesive or loosely connected. We could dare to hypothesize that research in this area is slowly accomplishing more and more integration across frameworks and approaches.

Overall Structure of Educational Research Within Social Sciences and Humanities

We can further deconstruct the notion of "place" of educational research in the larger context of social sciences and humanities research if we begin by illustrating the top-down characterization that Scopus proposes through its categorization of articles by Scopus subject area. The cluster map in **Figure 3** was formed by gathering all "social science and humanities" articles published in year 2000 from the Scopus database. This yielded 122,936 articles, and the colors signify the Scopus subject areas in which the journals publishing articles are categorized. We performed our cluster detection analysis on these articles, which allowed us to combine our bottom-up approach of bibliometric coupling to the top-down approach subject area categorization of Scopus. But we need to figure out where research in education is situated, within the network. There are five major regions which we can characterize according to how Scopus puts publication sources into subject areas: neuroscience (top right, orange), psychology (also orange, but in addition green, pink, middle red and brown), general social science/sociology (light blue, in the middle/right), business and management (yellow) and economics and finance (bottom, pink/purple)[5].

Education does not automatically appear as a single region with its own specific color, because Scopus does not have a subject area for education (see the materials and methods section where we explain how AERES and EERQI help us create this subject area). This paper can be viewed as an attempt to answer why Scopus might not have such a subject area called education. In order understand the extent to which research in education was present in the Social Sciences and Humanities corpus, we examined the relative proportion of the articles from the "education journals" database in each of the clusters of **Figure 3**; this is plotted in **Figure 4** on a red scale. To simplify the analysis, we restricted it to the social science and humanities clusters that contain more than 100 articles.

Results revealed four major and nine minor education sub topic clusters within the social sciences and humanities map (labeled and shown in red in **Figure 4**). If education had been a subject area that was well-defined in relation to the boundaries it shared with other Scopus subject areas, we could have expected there to be a clearly defined region. Instead, education is distributed over many different regions, which points to a loosely connected field with multiple areas

[5]See http://sebastian-grauwin.com/XYZ_EDUCMAP/BCclusters.html for more details.

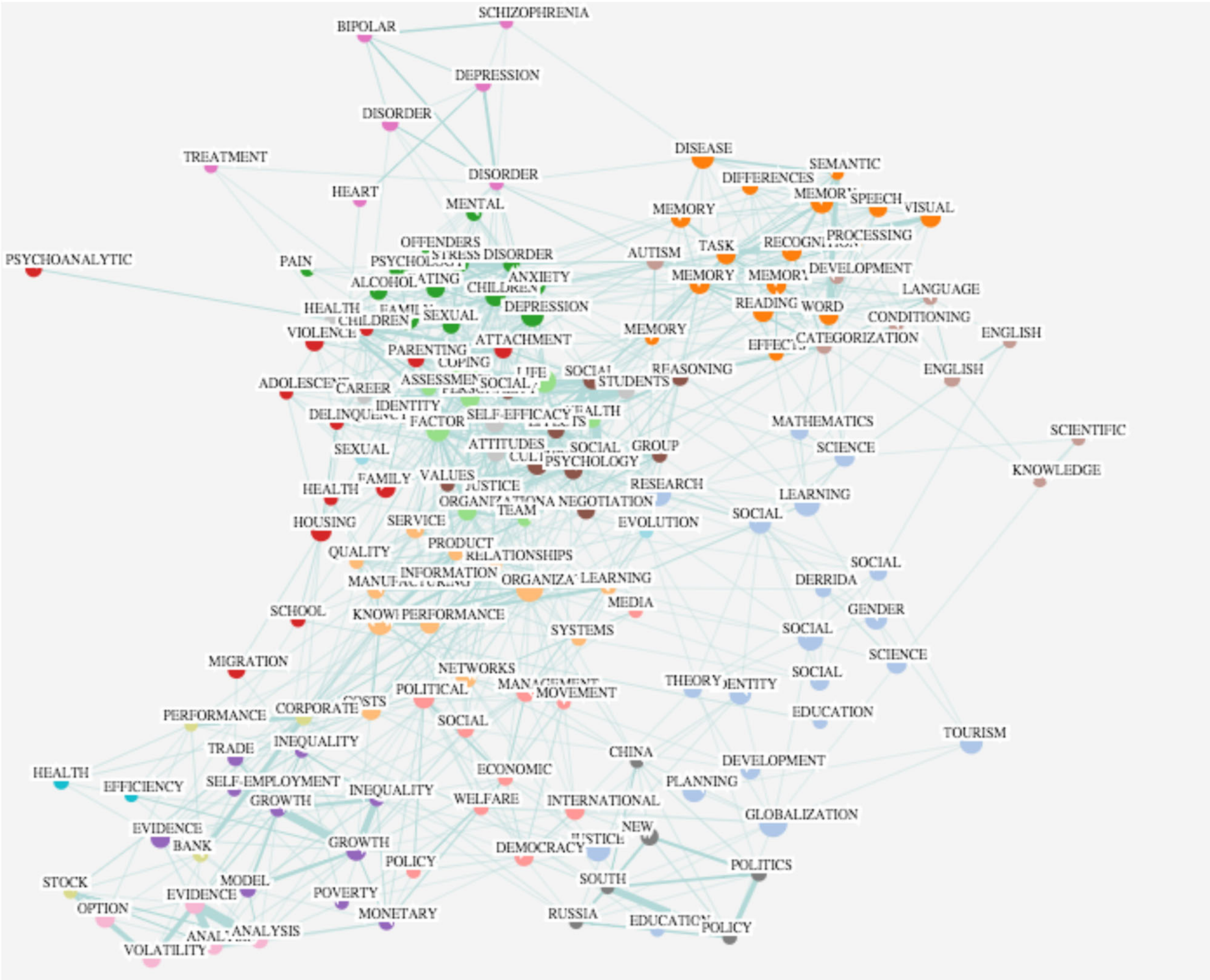

FIGURE 3 | Social Sciences and Humanities network sub topic clusters (in colors) obtained by the partitioning algorithm on the bibliographic network built on around 123,000 articles published in year 2000 in the Scopus "Social sciences and humanities" broad category. The area of the nodes corresponds to the number of articles of that cluster and the colors to the corresponding topic clusters, each of which would subsume all of colored sub topic clusters into one colored cluster. The labels correspond to the most common title keyword of the source publication. A few sub topic clusters, disconnected from the main network, containing mostly biology articles from Nature, Science and Proceedings of the National Academy of Sciences (PNAS), are not shown.

of study connecting to different subfields of social sciences and humanities, corresponding in nature to the bottom clusters of **Table 2**, within Education corpus 1 and Education corpus 2.

The four major sub topic clusters we see in **Figure 4** refer to social sciences and humanities clusters in which more than 50% of the articles are published in education journals. These are the sub topic clusters labeled "Mathematics" (87% of articles published in education journals), "Science" (72%), "Learning" (61%) and the larger "Education" sub topic cluster at 59%. Given their high rate of publication in journals that are labeled as education journals, these four sub topic clusters have a strong education coloring.

It is noteworthy that overall, the highlighted sub topic clusters that tend toward red in the social sciences and humanities map can be matched to a single Education corpus 1 sub topic cluster in **Table 3**, the "Science Education" sub topic cluster. Seventy-two percentage of the articles within the social sciences and humanities sub topic cluster "Science" are published in a journal in our list of education source publications. And among these, 80% can be found in the "Science Education" sub topic cluster of **Table 3** while the rest are distributed among other sub topic clusters. This observation shows that Scopus' top-down method of categorizing articles according to the Scopus subject area of the journal in which they are published and our bottom-up method of clustering articles together if they share references can triangulate. They both show a focus on Science in the year 2000 that come together in the respective analyses[6].

[6]Compare the clusters in Sociomap 2000 at the topic level to the clusters in Educmap 2000-2004 at the topic or sub topic level http://sebastian-grauwin.com/XYZ_EDUCMAP/BCclusters.html.

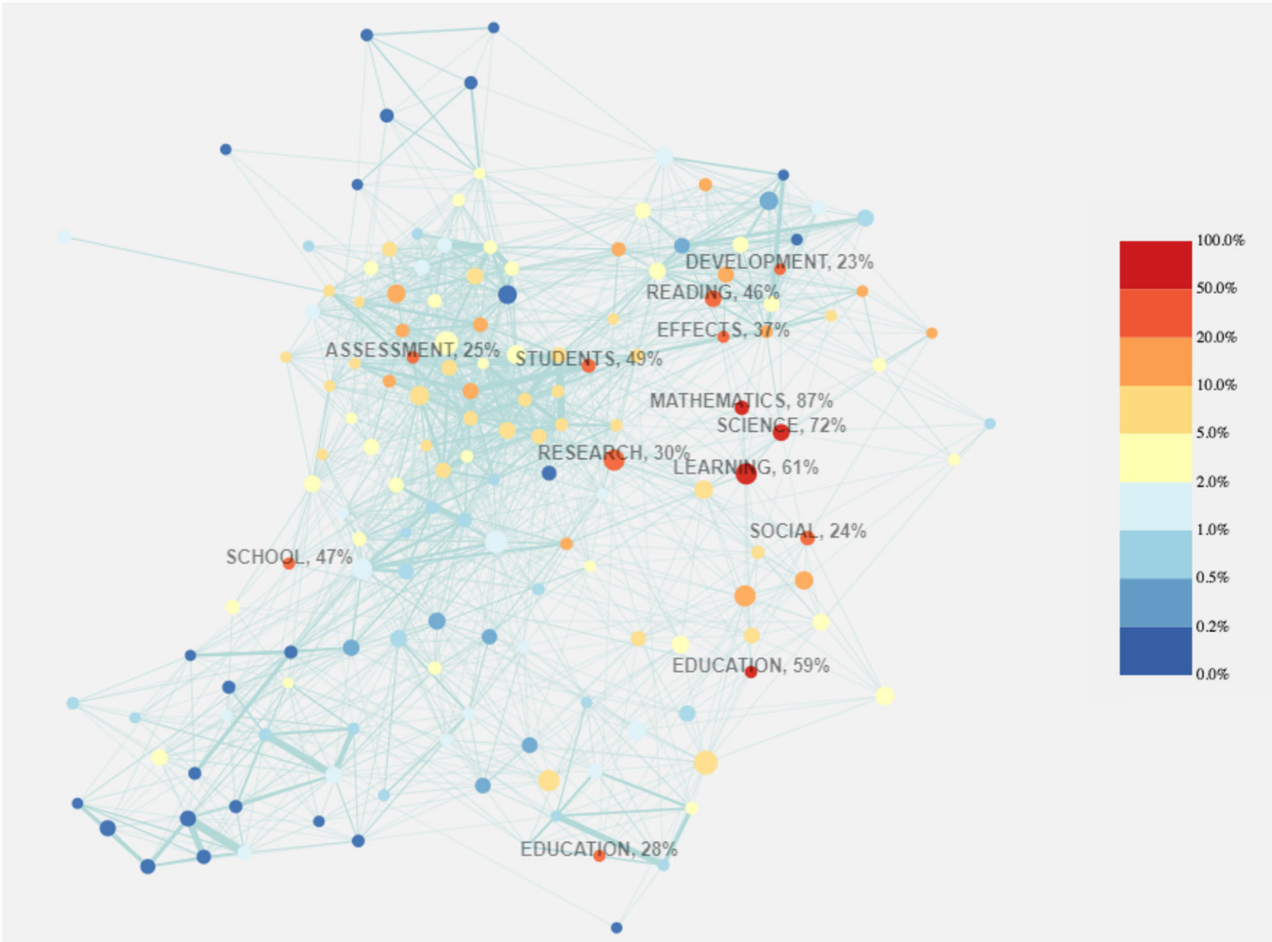

FIGURE 4 | Social Sciences and Humanities clusters (**Figure 3**) according to the proportion of articles of each sub top cluster that are also found in the Education corpus. The closer to red, the higher the proportion, the closer to blue, the lower. Four sub topic clusters contain more than half of the articles from education: "mathematics" (87%), "science" (72%), "learning" (61%), and "education" (59%).

The nine minor education sub topic clusters in **Figure 4** refer to clusters containing between 20 and 50% of education articles and reach toward other Scopus subject areas in terms of the nature of references of articles that are shared. They therefore have a weaker education coloring. There are three sub topic clusters within the orange neuroscience nodes of **Figure 3** ("Reading" 46%, "Effects" 37%, and "Development" 23%), four within the green, pink, brown, and red nodes of psychology ("Students" 49% "Research" 30%, "School" 47%, and "Assessment" 25%) and two within the light blue nodes of general social science and sociology ("Social" 24% and the smaller "Education" cluster at 28%).

Educational Research in Social Science Research at Large: A Deep Multidisciplinary Interface

In this section, we set out to answer our second research question:

How does educational research compare to the disciplinary composition of other subject areas?

As can be seen in **Figure 4**, education clusters are scattered over a wide region of the map, suggesting that educational research is carried out in close connection with a number of other research areas in social sciences and humanities research, rather than forming an isolated region. This scattering of education clusters over multiple disciplinary areas of the Social Sciences and Humanities corpus suggests that research in education is quite multi-disciplinary and wide-reaching, in that different intellectual traditions and areas of study coexist in a loosely connected fashion. In what follows, we quantify this interpretation in three ways. First, we compute the extent to which articles are connected via references to articles of the same Scopus subject area. We use the Scopus subject areas of the journals where the articles are published and remind the reader that journals may belong to several Scopus subject areas. The main Scopus subject areas used here are: "Business, Management, and Accounting," "Decision Sciences," "Economics, Econometrics and Finance," "Medicine," "Psychology." We add "Education" for those articles published in the list of education journals we built, for as we have mentioned, the Education subject area does not exist in Scopus. The results confirm that articles having to do with education are more likely to connect to articles in other Scopus subject areas than articles of other Scopus subject areas. For example, “Education” articles are equally likely to be linked to articles belonging to Education and to articles belonging to Psychology (39% of the links for both Scopus subject areas). And on the contrary, “Economics” articles are mostly linked to other Economics articles (60%).

A second way is to look at the sub topic cluster level. We qualify whether these sub topic clusters gather articles from different Scopus subject areas or not. The multidisciplinary indicator M (d1, d2) quantifies the co-presence of two Scopus subject areas d1 and d2 within sub topic clusters. By extension, one can define a single Scopus subject area indicator M (d, d) to quantify the proportion of articles of a single Scopus subject area d within the clusters. Formally:

$$\mathrm{M}\,(\mathrm{d1}, \mathrm{d2}) = \sum(\text{sub topic cluster i})[\text{fi}(\mathrm{d1})^{*}\text{fi}(\mathrm{d2})^{*}\text{size i}] / \sum(\text{sub topic cluster i})[\text{fi}(\mathrm{d1})^{*}\text{size i}]$$

where fi (d1) represents the percentage of articles of Scopus subject area d1 within sub topic cluster i and sub topic size i its size. This indicator computes the proportion of articles of Scopus subject area 2 for the clusters where there are articles of Scopus subject area 1 [i.e., fi (d1) is not 0], weighting the contributions by the number of articles of d1 [the factor fi (d1) * size i]. In other words, our indicator quantifies whether articles from different Scopus subject areas share many references with articles from other Scopus subject areas, which would indicate commonalities. Note that since many articles are attributed by Scopus to several Scopus subject areas, the sum of the frequencies of the different Scopus subject areas is not exactly equal to 1. Note that this indicator is independent of the number of articles of the Scopus subject area.

One obtains the following single-subject area relations: M (Education, Education) = 0.34; M (Economics, Economics) = 0.52; M (Psychology, Psychology) = 0.65; M (Business, Business) = 0.44; M (Decision, Decision) = 0.16 (see **Figure 5**). The value is the highest in psychology followed by economics, indicating that the psychology and economics sub topic clusters tend to be mono-disciplinary, i.e., articles from these Scopus subject areas share references mostly with articles from the same Scopus subject areas. On the other hand, education articles belong to social science and humanities sub topic clusters that contain only a third of articles from education. Similarly, “Decision sciences” articles belong to sub topic clusters that contain only a quarter of articles from Decision Sciences, showing that “Decision sciences” articles are more connected to articles belonging to Business and Economics Scopus subject areas.

From which Scopus subject areas do the rest of the articles concerning education come? As can be seen in **Figure 5**, the two most important multi-disciplinary indicators for Education are: M (Education, Psychology) = 0.35 and M (Education, Medicine) = 0.13. This means that, on average, education articles belong to clusters that contain, beyond the third of education articles, a third of Psychology articles and 13% of articles belonging to Medicine. In other words, Education articles share many references with articles of these Scopus subject areas, which indicates commonalities. These results suggest that Education as well as Decision sciences, instead of forming a single, well-delineated closed discipline, tend to form a loosely-connected and distributed discipline that actively incorporates diverse knowledge bases in social sciences and humanities fields.

The third way of confirming the multidisciplinary character of Education uses the clustering coefficient (cc) to quantify the disciplinary breadth to which articles are linked. This standard characterization of the nodes of a graph is whether the neighborhood of a node is tightly connected, that is, whether an article connects articles that are already connected or whether it connects otherwise unconnected articles. Formally, it is defined as the ratio of the number of links among the neighbors of a node to the number of possible links, that is, d*(d-1)/2 where d is the degree of the node. For a given article, cc = 0 means that there is no connection (no common references) among its neighbors, while cc = 1 means that all the articles connected to it are also connected among themselves. In most social networks (Newman, 2010), average cc values are around 0.2–0.5, meaning that there is a high probability that two of my “acquaintances” also know each other. In our network, cc corresponds to the proportion of articles in the neighborhood of an article that share more than two references. One can expect that cohesive subject areas show high values of cc, since many articles share similar references, while subject areas that connect distinct thematic areas are likely to have low cc values, as articles in those distant subject areas are not often connected.

We have computed the cc for all the articles in the social sciences and humanities network. The average cc is 0.29, while that of articles published in Economics journals is significantly higher (0.32, $p < 0.001$), and that of articles published in Education journals is significantly lower (0.28, $p < 0.02$). This means that the neighbors of articles in Economics also share references between each other whereas the neighbors of articles in Education do not necessarily share references between each other, and may share references with distant articles, in terms of Scopus subject area. We note that the Scopus subject area with the highest cc is Psychology (0.34, $p < 0.001$) which means that an article in Psychology has the most neighbors who share references. The same result is found when looking at the cc at the level of the journals, where nodes are entire journals that are linked whenever they contain articles sharing at least 2 references. In this aggregated network, the average cc is 0.38, Economics journals have an average cc of 0.41 ($p < 0.001$), Psychology journals have an average cc of 0.42 ($p < 0.001$), while Education journals are below average (0.36, $p < 0.01$). The same result is obtained when averaging cc for the articles in the different clusters of **Figure 4**: clusters that contain mostly Psychology articles show a larger average cc than those in Economics and Education.

In summary, education related research is widely scattered across research in social sciences and humanities, sharing references with different subject areas. First, articles having to do with education are more likely to connect to articles in other Scopus subject areas (in particular, psychology) than articles of

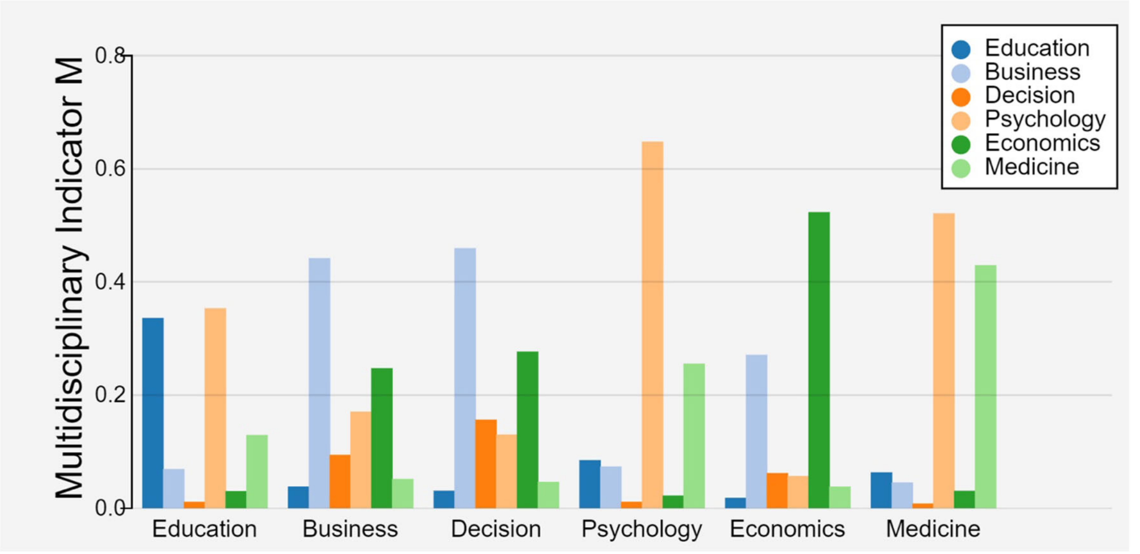

FIGURE 5 | Cumulative multidisciplinary indicator (Scopus subject area) for each subtopic cluster.

other Scopus subject areas. Secondly, in qualifying whether sub topic clusters gather articles from different Scopus subject areas or not, we see that Education first gathers articles essentially equally from both Education and Psychology and then next from Medicine. Thirdly, in quantifying the disciplinary breadth to which articles are linked, we see that Psychology has the least breadth, but the most cohesiveness whereas Education has the most breadth and the least cohesiveness. In conclusion, Education is the most multidisciplinary.

The Challenges That Face Researchers in Education

Given that we have empirically shown the cross-disciplinary connections between research and education and social sciences and humanities, in this section we discuss our third research question, which arises naturally from this state of affairs:

3. What can the disciplinary composition of the subject area of Education tell us about the challenges facing educational research?

Earlier in the paper, we presented the different types of cross-disciplinary relations, including multi, inter, and transdisciplinarity. In what follows, we discuss both the drawbacks of working across disciplines, and the reasons for venturing outside of disciplinary boundaries (Lund, 2016). Each of these are first presented from a general point of view and then specifically from the point of view of education related research. In this paper, we will argue in favor of working across disciplines in ways that are adapted to the objectives for and the context of education research. We will propose a model in response to our fourth research question that deals with meeting the challenges that we detail below.

Drawbacks of Working Across Disciplines

Klein (1990) notes three major difficulties facing interdisciplinary scholarship: general uncertainly over definition, lack of professional identity and dispersion of discourse. First, is interdisciplinarity just nostalgia for a lost wholeness or is it a new stage in the evolution of science? Is it a historical quest for unified knowledge or is its goal to develop the frontiers of knowledge? Second, some proponents of such scholarship are wary of organizing professional interdisciplinary movements because institutionalization may bring about insularity, and avoiding insularity was one of main reasons they were attracted to interdisciplinarity in the first place. Third, the discourse on interdisciplinarity is widely dispersed and so commonalities that could be shared are simply not available for those who could benefit. That said, the explorable on-line maps that we provide are meant to remediate this.

One of the more practical difficulties of working across disciplines is the time it takes to perform the intellectual work necessary to consider their compatibility or their incommensurability (Latour, 2005). It may be more efficient to stay within the boundaries of a discipline where the frameworks are well-defined and where the type of maneuvering is well-understood. Kuhn (1962/1970) called it doing "normal science" where details are slowly accumulated in accordance with an established broad theory and where there is no need to question or challenge the underlying assumptions of that theory. In this way, a researcher's energy can be put toward reaching specific disciplinary scientific objectives. It's difficult to use a framework in an effective way while you are questioning it.

Another drawback of working across disciplines is difficulty in framing research to be published so that it both fits the aims and scope of existing journals and so that it does not fall victim to be found lacking in one way from one disciplinary perspective and in another way from another disciplinary perspective.

Reasons for Venturing Outside of Disciplinary Boundaries

Klein (1990) notes a wide range of objectives that educators, researchers and practitioners have pursued through interdisciplinary work: answer complex questions, address broad issues, explore disciplinary and professional relations, solve problems that are beyond the scope of any one discipline, and achieve unity of knowledge, whether on a limited or grand scale.

In addition, communities of researchers may work separately on seemingly similar objects of study while ignoring each other's efforts. Asking how the efforts of others can relate to one's own efforts is another reason for venturing into another discipline's territory. In fact, the key word here is "seemingly." Once researchers gather data and make explicit their purpose of analysis and methods, it becomes clear that what is considered the same phenomenon is in fact very different. Although the phenomenon in question may seem to be the same at a general level (e.g., group interactions), it often is not the same phenomenon once the data has been gathered and the purpose of analysis is made explicit. For example, experimental psychologists will typically gather data on group interactions in controlled laboratory situations that are specifically designed to test a hypothesis concerning how a variable affects either group process or outcome whereas interactional linguists will more likely gather data on group interactions in naturally occurring situations with the goal of describing the ways that participants co-organize their actions. At first glance, it may not seem obvious what these psychologists and linguists would say to each other.

But are they missing opportunities for the advancement of scientific knowledge on group interactions by staying anchored in their respective communities? How is it different to do research within one discipline vs. in a way that reaches across disciplines? Could the latter be more productive or at least provide new opportunities for innovative research questions? If the group interactions that are in question have learning as a process as their goal, there are indeed ways that psychologists and linguists can bring their unique focus to bear on this question. The former can pinpoint which variables have effects on process in particular situations and the latter can describe in detail how these effects are manifested and co-constructed in the process. Such an approach gives a broader and richer view of the data.

Meeting the Challenges That Arise From the Diversity of Research in Education

Given the nature of research in education and how it is strongly connected to research in social sciences and humanities, and also given the various challenges that this situation provokes for the researcher, in this section, we discuss our fourth research question:

4. What kind of model for broader education-related collective knowledge advancement can meet these challenges?

So far, we've seen that research in education is carried out from within different institutional vantage points that may or may not have particular advantages. Different disciplinary viewpoints are taken on the same object, different theoretical frameworks are mobilized, data of different natures are focused on, and a variety of methodological approaches are used. Even if such diversity makes for fascinating research, it also brings about missed opportunities.

We begin by giving five examples of missed opportunities involving communication between researchers in fields relating to research in education and then we present a model that is supplemented with the exploration of our Education and Social Sciences and Humanities corpora. The combination of this model and the exploration of our corpora is designed to build broader collective knowledge by bringing together existing expertise to address missed research opportunities, given the diversity of research in education related topics.

Absence of Communication

A first type of missed opportunity is the absence of communication between disciplines or between sub-domains in a given field of inquiry. In what follows, we note five examples that all have to do with different deficiencies in communication between researchers. First, delving into our scientometric maps can illustrate the different ways in which theoretical constructions are used. Scientific analyses are fragile when they are carried out from one point of view. This approach radically limits conclusions. Operationalizing theoretical constructions under different foci makes them more robust (Rosé and Lund, 2013). Our approach allows us to identify zones where theoretical constructions are operationalized differently. For example, the research that treats personal epistemology, epistemic cognition and development, beliefs, theories and epistemological resources are dispersed in different clusters and do not share the same references, although these constructions could benefit from being compared.

A Divided Field That Leads to Pursuing Different Objectives

A second example of missed opportunity around communication occurs when a field divides in order to pursue specific objectives, but does not maintain contact with the evolution of the other part of the field. The two domains Computer Supported Collaborative Work (CSCW) and Computer Supported Collaborative Learning (CSCL) are an example. This division allowed them to concentrate, respectively, on work and learning. But now there are attempts to re-integrate findings through the study of CSCL @ work (Goggins et al., 2013).

Foundational Differences That Hinder Collaboration

A third example of missed opportunity around communication occurs when two fields share a goal, but due to differences in foundational theories and models, do not build on each other's research (Kirby et al., 2005). For example, Learning Sciences and Instructional Systems Design share an interest in the application of technology for advancing human learning. But Learning Sciences is founded on theories and methods from

Cognitive Science and Psychology (e.g., information processing models of learning; constructionist models of learning) whereas Instructional Systems Design is founded on theories and methods from design (instructionist models of learning). Kirby et al. (2005) showed that although there was a trend for increased cross-field citation, this trend was led by a small number of prominent researchers in both communities. The authors argued that combining the strengths of both communities—cognition in context on the one hand (Learning Sciences) and design (Instructional Systems Design) on the other—would give researchers a better chance of "effecting meaningful change in education through the creation and effective application of technology-enhanced learning environments" (op. cit., p. 46).

Specific Foci Lead to Researcher Isolation

A fourth example of missed opportunity around communication is when researchers who focus on particular aspects of a topic (e.g., human behavior) remain isolated from researchers who focus on others aspects. For example, in behavioral research in psychology, factors influencing dispositions of individuals are seen as more important than questions about the variation across differently situated populations (Longino, 2013). It follows that the set of references cited by a paper is very specific to the different disciplines and communities of readership for which the research is meant (op. cit.), despite the interest of combining approaches to gain broader understanding of the topic.

Same Broad Focus but Different Choices of Explanada

A fifth example of missed opportunity around communication is when researchers focus on the same phenomenon, but give attention to different explanatory aspects of it; an explanandum/explananda is the object(s) to be explained (Hempel and Oppenheim, 1948). Chevrot and Foulkes (2013) have noted the approaches of two fields in the study of language variation: Cognitive Sociolinguistics and Sociolinguistic Cognition. The first field explores cross-linguistic variation linked to social dimensions (Kristiansen and Dirven, 2008) and views linguistic knowledge and patterns of thought as properties shared by communities. These communities are seen as heterogeneous and are described at sociological, cultural, ideological or political levels thus sharing assumptions with sociolinguistics and employing a social approach to cognition (Kaufmann and Clément, 2011). The second field focuses on individual cognitive and cerebral mechanisms that underpin a person's ability to produce sociolinguistic variation and to process it during perception (Chevrot and Foulkes, 2013). For example, efforts include understanding the cognitive encoding of variants, describing the influence of social knowledge and its retrieval on the processing of variation, and studying the cerebral mechanisms that process indexical information. These authors argue that Sociolinguistic Cognition is close to psycholinguistics and represents the approach that Kaufmann and Clément (2011) call a cognitive approach to social facts. Indeed, it would seem that both Cognitive Sociolinguistics and Sociolinguistic Cognition should be considered in a relationship of reciprocal causality, in that knowledge about social life in relation to language and how we produce and understand language are mutually influencing each other.

Mobilizing Expertise at the Boundaries of Disciplines or Frameworks

The five examples of missed opportunities having to do with communication deficiencies discussed above also have another characteristic in common; they illustrate a deep expertise at boundaries between disciplines or frameworks. Such expertise includes building theoretical constructs, choosing which application domain to apply research, deciding on what part of analytic objects to focus, deciding which aspect of a topic to pay attention to, and choosing to focus on different explanatory aspects of the same phenomenon. Gaining this expertise takes time and effort.

In this final results section, inspired by Klein (1990) and deWachter (1982), we propose a slightly modified model (see **Figure 6**) that is set up to meet the challenges detailed in the previous section regarding broader education-related collective knowledge advancement. It is therefore a prospective model.

The interdisciplinary researcher mobilizes multiple theories that come from different disciplines and from different frameworks within the same discipline, according to analytic need. Each theory allows for observations that are oriented in particular ways and methods of interpretation that are based on specific assumptions. It is important to note that research questions should drive the choice of theories and methods, but that if researchers put the cart before the proverbial horse by choosing a particular theory and methods before thinking about research questions, this will seriously limit possibilities. In the same way, if a researcher only focuses on the specific theories and original methods with which they were trained, then answers will be limited in scope (Lund et al., 2013). Working from within bioethics, deWachter (op. cit.) proposes five stages to the interdisciplinary process, the foundation of which is a temporary suspension of all known methods, which he calls *methodical epoché:*

1. Accept that all involved disciplines abstain from approaching the topic using their own monodisciplinary methods;
2. Formalize the global question in an interdisciplinary way, acknowledging all of its aspects, as well as the total network;
3. Translate the global question into the specific language of each participating discipline;
4. Check the answer to this translated question to verify its relevance to the global question;
5. Agree upon a global answer which is not produced by any one disciplines, but integrates all particular available answers.

In **Figure 6**, we suggest that in each of the cases of missed opportunities around communication, researchers can follow this process model of interdisciplinarity in order to reach broader education-related collective knowledge advancement. Our proposed contribution to the model appears in phase 3 where researchers translate the global question into the specific language of each participating discipline. Finding out how to do this for different questions can be achieved by exploring the

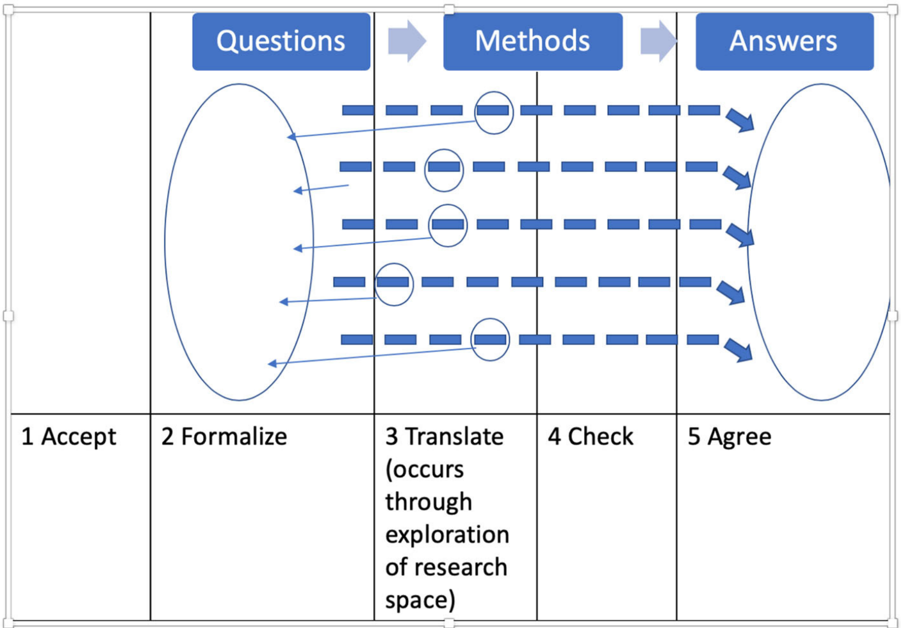

FIGURE 6 | Inspired by deWachter (1982) and Klein (1990), an abstract model of the interdisciplinary process redrawn to integrate exploration of the maps of the research space of education within social sciences and humanities.

bibliometric coupling maps and the analyses thereof that we provide. One can explore the maps we used to generate the clusters in many other ways, by searching for most frequent keywords, most frequent authors' keywords, most frequent title words, most frequent subject categories, most frequent journal sources, most frequent countries, most frequent institution, most frequent references, most frequent references' sources, most representative papers, most cited papers, or most cited authors. For a given sub topic cluster, one can see the collective answers to all of these searches and be pointed to both papers and authors on google scholar. Such a target search tool specifically scaffolds the reformulation of questions in terms of different frameworks and disciplines.

In summary, we propose that this model be used, coupled with exploration of our maps, to meet the challenges that we previously defined and that in general require mobilizing expertise at the boundaries of disciplines or frameworks, a phenomenon that our analyses show, both at the Education corpus level and at the Social Sciences and Humanities corpus level. We argue that — given the will — operationalizing this proposal will help researchers to connect who have the same broad focus but different choices of explanada. It may also alleviate researcher isolation due to specific foci, if bridges can be built to other foci in order to gain a broader understanding. Given that a divided field leads to pursuing different objectives, touching base with other community branches can help a larger field to integrate results and perspectives. Finally, foundational differences that hinder collaboration can be specifically addressed, but again, given the will of the involved researchers.

CONCLUSION

In this paper, we set out to understand the nature of education research and how it is connected through citations to the larger social sciences and humanities. We answered four research questions regarding (1) the place of research in education in this larger context, (2) how research in education compares to the disciplinary composition of other fields, (3) what such a composition can tell us about the challenges that research in education faces, and (4) what kind of model for broader education-related collective knowledge advancement can meet these challenges.

We argued that education cannot be considered to be a single discipline and is instead very multidisciplinary. We illustrated this in multiple ways. First, bibliographic coupling shows that research in education exists on a continuum, from highly connected, cohesive clusters of articles on a particular topic to loosely connected, less cohesive clusters. The former focus more directly on learning and teaching and specifically address science whereas the latter reach out to different subject areas such as linguistics, sociology, and psychology, with education being only one object of inquiry, amongst others. Secondly, this result is mirrored at the social sciences and humanities level in that research in education is distributed over many subject area regions, corresponding to the loosely connected, less cohesive clusters of the Education corpus. That said, we also see a focus on Science at this level because a large percentage of the education related articles within the social sciences and humanities sub topic cluster "Science" are published in Education corpus journals and a large percentage of these are in the

"Science Education" sub topic cluster in the Education corpus. Thirdly, articles having to do with education are more likely to connect to articles in other Scopus subject areas (in particular, Psychology) than articles of other Scopus subject areas. Fourthly, in qualifying Education related research first gathers articles essentially equally from both Education and Psychology and then next from Medicine subject areas. Fifthly, the Psychology subject area has the least breadth in terms of linked articles, but the most cohesiveness whereas the Education subject area has the most breadth and the least cohesiveness.

All of these measures point to the Education subject area as the most multidisciplinary both in terms of being less cohesive (lack of closely linked neighboring articles) and in terms of linking widely to different subject areas. A lack of cohesiveness may seem negative, but it is not necessarily a value judgement. And linking widely to different subject areas may seem positive in the sense of illustrating interest in theories and methods from other subject areas, and thereby exhibiting openness, but it may also be the curse of diversity we referred to in the introduction. If participating researchers have difficulty communicating due to no shared vision or common understanding of the literature, there will be dispersion of discourse and an unstable epistemic community. This seems to be the case today in research that is related to education, in light of its wide spread multidisciplinarity. We gave examples of the challenges that needed to be met and examples of the dispersion of discourse in education research. We then proposed a method based on the exploration of the Education and Social Sciences and Humanities corpora that can reduce dispersion, while building on expertise, and contribute to achieving broader education-related collective knowledge advancements, despite an unstable epistemic community.

LIMITATIONS AND PERSPECTIVES

The main limitation of our work is that although we were able to compare publications in education from two periods (2000–2004 and 2015–2018), we were only able to compare the first period to publications in social sciences and humanities to a year within that period (i.e., 2000). We were not able to compare the second period to publications in social sciences and humanities for the same period, due to the explosion in publication numbers rendering our method unworkable. Our results from 2000 to 2004 illustrate a snapshot in time before an initial major increase in new journals being indexed, so it is both a time before such a comparison may drastically change, and it's a way to keep the number of bibliographic records manageable. However, given the changes in the two Education corpora (2000–2004 and 2015–2018), coupled with more current literature that discusses the broader state of research in the social sciences and humanities, we hypothesized that the pressing need for methods to scaffold integration between disciplines and frameworks has not let up.

The second limitation is that articles from Scopus only give a partial view of the research in education, Scopus essentially only gives an anglophone view. Similar work using google scholar would be more comprehensive, but meta data is lacking. In addition to continuing to use our maps for the study of fine-tuning research questions from an interdisciplinary standpoint, perspectives include using them for other activities, such as training early career researchers, finding partners for collaboration, co-developing primers with practitioners that target conceptual constructs that are useful for teaching, and pinpointing gaps in the literature and thus opportunities for new research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: http://sebastian-grauwin.com/XYZ_EDUCMAP/.

AUTHOR CONTRIBUTIONS

PJ and SG provided the knowledge and methods regarding bibliometric coupling. SG built the algorithms, produced **Figures 1–4**, **Tables 1–4**, and also constructed the website where the underlying data can be visualized. PJ built the algorithms, calculated the data, and produced **Figure 5**. HJ and KL provided the expertise needed to interpret the data in terms of research on education. KL contributed the challenges and benefits of multidisciplinarity, and the connection with deWachter's model, (**Figure 6**) the bibliometric maps and their interpretation to aid researchers to use interdisciplinarity in order to alleviate missed opportunities. All authors worked together weekly for a number of months, discussing how to frame the paper and re-read, and made extensive comments on the manuscript. All authors approved the submitted version.

FUNDING

The authors would like to acknowledge support in 2014 and 2015 from the Programme Exploratoire Premier Soutien (PEPS) COMUE à Projets Interdisciplinaires Université de Lyon – CNRS. In addition, the authors are grateful to the ASLAN project (ANR-10-LABX-0081) of Université de Lyon, for its financial support within the program Investissements d'Avenir (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). This work was also supported by the National Research Foundation of Korea (Grant No. NRF-2016R1D1A1B03935697, 2016).

ACKNOWLEDGMENTS

The authors would like to acknowledge colleagues and students who have participated in various workshops with us on the EducMap project, among them Sanne F. Akkerman, Christine Detrez, Stephen M. Fiore, Judith Harackiewicz, Barbara K. Hofer, Emmanuelle Picard, and Romane Rozencwajg.

REFERENCES

AERES website: http://www.aeres-evaluation.fr/ AERES list of publication sources in educational research (2014). Available online at: http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-de-l-AERES

Apostel, L., and Centre for Educational Research and Innovation, Organisation for Economic Co-operation and Development (1972). *Interdisciplinarity Problems of Teaching and Research in Universities*. (Paris: Organization for Economic Cooperation and Development), 75–139.

Avis, J. (2003). Re-thinking trust in a performative culture: the case of education. *J. Educ. Polut*. 18, 315–332.

Birnbaum, P. H. (1981). Academic interdisciplinary research: characteristics of successful projects. *J. Soc. Res. Adminis.* 13, 5–16.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008:P10008. doi: 10.1088/1742-5468/2008/10/P10008

Chevrot, J.-P., and Foulkes, P. (2013). Introduction: language acquisition and sociolinguistic variation. *Linguistics* 51, 251–254. doi: 10.1515/ling-2013-0010

Choi, B. C. K., and Pak, A. W. P. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clin. Invest. Med.* 29, 351–364. doi: 10.25011/cim.v30i6.2950

Cotterell, R. (1979). Interdisciplinarity: the expansion of knowledge. *Higher Educ. Rev*. 11, 47–56.

deWachter, M. (1982). Interdisciplinary bioethics: but where do we start? A reflection on Epoché as Method. *J. Med. Philos*. 7, 275–287.

Epton, S. R., Payne, R. L., and Pearson, A. W. *(eds.)*. (1983). *Managing Interdisciplinary Research*. (New York, NY: Wiley).

Fortunato, S. (2010). Community detection in graph. *Phys. Rep.* 486, 75–174. doi: 10.1016/j.physrep.2009.11.002

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., and Trow, M. (1994). *The New Production of Knowledge*. (London: Sage).

Goggins, S., Jahnke, I., and Wulf, V. (eds.). (2013). "CSCL@Work: case studies of collaborative learning at work," in *Computer Supported Collaborative Learning Series, Vol*. 14, eds C. Hoadley, and N. Miyake (New York, NY: Springer), 243–268. doi: 10.1007/978-1-4614-1740-8

Good, B. H., Montjoye, Y.-A., and Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Phys. Rev. E* 81:046106204010. doi: 10.1103/PhysRevE.81.046106

Grauwin, S., Beslon, G., Fleury, E., Franceschelli, S., Robardet, C., Rouquier, J. B., et al. (2012). Complex systems: dreams of universality, reality of interdisciplinarity. *J. Am. Soc. Inform. Sci. Tech.* 63, 1327–1338. doi: 10.1002/asi.22644

Grauwin, S., and Jensen, P. (2011). Mapping scientific institutions. *Scientometrics* 89, 943–954. doi: 10.1007/s11192-011-0482-y

Hall, K. L., Vogel, A. L., Stipelman, B. A., Stokols, D., Morgan, G., and Gehlert, S. (2012). A four-phase model of transdisciplinary team-based research: goals, team processes, and strategies. *Transl. Behav. Med.* 2, 415–430. doi: 10.1007/s13142-012-0167-y

Hanisch, T. E., and Vollman, W. (eds.). (1983). *Interdisciplinarity in Higher Education, United Nations Educational, Scientific, and Cultural Organization*. Bucharest. European Centre for Higher Education.

Hausman, C. R. (1979). "Introduction: disciplinarity or interdisciplinarity?," in *Interdisciplinarity and Higher Education*, ed J. Kockelmans (University Park, PA: The Pennsylvania State UP), 1–10.

Hempel, C. G., and Oppenheim, P. (1948). Studies in the logic of explanation. *Philos. Sci.* 15, 567–579. doi: 10.1086/286983

Hermeren, G. (1986). "Interdisciplinarity revisited promises and problems," in *Interdisciplinarity Revisited: Re-assessing the Concept in Light of Institutional Experience*, eds L. Levin, and I. Lind (Stockholm: Organisation for Economic Cooperation and Development, Swedish Board of Universities and Colleges), 15–25.

Jäger, J. (2007). "*Foreword*," in *Handbook of Transdiscplinary Research*, eds G. Hirsch Hadorn, H. Hoggmann-Riem, S. Biber-Klemm, W. Grossenbacher-Mansuy, D. Joye, C. Pohl, et al. (Bern: Springer), vii–viii.

Jensen, P., and Lutkouskaya, K. (2014). The many dimensions of laboratories' interdisciplinarity. *Scientometrics*. 98, 619–631. doi: 10.1007/s11192-013-1129-y

Kaufmann, L., and Clément, F. (2011). "L'esprit des sociétés bilan et perspectives en sociologie cognitive," in *La Sociologie Cognitive*, eds F. Clément, and L. Kaufmann (Paris: Orphys et Maison des Sciences de l'Homme), 7–40.

Kessler, M. M. (1963). Bibliographic coupling between scientific articles. *Am. Document*. 24, 123–131.

Kirby, J. A., Hoadley, C. M., and Carr-Chellman, A. A. (2005). Instructional systems design and the learning sciences: a citation analysis. *Educ. Technol. Res. Dev*. 53, 37–48. doi: 10.1007/BF02504856

Klein, J. T. (1990). *Interdisciplinarity: History, Theory, and Practice*. (Detroit, MI: Wayne State University).

Klein, J. T. (1996). *Crossing Boundaries: Knowledge, Disciplinarities, and Interdisciplinarities*. (Charlottesville: University of Virginia Press).

Kocklemans, J. (1979). "Why interdisciplinarity?" in *Interdisciplinarity and Higher Education*, ed J. Kocklemans, (University Park, PA: Pennsylvania State UP), 123–160.

Kristiansen, G., and Dirven, R. (eds.). (2008). *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. (Berlin: Mouton de Gruyter).

Kuhn, T. S. (1962/1970). *The Structure of Scientific Revolutions (Edition 1970)*. (Chicago, IL: The University of Chicago Press).

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. (Oxford: Oxford University Press).

Lattuca, L. R. (2003). Creating interdisciplinarity: grounded definitions from College and University faculty. *Hist. Intell. Cult*. 3, 1–20.

Longino, H. E. (2013). *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. (Chicago, IL: University of Chicago Press).

Lund, K. (2016). *Modeling the Individual Within the Group: an Interdisciplinary Approach to Collaborative Knowledge Construction*. Education: Université Grenoble – Alpes.

Lund, K., Jeong, H., Grauwin, S., and Jensen, P. (2015). *A Scientometric Map of Global Educational Research*. Working paper.

Lund, K., Jeong, H., Grauwin, S., and Jensen, P. (2017). Une carte scientométrique de la recherche en éducation vue par la base de données internationales Scopus. *Les Sci. Éduc.* 50, 67–84. doi: 10.3917/lsdle.501.0067

Lund, K., Rosé, C. P., Suthers, D. D., and Baker, M. (2013). "Epistemological encounters in multivocal settings," in *Productive Multivocality in the Analysis of Group Interactions*, eds D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, N. *Law, C. Hoadley,* et al. (New York, NY: Springer), 659–682.

Miller, R. C. (1982). Varieties of interdisciplinary approaches in the social sciences: a 1981 overview. *Issues Integrat. Stud.* 1, 1–37.

Mingers, J., and Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *Eur. J. Operation. Res.* 246, 1–19. doi: 10.1016/j.ejor.2015.04.002

Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* 69:026113. doi: 10.1103/PhysRevE.69.026113

Newman, M. E. J. (2010). *Networks: An Introduction*. (Oxford: Oxford University Press).

Rosé, C. P., and Lund, K. (2013). "Methodological pathways for avoiding pitfalls in multivocality", in *Productive Multivocality in the Analysis of Group Interactions*, eds D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, and N. Law *(New York, NY: Springer),* 613–637.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Am. Soc. Inform. Sci.* 24, 265–269. doi: 10.1002/asi.4630240406

van den Besselaar, P., and Heimeriks, G. (2001). "Disciplinary, multidisciplinary, interdisciplinary: concepts and indicators," in *ISSI 2001, 8th International Conference of the Society for Scientometrics and Informetrics*, eds M. Davis and C.S. Wilson (Sydney, NSW*: UNSW*), 705–716.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 14 January 2021
doi: 10.3389/fpsyg.2020.579986

Benefits and Challenges of Interdisciplinarity in CSCL Research: A View From the Literature

Cindy E. Hmelo-Silver[1]* and Heisawn Jeong[2]

[1] Center for Research on Learning and Technology, School of Education, Indiana University Bloomington, Bloomington, IN, United States, [2] Department of Psychology, Hallym University, Chuncheon, South Korea

Computer-supported collaborative learning (CSCL) has a history of being interdisciplinary from its conception. Its beginnings have included computer scientists, psychologists, cognitive scientists, and educational researchers. These collaborations have been fruitful but have also posed challenges (Suthers et al., 2013). This article builds on the authors' extensive review of the CSCL literature to examine the nature of interdisciplinary collaboration in CSCL research as well as an interdisciplinary CSCL workshop. Using a corpus of more than 700 CSCL articles, we reported an updated analysis for the theories and methods used in CSCL research. In addition, bibliometric analyses examined journals that publish CSCL research and are cited by CSCL research. CSCL research is published in journals that are aligned with interdisciplinary research with large contributions from educational research followed by technology related fields and social sciences. The contributions from domain knowledge journals are relatively weak. These analyses revealed disciplinary influences and uptakes of CSCL research and how they might differ across CSCL research clusters. Lastly, we provide a case example of a CSCL workshop to further demonstrate the interdisciplinary nature of the field. Through these analyses we aim to characterize the benefits and challenges of interdisciplinary collaboration in CSCL research. Interdisciplinarity has helped CSCL research to adopt multiple theories and methods to understand CSCL. While cultivating diversity, we also need to be mindful that research outcomes are exchanged and appropriated actively across participating disciplines so that our understanding of CSCL rises above individual disciplines.

Keywords: computer-supported collaborative learning, interdisciplinarity, bibliometric analysis, systematic review, educational technology

OPEN ACCESS

Edited by:
Matthias Stadler,
Ludwig Maximilian University of Munich, Germany

Reviewed by:
Zachari Swiecki,
Monash University, Australia
Viktoria Pammer-Schindler,
Graz University of Technology, Austria

***Correspondence:**
Cindy E. Hmelo-Silver
chmelosi@indiana.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

***Received:** 03 July 2020*
***Accepted:** 04 December 2020*
***Published:** 14 January 2021*

Citation:
Hmelo-Silver CE and Jeong H (2021) Benefits and Challenges of Interdisciplinarity in CSCL Research: A View From the Literature. Front. Psychol. 11:579986. doi: 10.3389/fpsyg.2020.579986

INTRODUCTION

Computer-supported collaborative learning (CSCL) has a history of being interdisciplinary from its conception (Stahl et al., 2014). Its beginnings have included computer scientists, psychologists, cognitive scientists, and educational research disciplines. These collaborations have been fruitful but have also posed challenges (Suthers et al., 2013). This interdisciplinary nature of the CSCL research has been reflected in the diversity of theories and methodological frameworks used in CSCL (Jeong et al., 2014). In this article, we build on this research and attempt to examine and characterize the nature of interdisciplinarity in CSCL research. We begin by examining the historical roots of research traditions that comprise CSCL research. We will then examine the interdisciplinarity of CSCL research from multiple perspectives: (1) composition of CSCL research methods and theoretical frameworks, (2) bibliometric research clusters that emerged based on shared reference citations, (3) disciplinary associations of the journals that publishes CSCL research and is cited by CSCL research, (4) disciplinary affiliations/compositions of the contributing authors to International Journal of Computer-Supported Collaborative Learning (ijCSCL), and 5) a case example of an interdisciplinary CSCL workshop. These analyses rely on an updated corpus of CSCL literature that covers ten years of research between 2005 and 2014 and recent publications in the *ijCSCL*. We also relied on a range of analyses from content analysis, bibliometric analyses, and a qualitative case example. The case example moves from a bird's eye view of the field to a ground level description of how interdisciplinary collaboration results in new insights for the field. To set the context for the research questions and analyses that follow, we begin with a historical overview of the interdisciplinary beginnings of CSCL.

HISTORICAL OVERVIEW

The origin of the CSCL field dates back to the 1980s. Part of what makes this fundamentally interdisciplinary as a field is the relationship between the technology in the form of computational objects and the social interactions involved in learning (Stahl et al., 2014; Ludvigsen et al., in press). CSCL is the result of several converging forces. First, it was propelled by the research in developmental and social psychology and educational research that demonstrated that students working in pairs often performed better than those who worked alone (O'Donnell and O'Kelly, 1994; Miyake, 2007). These findings propelled researchers to examine underlying mechanisms of collaboration. Educators were also keen to develop instructional arrangements to promote the effects of collaborative work (Cohen, 1994; O'Donnell and King, 1999).

Another force that has contributed to the development of CSCL is the development of technology. It connected learners across geographical regions, enabling them to interact with learners and experts who are outside the geographical and temporal range of their social interactions. This interaction was mediated by a number of computational artifacts. A number of technologies and tools had been developed to help learners engage in collaborative sensemaking activities (Miyake, 2007). Other technologies provided opportunities for rich contexts that support collaboration (Roschelle, 1992; Goldman-Segall and Maxwell, 2002; Hmelo-Silver et al., 2016). Lastly, the integration of socio-cultural theory was critical. It provided a framework to incorporate both collaboration and tool mediation.

These forces created tensions and conflicts from the beginning. Some found the significance of CSCL in its epistemological underpinnings and argued that it initiated a new paradigm of learning research (Koschmann, 1996). In contrast, others took a more pragmatic approach and saw CSCL as a way to promote learning without necessarily signing up for its radical epistemological underpinnings. Educators, for example, saw that as an instructional intervention to promote cognitive learning outcomes both within and outside of the classrooms. Technology has provided ways to make classroom collaboration more engaging and meaningful. And yet to realize its promise, CSCL needs to build on both advanced technological innovation and deep understanding of how people learn. These multiple motivations and visions for CSCL have made CSCL research interdisciplinary and prompted CSCL to adopt a diversity of theoretical and methodological approaches (Jeong et al., 2014).

In this article we address two research questions:

(1) To what extent is CSCL an interdisciplinary research community?
(2) How can different sources of evidence be used to paint a picture of interdisciplinary collaboration?

METHODS

Article Selection and Screening

Much of the literature discussed is a secondary analysis of systematic reviews presented in earlier publications (Jeong et al., 2014, 2019a; McKeown et al., 2017) and builds on a corpus of CSCL literature collected for that purpose, the dates of that review being from 2005 to 2014. The corpora used for the systematic reviews of CSCL literature were constructed based on two databases, ERIC and Web of Science, in addition to seven key journals regarded by experts (Jeong et al., 2014) to be leaders in publishing CSCL research: *Computers and Education*, *Computers in Human Behavior*, *International Journal of Computer Supported Collaborative Learning*, *International Journal of Artificial Intelligence in Education*, *Journal of Computer Assisted Learning*, *Journal of Learning Sciences*, and *Learning and Instruction*. We screened over 1,600 articles published between 2005 and 2014 to ensure each article met the following criteria: (a) STEM education, (b) empirical research, and (c) use of technology to support collaborative learning. CSCL research refers to research articles in which participants learned collaboratively with the support of computers and/or other technologies. The technology also needed to be specific so that studies examining technology integration or adoption in general were not included. Studies about students with physical or learning disabilities were excluded because these can involve special technologies not typical in CSCL. Learners needed to interact in small groups or in some ways with peers at some point during the learning process. Studies needed to address learning, broadly defined (see Jeong et al., 2014 for additional details). We defined empirical research to refer to studies that relied on first-hand data to validate a theory, hypothesis, research question, and/or design. Although we broadly used the same criteria to screen and select articles for the corpus, changes in the research question over the years had led to the construction of a corpus with a slightly different scope and nature. In Jeong et al. (2014), we covered CSCL articles from 2005 to 2009 and included CSCL research both in STEM and non-STEM (n = 400). Funded by the United States National Science Foundation, McKeown et al. (2017) aimed to examine CSCL in STEM domains and thus focused on only STEM CSCL research, but expanded the corpus up to 2014 to include ten years of research (n = 735). Jeong et al. (2019b) combined the two corpora for bibliometric analysis (n = 869), which is basically the articles in the corpus used by McKeown et al. (2017) with an addition of non-STEM articles in the earlier corpus by Jeong et al. (2014) that were excluded in the McKeown et al. (2017) corpus. It was done to ensure a large enough corpus for the bibliometric analysis, but made the corpus used for bibliometric analyses unbalanced because non-STEM articles were not present for the 2010–2014 period. These features of the corpora need to be taken into consideration when interpreting the results.

For the *International Journal of Computer-Supported Collaborative Learning* journal analysis, we selected multiple-authored ijCSCL articles from three time points which are 2006–2007, 2012–2013, and 2018–2019. These three 2-year periods were selected from the total number of complete issues of the journal which began its publication in 2006. This allowed examining historical trends over time and provided a view of the early issues and most recent with the 2012–2013 providing a midpoint.

Coding

The articles that met our initial criteria were coded based on several dimensions. In our earlier examination of CSCL research practices, we examined CSCL research methods in terms of research design, research settings, data sources, and analysis methods (Jeong et al., 2014). We additionally examined how these methodological practices related to theoretical frameworks of the research. Theoretical frameworks referred to perspectives that guided the research (Danish and Gresalfi, 2018). The initial list of frameworks was derived from keywords used for the CSCL 95 conference and then expanded based on the frameworks that were represented in thearticles. Information processing theory referred to traditional cognitive theories with a strong emphasis on individual cognitive processes. Socio-cognitive theory referred to theories related to constructs of cognitive conflict and conceptual change (DeLisi and Golbeck, 1999). Constructivism referred to a broad range of theoretical approaches that emphasize active learner processing and knowledge construction in individualistic and collaborative settings (von Glasersfeld, 1995; Chi and Wylie, 2014). Socio-cultural theory referred to a diverse range of theories such as Vygotskian approaches, distributed cognition, or activity theory that emphasizes the fundamental role of tools, activities, social norms and systems (Danish and Gresalfi, 2018). Communication theory referred to theories addressing linguistic and communicative aspects of collaboration (Krauss et al., 1990). Social psychology theory referred to theories that focused on social aspects of collaboration such as status difference, gender, and/or group dynamics (Levine and Thompson, 1996). Motivation theory referred to theories with a focus on motivational aspects of learning, addressing issues such as attribution or self-regulation (Pintrich, 1999). The Other theory category referred to theories that did not fit into any of the categories that we have described (e.g., constructionism). Studies coded as Atheoretical referred to investigations that were primarily guided by practical concerns (e.g., program evaluations). Boundaries of different theoretical frameworks were not always clear-cut. If authors explicitly named their theoretical frameworks, we coded them as such. If they were not, we relied on references and major variables examined in the study (e.g., conceptual change is a typical variable or topic of study strongly associated constructivism). Studies could have more than one theoretical framework. Methodological practices refer to research design approaches (e.g., experimental, descriptive, and design-based research) and analysis methods (e.g., qualitative, quantitative, mixed methods; see Jeong et al., 2014; Hmelo-Silver and Jeong, in press, for further details). Two raters independently coded 20% of the sample with an overall kappa of.68, showing substantial agreement.

For the case study of the workshop, the first author went through the list of participants and the workshop report, identifying the academic disciplines of each of the participants based on their academic departments. Additional information regarding the workshop is drawn from the workshop report (Hmelo-Silver, 2019).

THEORIES AND METHODS USED IN CSCL

The different methods and the need to incorporate them is part of what makes CSCL a multidisciplinary field (Stahl et al., 2014). Analysis of methodological traditions in the field demonstrate that this has been the case from the beginning (Jeong et al., 2014). We coded these features of research for five years of CSCL research from 2005 to 2009 and found that the overall CSCL research practices are quite diverse, likely to reflect the diverse traditions that contributed to the formation of CSCL. We also found that this led to the use of research methods that are quite eclectic. For example, experimental work in classroom or online settings and wide usage of mixed studies. These trends were observed widely regardless of research traditions, but there was a clear alignment between research methods and theoretical frameworks. According to Jeong et al. (2014), four clusters of research emerged. Two of them were clearly guided by theoretical frameworks such as sociocultural and constructivists perspectives. While these traditions mostly relied on descriptive designs in classroom settings, there is a small cluster of CSCL research that strongly relies on experimental approaches. An updated analysis described in Hmelo-Silver and Jeong, in press) shows that the trends to use diverse methodological framework continued throughout the expanded time period (2005–2014), with a mix of methods drawn from psychology, linguistics, anthropology, and human–computer interaction.

Table 1, an updated table of theoretical frameworks that includes publications through 2014 (in STEM domains) indicates that articles with multiple theoretical frameworks account for 30% of the articles in the corpus. The largest overlap was among

TABLE 1 | Co-occurrence of theoretical frameworks.

	Single	Multiple
Information processing	33	17
Socio-cognitive	22	22
Constructivism	188	71
Sociocultural	109	54
Communication	19	19
Social Psych	44	24
Motivation	26	18
Other	53	24
Atheoretical	77	1
Total	571	250
	69.55%	30.45%

articles coded as constructivism and those coded as sociocultural (n = 26). Thus, the field continues to use diverse methodological frameworks in CSCL but since the earlier analysis, more individual articles use multiple frameworks.

The presence of diverse theoretical and methodological frameworks confirms that different disciplines contribute to CSCL research. The co-existence of these frameworks is reflective of the diverse research traditions that converge on CSCL and the interdisciplinarity of CSCL, but it is only a small piece of the picture. Another way to take advantage of this corpus to examine the (interdisciplinary) nature of CSCL as a field is through an analysis of the bibliometric data.

NATURE OF CSCL INTERDISCIPLINARITY: BIBLIOMETRIC ANALYSIS

Bibliometrics or scientometrics in particular analyzes scientific publications to measure and understand scientific research practices. It relies on citation or other statistical data related to academic publications. The development of digital technology and large databases such as Web of Science (WOS) and Scopus has contributed to its recent rise. It is increasingly used to understand questions such as the impact of specific research fields, a set of researchers or particular publications that connects different research fields, and/or publications with large impacts (Mingers and Leydesdorff, 2015). Bibliographic coupling (BC) analysis is a kind of bibliometric analysis that analyzes references of publications and identifies clusters of articles with shared references. This technique has been used to successfully map the networks of researchers in scientific institutions or in a given research field (Grauwin and Jensen, 2011; Grauwin et al., 2012).

Jeong et al. (2019b) have applied BC analysis to understand how CSCL research publications are organized. Using the extended corpus of CSCL articles (n = 869), they identified clusters of CSCL research that are linked by shared references. This is the expanded corpus mentioned in the previous section that included non-STEM articles to capture CSCL research clusters more widely. The BC analysis creates links between articles when they shared references (Kessler, 1963). A community detection algorithm based on modularity optimization (an implementation of the Louvain algorithm) was then applied to partition networks of linked articles into clusters in a map, in which a node represents a cluster with its size proportional to the quantity of articles within the clusters. Note that not all articles shared references with other articles. Clusters were not always connected to the rest of the clusters. In the end, 735 articles were included in the BC map shown in **Figure 1**. The rest of them (n = 134) did not share references with other CSCL articles, suggesting that there is some research that we have classified as CSCL that does not build on this literature and may draw on other research foundations.

The cluster labels were derived automatically based on the most frequently used keywords. Each cluster represents sub-areas of CSCL research that are linked by distinct sets of shared references, suggesting that they are referencing different knowledge bases in CSCL research. Keywords frequently used by the articles within the clusters are used as cluster labels. The "knowledge building" cluster in **Figure 1** means that articles in the clusters are likely related to research relevant to knowledge building in some way. The clusters also differ in size, suggesting that some topics have been the subject of more published research than others. As **Figure 1** shows, more research has been published on knowledge building (n = 145) and argumentation (n = 127) than topics such as peer assessment (n = 13) and gross anatomy education (n = 7). Jeong et al. (2019b) identified the five biggest clusters as major and the rest as minor CSCL research clusters. The five major clusters represent major areas of CSCL research such as knowledge building and argumentation, whereas the five minor clusters represent less well represented areas of CSCL research such as peer assessment and gross anatomy education. They differed in the references they share as well as in the publications sources in which their references were published.

References can reveal the intellectual traditions and disciplinary knowledge base that CSCL research draws upon. Articles in the same disciplines or sub-areas tend to cite similar publications. Another important marker of disciplinary association is the journals in which the article is published. Journals are outlets of academic research conducted in a particular field of research. They serve as gatekeepers of research and decide whether a particular piece of research is appropriate to their mission in terms of topics as well as quality (Crane, 1967). In this section, we examined the sources of research that CSCL cites and outlets of research that CSCL publishes to understand the disciplinary influences and composition of CSCL research and whether and/or how this reliance on particular disciplines may differ across sub areas of CSCL research. The historical disciplinary influences might still be visible to some degree as we have witnessed in the different theoretical approaches and methods.

To examine these questions, we extracted the following information from the articles (n = 735) included in the ten CSCL clusters: (1) authors, (2) year of publication, and (3) publication source (i.e., journals in which the article is published) (4) reference sources in the reference list (e.g., books, journals, and other sources that the article cites), (5) discipline categories

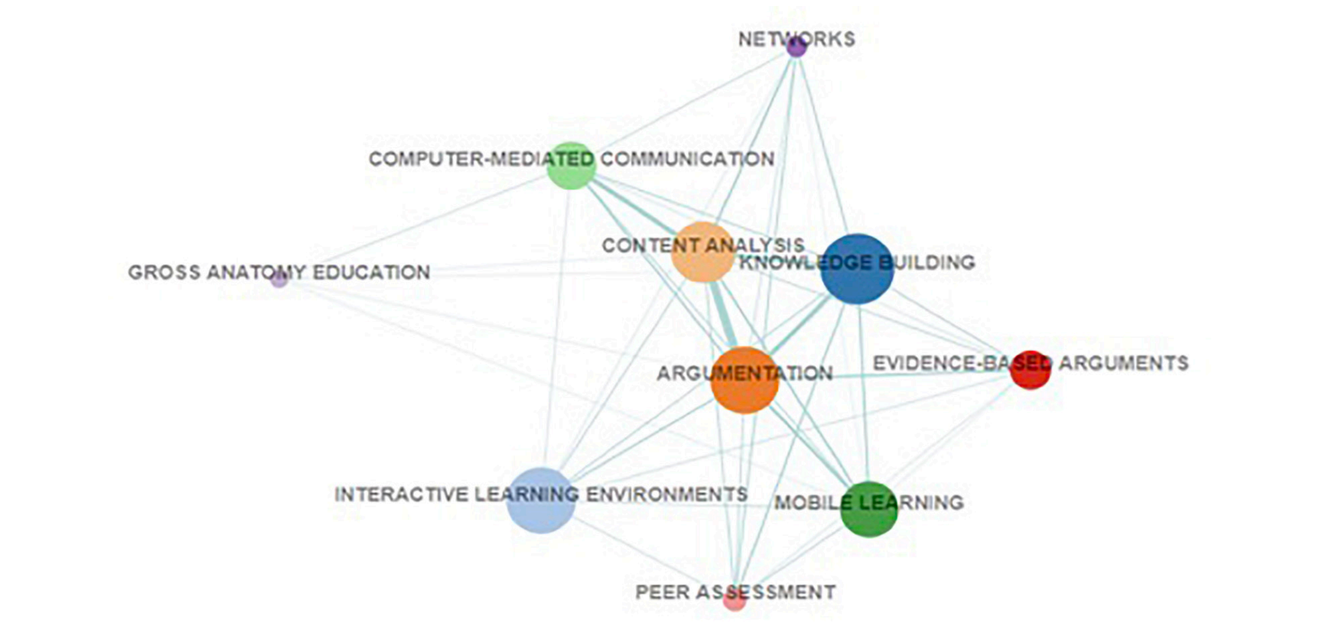

FIGURE 1 | Ten CSCL BC clusters (cluster labels represent most frequently used keywords by the articles in the clusters).

assigned by WOS called "WOS categories." We extracted the meta-data of the indexed articles from WOS, but hand coded the meta-data using the pdf files for the articles that were not indexed in WOS.

Journals That Publish CSCL

Computer-supported collaborative learning journals refer to those journals that are the sources of CSCL research articles reviewed here. We identified 33 such journals based on the most frequent publication sources across the ten CSCL research clusters. **Table 2** lists the top ten journals that published CSCL research during this period. They are the major outlets for CSCL research. Different journals publish different numbers of issues and articles each year. Journals with higher numbers and percentages are likely to be those journals that publish more issues and articles over the years.

We first examined the "aims and scope" statement of these journals as listed on the journal homepage to understand the disciplinary or interdisciplinary associations of the journals as identified by the editorial teams of the journals. These statements anchor the positions and directions of the journals and can serve as important guidelines for both authors and readers of the journals. We did not engage in formal coding, but looked for words or phrases that signaled associations with specific disciplines or interdisciplinary research. Most of the journals emphasize problems or research topics (e.g., "application of AI to education") rather than disciplinary associations. *Computers and Education*, for example, states that it welcomes research articles on the "pedagogical uses of digital technology, where the focus is broad enough to be of interest to a wider education community." Such emphasis on research problems and topics are indicative of the openness to approaches coming from different disciplines. Some journals go a step further and are explicit about this. *ijCSCL*, for example, states that it "aims to serve as a forum for a diverse range of disciplines such as education, computer science, information technology, psychology, communications, linguistics, anthropology, sociology, and business." Not all journals have multidisciplinary orientations, however. *Computers in Human Behavior* (CHB), for example, was clear that this journal is "dedicated to examining the use of computers from a psychological perspective." This disciplinary focus was more likely to be the case in journals focused on science and anatomy education. Still, such explicit mono-disciplinary association is an exception rather than a rule. In sum, it appears that most CSCL articles are published in journals that explicitly promote multidisciplinary approaches or emphasize research problems rather than specific disciplinary approaches.

Every journal or book indexed in WOS is assigned to at least one subject area category such as education or psychology. There are 256 WOS subject categories as of 2018[1]. WOS categories are quite detailed. There are three WOS categories for education, for example: "Education, Educational Research," "Education, Scientific Disciplines" and "Education, Special". Psychology has 11 WOS categories such as psychology, experimental, social, and so on. In order to examine the disciplines at a broad level more used in everyday discussion of disciplines, we grouped the WOS category of CSCL journals into four discipline groups: (1) Education (2) Technology (3) Social Sciences and Psychology, and (4) Knowledge Domains. **Table 3** shows how our discipline groups map onto the WOS categories with some example

[1]https://images.webofknowledge.com/images/help/WOS/hp_subject_category_terms_tasca.htm

TABLE 2 | List of top ten journals publishing CSCL research.

	Journals	Aims and Scope	*N* (%)
1	Computers and Education (C&E)	"Pedagogical uses of digital technology, where the focus is broad enough to be of interest to a wider education community"	333 (38%)
2	Computers in Human Behavior (CHB)	"dedicated to examining the use of computers from a psychological perspective"	133 (15%)
3	Journal of Computer-Assisted Learning (jCAL)	"…covers the whole range of uses of information and communication technology to support learning and knowledge exchange."	90 (10%)
4	International Journal of Computer-Supported Collaborative Learning (ijCSCL)	"A forum for a diverse range of disciplines such as education, computer science, information technology, psychology, communications, linguistics, anthropology, sociology, and business"	87 (10%)
5	Journal of the Learning Sciences (JLS)	"A multidisciplinary forum for research in education and learning…"	28 (3%)
6	Learning and Instruction (L&I)	"As an international, multi-disciplinary, peer-refereed journal,…a platform for the publication of the most advanced scientific research in the areas of learning, development, instruction and teaching"	25 (3%)
7	International Journal of Artificial Intelligence in Education (ijAIED)	"…publishes articles concerned with the application of AI to education"	24 (3%)
8	Educational Technology and Society (ET&S)	"…publishes the research that well bridges the pedagogy and practice in advanced technology for evidence-based and meaningfully educational application"	10 (1%)
9	Journal of Science Education and Technology (JSET)	"An interdisciplinary forum for the publication…that address the intersection of science education and technology with implications for improving and enhancing science education at all levels across the world"	8 (1%)
10	Anatomical Sciences Education (ASE)	"An international forum for evidence-based exchange of ideas, opinions, innovations, and research on topics related to education in the anatomical sciences of gross anatomy, embryology, histology, neurosciences, biomedical, and life sciences."	8 (1%)

Percentage refers to base of total number of articles in the expanded corpus (n = 869).

TABLE 3 | Web of science (WOS) categories of journals.

Discipline groups	WOS subject category of Major CSCL journals	Example Journals
Education	Education and Educational Research Education, Scientific Disciplines	JCAL C&E*, ijCSCL*, JLS* IEEE Transactions on Education*
Technology	Computer Science, Interdisciplinary Applications Information Science and Library Science Computer Science, Hardware and Architecture Computer Science, Information Systems Computer Science, Software Engineering Computer Science, Theory and Methods	C&E*, ijCSCL* Computer Applications in Engineering Education* IEEE Transactions on Education*
Social Sciences and Psychology	Businesses Communication Management Psychology, Experimental Psychology, Multidisciplinary Psychology, Educational Sociology	American J of Sociology CHB Communication Research Organization Science JLS*, Learning and Instruction*
Knowledge Domains	Anatomy and Morphology Biology Engineering, Electrical and Electronic Engineering, Multidisciplinary Ergonomics Health Care Sciences and Services Medicine, General and Internal Physiology	Annals of Anatomy Behavior Information Technology Croatian Medical Journal American Biology Teacher* Computer Applications in Engineering Education*

"" Indicates journals listed in more than one disciplinary group.*

journals in each group. Two journals were not indexed in WOS and thus could not be assigned to a discipline group in **Table 3**, but the rest of the 31 journals were assigned to at least one discipline group.

One way to define the multidisciplinarity of a journal is to examine whether they belong to more than one discipline group. CSCL publishing journals often belong to more than one discipline group. For example, journals such as *Computers and Education* and ijCSCL both belong to the Education as well as the Technology discipline groups. Computer Applications in Engineering Education (CAEE) belongs to three discipline groups: Education, Technology, and Knowledge Domains (i.e., Engineering, Multidisciplinary). About one-third (12 out of 31) of the CSCL publications are multidisciplinary in this sense. They all belong to the Education group, but varied in their second disciplinary association.

The number (and percentage) of articles in each discipline group is presented in **Table 4**. These numbers should be interpreted cautiously as journals often belong to more than one discipline. Even so, **Table 4** shows that CSCL research is published most in journals associated with Education (81%), followed by Technology (55%), and Social Sciences (25%). Journals in the Knowledge Domains group do publish CSCL research, but only 2% of CSCL articles have been published in such journals. Considering that a quite sizable portion of CSCL research involves STEM education (Jeong et al., 2019a), this mismatch is puzzling. In spite of STEM domains dominating CSCL, CSCL may not be widely adopted as a useful pedagogical strategy and/or there might not be sufficient audiences for CSCL research in these journals. In addition, although a large number of articles are being published in technology domain journals, they are concentrated on three journals: *Computers and Education* (n = 333), *ijCSCL* (n = 87), and *Computer Applications in Engineering Education* (n = 1).

TABLE 4 | Discipline groups of the journals that publish CSCL research.

Discipline groups	*N* of journals	*N* (%) of articles
Education	24	568 (81%)
Technology	3	382 (55%)
Psychology and Social Sciences	5	176 (25%)
Knowledge Domains	10	14 (2%)

Technology disciplines are part of CSCL research, and yet again, CSCL research is not being published as widely in technology journals. This, however, may be an artifact of the corpus, since one of the criteria to be included in the corpus was that the articles need to be empirical articles (Jeong et al., 2014). Articles that focus purely on technical or design aspects of CSCL tools were not likely to be included in the corpus. A similar pattern can be observed in Social Sciences domain journals. Most of the articles were published in five journals, most of which publish psychology or educational psychology research as Computers in Human Behavior ($n = 133$) and JLS ($n = 28$). In sum, while CSCL research is multidisciplinary in its historical origin and participating members' disciplines, CSCL research may not be relevant to participating disciplines to the same extent. The main audiences for CSCL research are readers of education journals or journals that are at the intersection of education, technology, and neighboring disciplines.

The disciplinary composition of CSCL publishing journals may vary depending on the nature of the research question. When the use of the tools in the classroom and appropriate pedagogical interventions are the focus, it is more likely to be relevant to educational researchers and journals that publish such research. **Figure 2** presents the proportion of articles published in each discipline group across the ten CSCL research clusters. Clusters are ordered from the biggest on the left to the smallest in size on the right in the figure. The proportion of each discipline group fluctuates across the clusters, but educational journals play the biggest role in publishing CSCL research, followed by technology journals and then by psychology and disciplinary education journals, replicating the general trend that we observed in **Table 4**. A few deviations from this general trend are notable, however. Knowledge domain journals have a larger presence in clusters such as gross-anatomy education (14.29%) and evidence-based arguments (12.20%) clusters compared with the other clusters. The cluster with the highest proportion of Knowledge Domains journals is the gross anatomy education cluster which is the smallest in size along with a narrow research focus. The cluster with the second highest proportion of knowledge domain journal articles is the evidence-based arguments cluster (12.2%) which is also relatively small in size and indicates a narrower research focus, a specific sub-type of argumentation. Taken together, it appears that the disciplinary composition of CSCL publishing journals is more or less the same across the ten CSCL research clusters, although a few deviate from it mainly due to the size and research topics of the clusters.

Journals That CSCL Research Articles Cite

The interdisciplinarity of CSCL research can also be examined based on the journals that it cites. There are 1,885 distinct reference sources cited by the CSCL research in the corpus, which include books and book chapters, but journal articles turn out to be major citing sources of CSCL research. **Table 5** below lists the top ten journals that articles in CSCL research cite. As shown in **Table 5**, *Computers and Education* is cited by about half of the CSCL articles; *Journal of the Learning Sciences* (JLS) and *Journal of Computer Assisted Learning* (JCAL) are cited by about one-third of the CSCL articles. *Computers and Education* (C&E) continue to be the top referenced journals as well as publication outlet.

Comparison between **Tables 2** and **5** shows that a group of journals such as C&E and JLS appear in both tables, indicating that they play an important role both as an outlet and reference source of CSCL research. At the same time, a group of journals emerged as a major reference source of CSCL research in **Table 5** although they did not appear in **Table 2**. ETR&D, Instructional Science, and Review of Educational Research fall into this category. In the case of Review of Educational Research, it publishes only review articles and thus is not likely to be a publication outlet for primary empirical articles included in our corpus. ETR&D and Instructional Science do publish CSCL research (ranked 9th and 19th in the publishing journal list), but did not appear in **Table 2** likely due to their low volume of publications. Yet there is another group of journals that appear in **Table 2**, but not in **Table 5**. For example, ijCSCL appears in **Table 2**, but

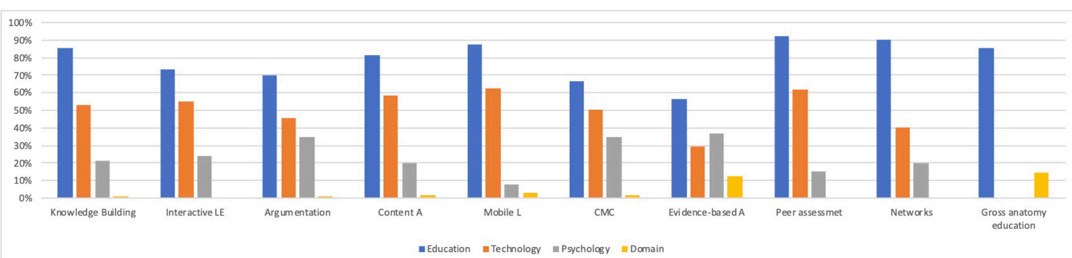

FIGURE 2 | Disciplinary composition of the publishing journals by clusters.

TABLE 5 | List of major journals that CSCL research cites.

	Journals	Frequencies (%)
1	Computers and Education (C&E)	468 (54%)
2	Journal of the Learning Sciences (JLS)	290 (33%)
3	Journal of Computer-Assisted Learning	280 (32%)
4	Computers in Human Behavior	265 (30%)
5	Learning and Instruction	233 (27%)
6	ETR&D-Educational Technology Research and Development*	218 (25%)
7	Instructional Science*	208 (24%)
8	Review of Educational Research*	205 (24%)
9	British Journal of Educational Technology*	193 (22%)
10	Journal of Educational Computing Research*	175 (20%)

"" Indicates journals that were not in **Table 2**.*
Frequencies and percentages refers to the percentages of the articles in the expanded corpus (n = 869).

not in **Table 5**. This is likely because it began publishing in 2006 and there is likely to be a time lag until researchers start reading and referencing articles from. ijAIED and ET&S are also journals that publish CSCL articles, but they are not referenced frequently in CSCL research. It may indicate uneven readership interests so that there is likely to be more interest in the application of technology to support CSCL in the AIED community, although computer science and technology articles may not be actively cited and referenced in the rest of the CSCL research.

Nonetheless, the disciplinary composition of the citing journals largely remains more or less the same as the disciplinary composition of the publishing journals (see **Table 6**). We analyzed the disciplines of the 39 citing journals included in the CSCL BC Map, excluding books or non WOS journals and journals cited by little CSCL research. Most of its citations are from education journals, indicating that CSCL research substantially builds on educational research. This does not mean that research from other disciplines does not contribute to CSCL research. A sizable proportion of the citation comes from journals in technology and/or social sciences journals as well as from journals in the knowledge domains, even though it is a proportionally small part of CSCL citations. The diverse historical origins of CSCL is visible from the disciplines of the citing journals, but knowledge uptake across disciplines appears to be uneven.

TABLE 6 | Discipline groups of the journals that CSCL research cites.

Discipline groups	*N* of journals	*N* (%) of citations
Education	25	2,019 (61%)
Technology	5	482 (15%)
Social Sciences	11	759 (23%)
Knowledge Domains	7	42 (1%)

We further examined the pattern of research uptake across CSCL research clusters. **Figure 3** presents the percentage of citations that journals in each discipline group received across the ten CSCL research clusters. Clusters vary in terms of disciplines of the journals the research they cite belong to. Most clusters draw on research from at least three discipline groups. Peer assessment and evidence-based clusters draws on all four discipline groups. Interactive Learning Environment and gross-anatomy education draws from two discipline groups. All clusters heavily cite research in education journals, but the extent of reliance varies. In the networks cluster, it relies on technology and social science journals more and in the gross anatomy cluster domain journals were equally cited. Taken together, educational research is the major knowledge base in all CSCL research clusters, but exact disciplinary composition varies somewhat depending on the clusters.

HOW INTERDISCIPLINARY IS *ijCSCL*?

Another marker of CSCL as an interdisciplinary field is through the composition of the journal devoted specifically to CSCL research. This includes the editorial team as well as the authors of articles in the journal. Academic societies often have flagship journals which members consider to be representative of the research that they do in the community. In the case of CSCL, it is the *International Journal of Computer-Supported Collaborative Learning* (*iJCSCL*). It was established in 2006 as the field was being established. There are several major journals that publish substantial amounts of CSCL research, but it has quickly established itself as a major outlet of CSCL research. This is remarkable when we consider that ijCSCL publishes far fewer articles per year than those journals. In this section, we examine the interdisciplinarity of the journal in terms of the disciplinary associations of its editorial team and contributing authors.

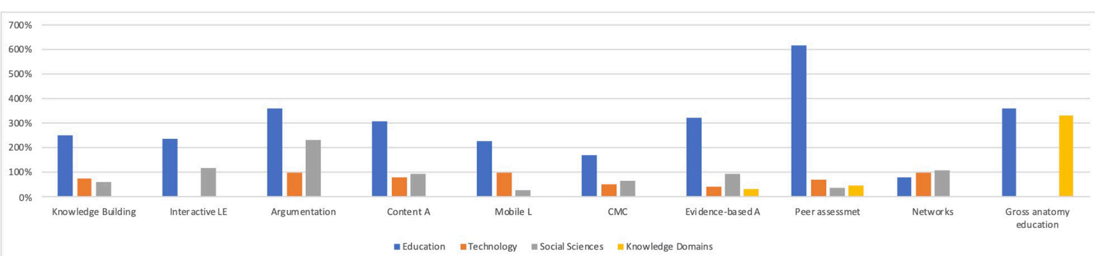

FIGURE 3 | Disciplinary composition of citing journals by clusters.

One example of the interdisciplinary nature of the journal is the editorial team of the *ijCSCL*. The founding co-editors included a computer scientist (Dr. Gerry Stahl) and a psychologist (Dr. Friedrich Hesse) with a similar composition among the newest co-editor team (Dr. Sanna Järvela, psychologist and Dr. Carolyn Rosé, computer scientist).

The interdisciplinary characteristics of ijCSCL can be inferred by looking at the range of the contributing authors' disciplinary associations in multi-authored ijCSCL articles. The number and percentages of the multi-disciplinary multi-authored articles are presented in **Table 7**. As this table shows, the percentage of articles (excluding editorials) has ranged from 14.29 to 29.42%, and including editorials has been roughly a third of the total multi-authored contributions. These interdisciplinary teams tend to be among social scientists (psychology and education), technology (computer and information sciences), and domain-specific (e.g., STEM departments and health sciences). These numbers are promising, but there is also a long way to go to promote more interdisciplinary collaboration that supports innovation in technology and sophisticated analysis of how the technology is a tool for CSCL supporting learning and engagement.

CASE EXAMPLE: BUILDING INTERDISCIPLINARY CAPACITY WORKSHOP

Although the journal citations and authorships provide some evidence of interdisciplinary collaboration, they may also underestimate the coherence in the community. Many workshops that try to solve CSCL problems are broadly interdisciplinary, a recognition that to build capacity in CSCL, a combination of technological, pedagogical and methodological approaches is needed. An example of this is the workshop organized by the first author (Hmelo-Silver, 2019). The workshop had an explicit goal of "Building Interdisciplinary Capacity for Understanding and Supporting Computer Supported Collaborative Learning." Although the particular team fluctuated over a series of four 1–2 day workshops, the regular contributors included 17 scholars who identified as education researchers and 13 who largely identified as computer scientists with some representation among industrial engineering and management sciences. This interdisciplinary group discussed actionable indicators in work on learning analytics and adaptive support for collaborative learning. Learning analytics work showed promise for informing collaborative learning but many of the indicators being used were shallow measures of participation and engagement. The next logical step for the workshops was to delve deeper and extract actionable indicators from research on collaboration, to determine whether or not supported by technology, they could be used to develop new technologies that would build models of collaboration that would be amenable to learning analytics, and ultimately lead to better adaptive support for collaborative learning, whether in stand-alone systems or to help teachers on a just-in-time basis. Many insights developed as behaviorally oriented researchers worked in small groups with more technically oriented researchers to identify what both saw as needed and interesting for driving research on CSCL forward. One outcome of this project was developing a common, shared language for talking about collaborative learning that can facilitate reporting and comparing research on collaborative learning as well as advancing joint research (e.g., Mott et al., 2019; Saleh et al., 2019; Chen et al., 2020). The workshop integrated across disciplines to show how CSCL researchers conceive of high-quality collaboration and indicators of lesser quality. This begins to provide a shared language to talk about aspects of collaboration that would be targets for automated analysis of collaboration, learning analytics, and adaptive support for collaborative learning. These discussions have led to further interdisciplinary collaboration towards just this end among learning sciences, instructional systems technology, and computer science in a team that is developing adaptive support for game-based learning (e.g., Mott et al., 2019; Saleh et al., 2020).

TABLE 7 | Interdisciplinary composition of *ijCSCL* multi-authored articles.

Year	Total # Articles including editorials	# Interdisciplinary	%	# excluding editorials	%
2006	24	7	29.17	3	14.29
2007	22	8	36.36	6	31.58
2012	27	6	22.22	5	21.74
2013	22	8	36.36	5	27.78
2018	22	7	31.82	4	22.22
2019	21	7	33.33	5	29.42

DISCUSSION

Educational research in general, and the learning sciences in particular is a multidisciplinary field that exists at the nexus of psychology, sociology, linguistics, anthropology, computer science, and technology (Lund et al., 2020; Pea and Linn, 2020). As an important branch of the learning sciences CSCL should be multidisciplinary as it needs to address educational, social, and psychological aspects of learning as well as technology designs and learning domains (e.g., disciplinary knowledge, skills, and practices) to be successful. This is reflected in the journals in which CSCL research is published and ways in which there are opportunities for interaction across disciplines. We have examined this interdisciplinarity through systematic review of the literature, bibliometric analyses, examination of editorial and authorship patterns in a major CSCL journal, and a case example from an interdisciplinary workshop. Together, these provide suggestions for ways that the CSCL research community works across disciplines, addresses interdisciplinary audiences, and where there is more that could be done.

Although CSCL began as a multidisciplinary endeavor with its research methodology and theoretical frameworks reflecting diverse traditions, the bibliometric analysis suggests that the main outlet and audience for CSCL research appears to be largely educational research journals, with some exceptions. In some sense, it is understandable. As a pedagogical strategy, there is a clear relevance to education. Still, the scarcity of journals that publish CSCL research devoted to disciplines other than education provide a barrier to CSCL research achieving its transdisciplinary goal. As the flagship journal of CSCL however, *ijCSCL* is a notable exception. Since its inception, this journal has had an international editorial team that includes computer scientists, learning scientists, educational psychologists, and discipline-based educational researchers (most notably in STEM education). This journal has generally had 20–30% of its articles composed of interdisciplinary collaborations. It is a venue that welcomes contributions from researchers across these domains (and a survey of the most recent volume of *ijCSCL* suggests at least one computer science contribution in each issue).

CSCL has been a collaboration between the technical and more socially oriented research fields. This research has appeared infrequently in journals that are dedicated to the teaching of specific disciplines such as STEM. These discipline-specific journals may be distributed across a range of fields and dilute the impact across any one field. We reported in our meta-analysis of CSCL that its effectiveness may vary depending on the learning domains and suggested that CSCL needs to be tailored to meet the needs of the knowledge domain (Jeong et al., 2019a). This may require active collaboration with disciplinary education researchers, and yet may not be well-reflected in authorship and journal outlets during the ten years of CSCL research that our corpus covers. Our analysis only covers active authorship whereas disciplinary expertise and collaboration may be reflected as contributions that are not authorship (e.g., as acknowledgments).

Nonetheless, we do see opportunities for collaboration. Many of the journals that authors publish in, the theories and research methods that draw from multiple fields, and the in-person interactions suggest that these interdisciplinary collaborations can and do occur with some regularity and are reflected in the diverse theoretical frameworks and research methods. From the early history of the field to the current journal editorship, the collaborations and contributions have been between computer scientists, educational technologists, and social scientists from the learning sciences, educational psychology, and other education fields. As a field, CSCL requires knowledge of design and pedagogy, technical expertise, classroom research strategies, and knowledge of multiple research methods. Collaborations between socially oriented researchers and technically oriented scholars can help bring more ambitious and forward thinking visions than either can alone. Computer scientists can help envision technical possibilities and advancements whereas social scientists can think about a pedagogical wish list but may not be able to envision what is technically possible.

The possibilities of these collaborations are exciting but also are challenging. Different disciplines have different standards for publication. Conference proceedings are more valued in technical areas (e.g., computer science) but less so in social sciences and education. We note that our analyses did not examine conference proceedings. In addition, the genre of research will be tied to particular disciplines (e.g., design and evaluation for computer science compared with empirical research in the learning sciences). University structures also tend to reward one publishing in one's own disciplinary field, providing further barriers and disincentives for cross-disciplinary work. However, in our analysis, there are clearly some high-impact journals at the intersection of disciplines. Bringing people together in workshops and face-to-face conferences is one way that these interactions have been promoted (Suthers et al., 2013; Hmelo-Silver, 2019). There are serious efforts underway to highlight and promote interdisciplinary work, most notable being the International Alliance to Advance Learning in the Digital Era[2] (IAALDE). This organization has promoted sharing research across disciplines that include behavioral, educational, and computer science fields. These organizations have committed to showcasing work across the disciplinary boundaries. This represents an interdisciplinary effort among leaders of these societies.

Computer-supported collaborative learning is a field with multidisciplinary foundations and origins (Hoadley, 2018). Through a range of analytic approaches, we have demonstrated the multidisciplinary theoretical and methodological foundations, the citations patterns, *ijCSCL* editorial and authorship collaborations, and workshop interactions to make an argument for ways in which there are influences to and from different disciplines and actual interactions among them. Although both social sciences and technical disciplines have been an important part of CSCL, the CSCL field has foundations it can build on for an even more interdisciplinary future.

AUTHOR CONTRIBUTIONS

CH-S worked on the sections on the systematic review, ijcscl analysis, and case example. HJ worked on the bibliometric analysis. Both authors contributed equally to the introduction and discussion.

FUNDING

This work was supported by the National Research Foundation of Korea (Grant No. NRF-2016R1D1A1B03935697, 2016) and the National Science Foundation Grant DRL # 1439227. The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Research Foundation of Korea. Open Access publication of this work was supported by the Indiana University Bloomington Open Access Article Publishing Fund.

ACKNOWLEDGMENTS

We thank Sebastian Grauwin for the bibliometric analysis and the CSCL BC Map reported in Jeong et al. (2019b).

[2]http://www.alliancelss.com/

REFERENCES

Chen, Y., Andrews, C., Hmelo-Silver, C. E., and D'Angelo, C. (2020). Coding schemes as lenses on collaborative learning. *Inform. Learn. Sci.* 121, 1–18. doi: 10.1108/ILS-08-2019-0079

Chi, M. T. H., and Wylie, R. (2014). The ICAP framework: linking cognitive engagement to active learning outcomes. *Educ. Psychol.* 49, 219–243. doi: 10.1080/00461520.2014.965823

Cohen, E. G. (1994). Restructuring the classroom: conditions for productive small groups. *Rev. Educ. Res.* 64, 1–35. doi: 10.3102/00346543064001001

Crane, D. (1967). The gatekeepers of science: some factors affecting the selection of articles for scientific journals. *Am. Sociol.* 2, 195–201.

Danish, J. A., and Gresalfi, M. (2018). "Cognitive and sociocultural perspective on learning: tensions and synergy in the cognitive and sociocultural perspective on learning," in *International Handbook of the Learning Sciences*, eds F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann (Abingdon: Routledge), 33–43. doi: 10.4324/9781315617572-4

DeLisi, R., and Golbeck, S. L. (1999). "Implications of Piagetian theory for peer learning," in *Cognitive Perspectives on Peer Learning*, eds A. M. O'Donnell and A. King (Abingdon: Routledge), 3–37.

Goldman-Segall, R., and Maxwell, J. (2002). "Computers, the Internet, and new media for learning," in *Handbook of Psychology: Educational Psychology*, Vol. 7, eds W. M. Reynolds and G. E. Miller (New York, NY: John Wiley and Sons, Inc), 393–427. doi: 10.1002/0471264385.wei0716

Grauwin, S., Beslon, G., Fleury, É, Franceschelli, S., Robardet, C., Rouquier, J.-B., et al. (2012). Complex systems science: dreams of universality, interdisciplinarity reality. *J. Am. Soc. Inform. Sci. Technol.* 63, 1327–1338. doi: 10.1002/asi.22644

Grauwin, S., and Jensen, P. (2011). Mapping scientific institutions. *Scientometrics* 89, 943–954. doi: 10.1007/s11192-011-0482-y

Hmelo-Silver, C. E. (2019). Workshop Report: Building Interdisciplinary Capacity for Understanding and Supporting Computer-Supported Collaborative Learning. ISLS Rapid Community Reports. Available online at: http://go.iu.edu/2crE doi: 10.1007/s11192-011-0482-y (accessed December 18, 2020).

Hmelo-Silver, C. E., and Jeong, H. (in press). "An overview of CSCL methods," in *International Handbook of Computer Supported Collaborative Learning*, eds U. Cress, C. Rosé, A. F. Wise, and J. Oshima (New York, NY: Springer).

Hmelo-Silver, C. E., Jung, J., Lajoie, S., Yu, Y., Lu, J., Wiseman, J., et al. (2016). "Video as context and conduit for problem-based learning," in *Educational Technologies in Medical and Health Sciences Education*, eds S. M. Bridges, L.-K. Chan, and C. E. Hmelo-Silver (Berlin: Springer), 57–77. doi: 10.1007/978-3-319-08275-2_4

Hoadley, C. (2018). "A short history of the learning sciences," in *International Handbook of the Learning Sciences*, eds F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann (New York NY: Taylor and Francis), 11–23. doi: 10.4324/9781315617572-2

Jeong, H., Hmelo-Silver, C. E., and Jo, K. (2019a). Ten years of computer-supported collaborative learning: a meta-analysis of CSCL in STEM education during 2005-2014. *Educ. Res. Rev.* 28:100284. doi: 10.1016/j.edurev.2019.100284

Jeong, H., Hmelo-Silver, C. E., and Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *Int. J. Com. supp. Collab. Learn.* 9, 305–334. doi: 10.1007/s11412-014-9198-3

Jeong, H., Seo, S., Jeong, J., Hmelo-Silver, C. E., and Grauwin, S. (2019b). "What can be learned about computer-supported collaborative learning from a bibliographic coupling analysis?," in *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer-Supported Collaborative Learning (CSCL) 2019*, Vol. 2, eds K. Lund, G. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, and M. Baker (Lyon: International Society of the Learning Sciences), 712–715.

Kessler, M. M. (1963). Bibliographic coupling between scientific articles. *Am. Document.* 24, 123–131.

Koschmann, T. D. (Ed.). (1996). *CSCL, Theory and Practice of an Emerging Paradigm*. Routledge.

Krauss, R. M., and Fussell, S. R. (1990). "Mutual knowledge and communicative effectiveness," in *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, eds J. Galegher, R. E. Kraut, and C. Edigo (Hillsdale, NJ: Erlbaum).

Levine, J. M., and Thompson, L. (1996). "Conflict in groups," in *Social Psychology: Handbook of Basic Principles*, eds. E. T. Higgins and A. Kruglanski (New York, NY: The Guilford Press).

Ludvigsen, S., Lund, K., and Oshima, J. (in press). "A conceptual stance on CSCL history," in *International Handbook of Computer Supported Collaborative Learning*, eds U. Cress, C. Rosé, A. F. Wise, and J. Oshima (New York, NY: Springer).

Lund, K., Jeong, H., Grauwin, S., and Jensen, P. (2020). Research in education draws widely from the social sciences and humanities. *Front. Educ.* doi: 10.3389/feduc.2020.544194 [Epub ahead of print].

McKeown, J., Hmelo-Silver, C. E., Jeong, H., Hartley, K., Faulkner, R., and Emmanuel, N. (2017). "A meta-synthesis of CSCL literature in STEM education," in *Proceedings of CSCL 2017*, eds B. K. Smith, M. Borge, E. Mercier, and K.-Y. Lim (Philadelphia, PA: International Society of the Learning Sciences).

Mingers, J., and Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *Eur. J. Operat. Res.* 246, 1–19. doi: 10.1016/j.ejor.2015.04.002

Miyake, N. (2007). "Computer supported collaborative learning," in *The SAGE Handbook of e-Learning Research*, eds R. Andrews and C. Haythornthwaite (Thousand Oaks, CA: SAGE), 248–265. doi: 10.4135/9781848607859.n11

Mott, B. W., Taylor, R. G., Lee, S. Y., Rowe, J. P., Saleh, A., Glazewski, K. D., et al. (2019). "Designing and developing interactive narratives for collaborative problem-based learning," in *International Conference on Interactive Digital Storytelling*, eds R. Cardona-Rivera, A. Sullivan, and R. Young (New York, NY: Springer), 86–100. doi: 10.1007/978-3-030-33894-7_10

O'Donnell, A. M., and King, A. (1999). *Cognitive Perspectives on Peer Learning*. Mahwah, NJ: Erlbaum.

O'Donnell, A. M., and O'Kelly, J. (1994). Learning from peers: beyond the rhetoric of positive results. *Educ. Psychol. Rev.* 6, 321–349. doi: 10.1007/bf02213419

Pea, R., and Linn, M. C. (2020). Personal perspectives on the emergence of the learning sciences: 1970s–2005 [Conceptual Analysis]. *Front. Educ.* 5:130. doi: 10.3389/feduc.2020.00130

Pintrich, P. R. (1999). The role of motivation in prompting and sustaining self-regulated learning. *Int. J.Edu. Res.* 31, 459–470. doi: 10.1016/S0883-0355(99)00015-4

Roschelle, J. (1992). Learning by collaborating: convergent conceptual change. *J. Learn. Sci.* 2, 235–276. doi: 10.1207/s15327809jls0203_1

Saleh, A., Chen, Y., Hmelo-Silver, C. E., Glazewski, K., Mott, B. W., Rowe, J. P., et al. (2020). Synergistic scaffolds for collaborative inquiry in game-based learning environments. *J. Res. Sci. Teach.* 57, 1490–1518. doi: 10.1002/tea.21656

Saleh, A., Hmelo-Silver, C. E., Glazewski, K. D., Mott, B., Chen, Y., Rowe, J., et al. (2019). Collaborative inquiry play: a design case to frame integration of collaborative problem solving with story-centric games. *Inform. Learn. Sci.* 120, 547–566. doi: 10.1108/ils-03-2019-0024

Stahl, G., Koschmann, T., and Suthers, D. D. (2014). "Computer-supported collaborative learning," in *Cambridge Handbook of the Learning Sciences*, ed. R. K. Sawyer (New York, NY: Cambridge University Press), 479–500. doi: 10.1017/CBO9781139519526.029

Suthers, D. D., Lund, K., Rosé, C. P., and Teplovs, C. (2013). "Achieving productive multivocality in the analysis of group interactions," in *Productive Multivocality in the Analysis of Group Interactions*, eds D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, and N. Law (Boston, MA: Springer), 577–612. doi: 10.1007/978-1-4614-8960-3_31

von Glasersfeld, E. (1995). "A constructivist approach to teaching," in *Constructivism in Education*, eds L. P. Steffe and J. Gale (Mahwah, NJ: Erlbaum), 3–15.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

POLICY AND PRACTICE REVIEWS
published: 23 March 2021
doi: 10.3389/feduc.2020.579982

FALKE: Experiences From Transdisciplinary Educational Research by Fourteen Disciplines

Anita Schilcher[1], Stefan Krauss[2], Petra Kirchhoff[3], Alfred Lindl[4], Sven Hilbert[4], Katharina Asen-Molz[4], Christina Ehras[5], Michael Elmer[6], Mario Frei[4], Lisa Gaier[1], Maria Gastl-Pischetsrieder[1], Eileen Gunga[1], Renate Murmann[7], Simone Röhrl[2], Anna-Maria Ruck[7], Matthias Weich[7], Arne Dittmer[5], Michael Fricke[7], Bernhard Hofmann[8], Josef Memminger[7], Astrid Rank[4], Oliver Tepner[6] and Christiane Thim-Mabrey[1]*

[1]Faculty of Languages, Literature and Culture, University of Regensburg, Regensburg, Germany, [2]Faculty of Mathematics, University of Regensburg, Regensburg, Germany, [3]Faculty of Philosophy, Erfurt University, Erfurt, Germany, [4]Faculty of Human Sciences, University of Regensburg, Regensburg, Germany, [5]Faculty of Biology and Pre-Clinical Medicine, University of Regensburg, Regensburg, Germany, [6]Faculty of Chemistry and Pharmacy, University of Regensburg, Regensburg, Germany, [7]Faculty of Philosophy, Art, History, and Humanities, University of Regensburg, Regensburg, Germany, [8]Faculty of Philosophy and Social Sciences, University of Augsburg, Augsburg, Germany

Edited by:
Matthias Stadler, Ludwig Maximilian University of Munich, Germany

Reviewed by:
Nicole Heitzmann, Ludwig Maximilian University of Munich, Germany
Nicolas Becker, Saarland University, Germany

***Correspondence:**
Anita Schilcher
anita.schilcher@ur.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *03 July 2020*
Accepted: *23 December 2020*
Published: *23 March 2021*

Citation:
Schilcher A, Krauss S, Kirchhoff P, Lindl A, Hilbert S, Asen-Molz K, Ehras C, Elmer M, Frei M, Gaier L, Gastl-Pischetsrieder M, Gunga E, Murmann R, Röhrl S, Ruck A-M, Weich M, Dittmer A, Fricke M, Hofmann B, Memminger J, Rank A, Tepner O and Thim-Mabrey C (2021) FALKE: Experiences From Transdisciplinary Educational Research by Fourteen Disciplines. Front. Educ. 5:579982. doi: 10.3389/feduc.2020.579982

This article details how the FALKE research project (**Fa**chspezifische **L**ehrer**k**ompetenzen im **E**rklären; Engl.: subject-specific teacher competency in explaining) integrates 14 heterogeneous disciplines in order to empirically examine the didactic quality of teacher explanations in eleven school subjects by bringing together trans-, multi-, and interdisciplinary perspectives. In order to illustrate the academic landscape of the FALKE project we briefly outline the nature of the transdisciplinary German "Fachdidaktiken" (Engl.: subject-matter didactics, i.e., special academic disciplines of teaching and learning specific school subjects). The FALKE project required the willingness of all researchers from eleven participating subject-matter didactics to rely on both the concepts and the methods of educational sciences as an overarching research framework (transdisciplinary aspect). All researchers of subject-matter didactics had to develop a shared conceptual, methodological, and administrative framework in order to empirically investigate commonalities in and differences between "good explanations" across the range of school subjects represented (multidisciplinary aspect). The additional perspectives of researchers in speech science and linguistics proved fruitful in recognizing rhetorical and linguistic aspects of teacher explanations (interdisciplinary aspect). Data management and statistical analysis were provided by the discipline methods of educational sciences. Rather than reporting empirical results, we here discuss opportunities and challenges as well as the lessons learned from the FALKE project regarding cognitive-epistemic reasoning, communication, and organization.

Keywords: transdisciplinarity, multidisciplinarity, interdisciplinarity, subject-matter didactics, FALKE research project, explaining, teacher competence, pedagogical content knowledge

INTRODUCTION

In matters of learning and education, the question of what makes a good explanation has been pondered for centuries. In his *Didactica magna* (1657), Comenius (1967) was already asking what a good explanation was and how a teacher could explain well. Didactically appropriate explanations are at the heart of high-quality teaching and learning experiences in any subject. According to Gage (1968), explaining is a core aspect of a teacher's professional competence:

> *Explaining may come close to being the essence of instruction, so that when a teacher is attempting to explain proportionality to his geometry class or irony to his English class, he is behaving more purely as a teacher than when he is attempting, say, to motivate, promote discussion, or maintain discipline (p. 3).*

When students are asked about the role of teacher competency in explaining, the empirical evidence is undisputed: based on a survey with more than 1,000 participants, Wragg and Wood (1984) reported that school students clearly considered explanation competency to be the most important skill of a teacher. A more recent study by Kulgemeyer and Peters (2016) demonstrates similar findings with regard to the subject of physics. But even though explaining has been demonstrated to play such a crucial role in teaching and learning in all instructional contexts, there is still a dearth of empirical research on this topic (Odora, 2014; Findeisen, 2017).

The question that really needs to be answered is which scientific discipline can best examine and analyze—theoretically and empirically—what a good explanation or act of explaining actually is, including its ultimate effect on learners. One could argue that educational psychology is the most appropriate research discipline for this task: it has both well-proven methodological tools (i.e., statistics and psychometrics) and a broad foundation of relevant conceptual models and empirical evidence. Ideally, these prerequisites can serve as a productive and reliable basis for general research on explaining. But in meta-studies, the greatest predictors seen thus far on effective teaching can be found in the domain-specific components of teaching (Seidel and Shavelson, 2007), and, in fact, the research community has acknowledged that the subject-related perspective is key to understanding teaching and learning. As early as the 1980s, in a comprehensive theoretical analysis much noticed by the international educational research community, Shulman (1986) had already pointed out the necessity of a stronger relationship between pedagogical processes and the content to be conveyed in research.

> *In their necessary simplification of the complexities of classroom teaching, investigators ignored one central aspect of classroom life: the subject matter. This omission also characterized most other research paradigms in the study of teaching. Occasionally, subject matter entered into the research as a context variable—a control characteristic for subdividing data sets by content categories (e.g., "When teaching 5th grade mathematics, the following teacher behaviors were correlated with outcomes. When teaching 5th grade reading…"). But no one focused on the subject matter content itself [...] Why this sharp distinction between content and pedagogical process?* (p. 6).

Over the last two decades, we have seen a substantial number of publications (especially on STEM education) answering Shulman's call for a closer look at subject matter in classroom-based teaching and learning processes. What is still missing, however, is a broader—or even joint—engagement in research on 1) teachers' professional competencies in teaching and 2) students' ensuing learning of subject-matter in all school disciplines. In recent years, even educational psychologists have been critical of the fact that empirical studies in this domain predominantly focus on mathematics and science and then generalize their findings to other school disciplines (Leutner et al., 2017; Praetorius et al., 2018). And one might well doubt whether such generalized statements really could apply to all school subjects. We rather need to ask ourselves—across disciplinary boundaries—which evidence and statements we might be able to generalize from one subject to others. Conversely, we need to consider when we should only look at teaching and learning through the lens of the specific subject, taking into account its highly complex content and domain-specific learning processes.

So, should the competency of explaining well be investigated exclusively within the corresponding subject-matter discipline? The mathematician determines what makes a good mathematical explanation for sine and cosine, the biologist how to explain evolution and the associated genetic changes, and the literary scholar how best to interpret texts or how to elaborate on the nature and function of Francis Underwood's infamous asides in *House of Cards*. Yet, is the academic subject-matter expert automatically an expert on teaching and learning, especially when not the academic but the school-related content of a discipline is considered?

Without a doubt, the only way to make a decision in each of those cases as to what makes a valid explanation of content requires discipline-specific knowledge. Following common sense, a *wrong* explanation (e.g., of why 1 + 1 = 3) can never be a "good explanation." However, it is also evident that not every valid explanation is, automatically, a high-quality explanation, for example, in terms of the learning gains of students. Therefore, One idea would be to bring the instructional knowledge of educational science (e.g., psychology, pedagogy, etc.) and expertise of the subject-matter disciplines (e.g., chemistry, English, geography, etc.) together. To this end, Kirschner et al. (2017) proposed an interdisciplinary cooperation between instructional designers and experts from the disciplines:

> *Assume that I, as a cognitive-psychologically based instructional designer, am designing a new learning environment in a particular subdomain of mathematics. I don't know if I need to have deep*

conceptual understanding of the topics to be taught. But of course I will need to have someone working together with me who does have that deep conceptual knowledge. And, of course, I will need some basic knowledge of the (sub)domain in order to make sure that the communication and cooperation with my partner works well. (p. 642).

Kirschner et al. (2017) seem to assume that *interdisciplinarity* (for a distinction between interdisciplinarity, multidisciplinarity, and transdisciplinarity, see below) is sufficient for successfully overcoming challenges in domain-specific teaching or explaining. Nevertheless, in our view, this approach to innovation overlooks the fact that the canon content of any school curriculum, and the corresponding understanding of how thoroughly this content should be taught, is subject to continuous change and negotiation (Kansanen, 2002). An academic discipline generates academic knowledge that is intended for the discussion that takes place within that highly specialized discipline. Which of these academic ideas should become part of the school curriculum and how learning from lower to higher levels of abstraction and complexity should take place is a question that scientific communities usually do not engage in (Abraham, 2019).

Educational psychologists possess general knowledge about teaching and learning and the relevant predictors of learning processes on the level of general constructs (such as cognitive activation of learners or classroom management, cf. Kunter et al., 2013). However, these principles have to be transferred to a school curriculum that has a variety of heterogeneous subjects and corresponding contents (Praetorius et al., 2020). Educational psychology tells us, for example, that a *clear structure* in an explanation helps students to gain a better understanding. To map this concept of clarity onto existing structures of teaching and learning specific subject matter, and particularly while keeping in mind real-life learners, is not as easy as it seems at first. Multiple layers of knowledge and expertise are needed to explain well the meaning of a word like "mansplaining" in a multilingual or multicultural class, or the process of creating a convincing argument in a written text, or the orchestration of instruments in a beginning brass band. What is more, which exact competencies should be acquired by school students in different subjects is an open question. Only the core literacies (e.g., reading, mathematics, science, etc.) in large-scale assessments like *Programme for International Student Assessment* (PISA) have been comprehensively defined and empirically validated (e.g., OECD, 2019). The complexity of these demands exceeds the potential of a solely interdisciplinary cooperation between experts in educational sciences and subject-matter research.

The science philosopher Mittelstrass (2011), too, sees interdisciplinary cooperation as not enough of a solution for complex problems (in FALKE: explaining subject-matter content) because in interdisciplinary research, the academic disciplines "contribute what they know, but they do not change themselves in their forms of knowledge or methodology" (p. 336). In order to find out how teachers can provide didactically good explanations, a *transdisciplinary* approach is indispensable. Mittelstrass sees transdisciplinarity as a form of cooperation that will "lead to an enduring and systematic scientific order that will change the outlook of subject matters and disciplines. Transdisciplinarity is a form of scientific work which arises in cases concerning the solution of non-scientific problems" and "a principle of research and science, one which becomes operative wherever it is impossible to define or attempt to solve problems within the boundaries of subjects or disciplines" (p. 331).

Thus far, truly transdisciplinary research has flourished in areas such as public health science (e.g., Rosenfield, 1992; Turnbull et al., 2019), environmental research (e.g., Hoffmann et al., 2009), sustainability research (e.g., Schneidewind, 2010), nanotechnology or the quantum-mechanic measurement process and the concept of information (e.g., Pohl et al., 2008; Mittelstrass, 2011). Given the importance of school education (e.g., for the prosperity of societies; Woessmann, 2016), it is surprising that we are not yet looking at a similar wealth of transdisciplinary research on educational science and subject-matter didactics.

The characteristics of transdisciplinarity directly apply to the FALKE research program (for details see, e.g., **Figure 1**): finding out what makes up a good explanation in a school context is a non-scientific, real-world problem. Hence, one discipline cannot resolve it on its own. Its untangling is, rather, an endeavor that touches multiple disciplines: First, knowledge of the corresponding (academic) subject-matter discipline is needed to be able to decide on the validity of the explanation. Second, educational psychology provides valuable insights at a general level, for example on learners' general cognitive development and information processing. Third, applied linguistics offers a sound understanding of the salient linguistic features of explanations (e.g., the recommended number of words per sentence, the limited use of relative clauses, or of the passive voice, etc.), and speech science might supply insights into embodied teacher performance (e.g., voice, body expression, etc.) and its effect on the learner's perception. In addition, psychometrics can point to how to operationalize the addressed constructs (e.g., by questionnaires or tests), which experimental design has to be implemented, and which statistical analyses have to be conducted for answering specific research questions. What is still missing from this scenario, however, is the expertise and unifying force of subject-matter didactics.

TRANSDISCIPLINARY SUBJECT-MATTER DIDACTICS

Subject-matter didactics were the driving force behind the setting up of the FALKE research program. As subject-matter didactics is not an internationally known academic discipline, we will briefly explain its development and current purpose (also see middle column in **Figure 2**). The disciplines of subject-matter didactics can be found, for example, in many European universities (cf. Kansanen, 2002; Rothgangel and Vollmer, 2020). In Anglo-American countries, the adjective "didactic" has a negative connotation, suggesting oversimplified ideas of teaching and learning or "recipe-book instructions" on teaching methodology

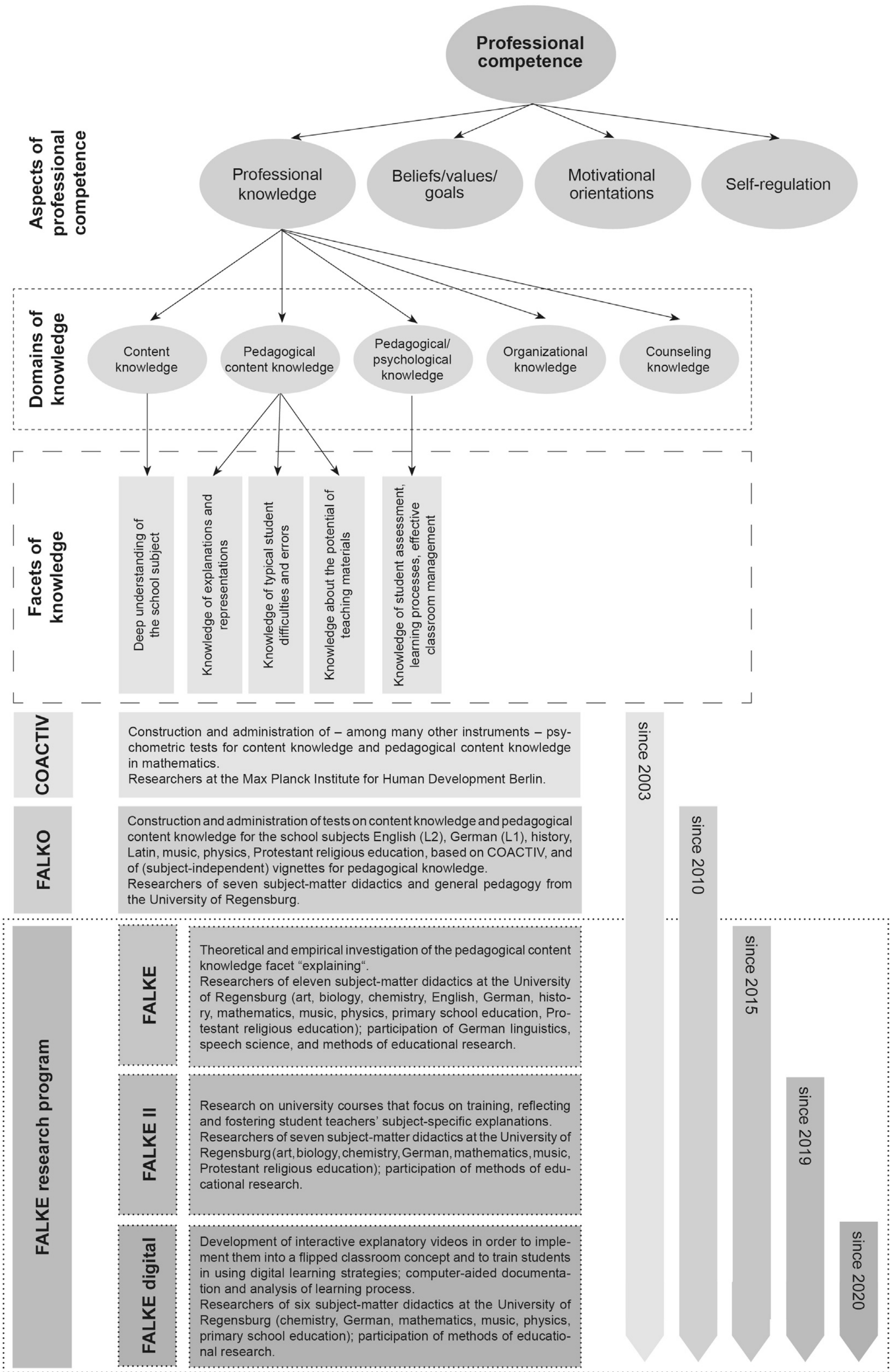

FIGURE 1 | The COACTIV model of teachers' professional competence, the preceding project FALKO (above) and the three projects of the FALKE research program (below) at the University of Regensburg.

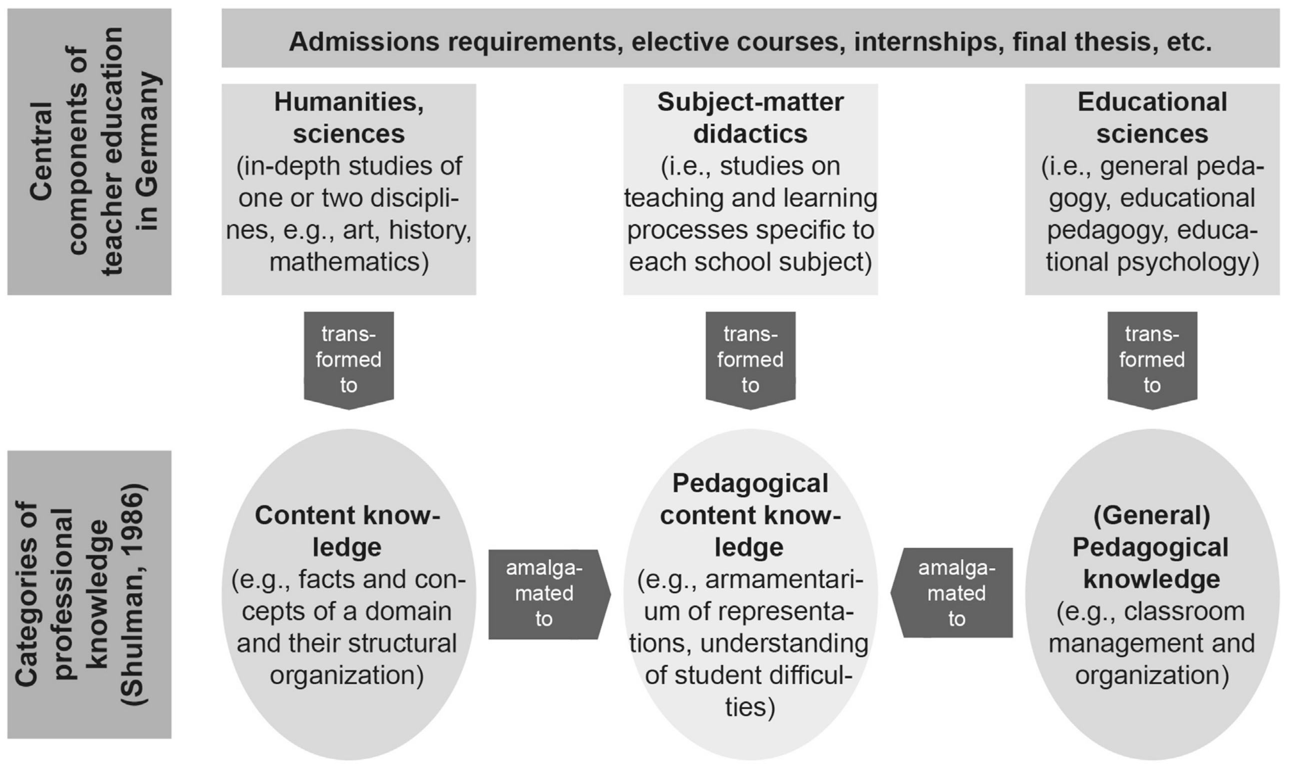

FIGURE 2 | Components of teacher education at German universities (first row) and teachers' corresponding professional knowledge categories according to Shulman (second row).

(Arnold, 2012), which may carry over to the noun "didactics." The idea of *didactics* originally stems "from the German tradition of theorizing classroom learning and teaching" (Arnold, 2012; p. 986). *Subject-matter didactics* disciplines (e.g., mathematics didactics, history didactics, music didactics) conceptualize teaching and learning as strongly situated in content.

Traditionally, the subject-matter didactics disciplines were asked to make normative decisions on the canon and to transform (academic) subject-matter content for (school-related) learning purposes. In German-speaking countries, professors of subject-matter didactics are therefore assigned for the most part to the faculties of the corresponding disciplines (e.g., biology didactics in the faculty of biology, etc). As a result, the subject-matter didactics disciplines tend to connect strongly with the respective subject-matter discourse (left column in **Figure 2**). To a great extent, the logic of the corresponding subject frames the thinking and informs the research interests of the individual researchers in the corresponding subject-matter didactics.

Lately, a growing number of researchers in subject-matter didactics has begun to see their disciplines as an evidence-based science having the following objectives in mind (Leutner et al., 2017): First, seeking to develop theories and models and to formulate (verifiable) hypotheses about subject-specific teaching and learning phenomena and challenges. Second, addressing these subject-specific phenomena and challenges on an empirical level, for instance by implementing quantitative correlational or experimental designs, or by following qualitative research paradigms such as conducting field-observations or interviews. Third, analyzing the data obtained and integrating the findings into the body of already existing evidence. Note that such attempts are not restricted to students' subject-specific learning processes but in the same way apply, especially over the last decade, to the subject-specific professional competence of teachers and its development (for a teacher competence model, see **Figure 1**).

This opening toward an evidence-based approach—while simultaneously maintaining the logic and the framework of the corresponding subject-matter discipline—comes with an increased orientation toward and integration of the concepts and methods of educational science (right in **Figure 2**) that provide both an understanding of statistical methods as well as an awareness of general concepts on teaching and learning. In this sense, the field of subject-matter didactics—for those who are open to this path—must address transdisciplinarity on two levels: Beyond becoming more transdisciplinary as academic disciplines themselves, they should—also in line with Mittelstrass (2011), see above—not only reach out toward enduring cooperation with educational scientists but also with educators at schools as well as appropriate governmental officials.

So far, this level of transdisciplinarity (i.e., integration of and cooperation with educational psychology) is not common practice for most researchers in subject-matter didactics, except perhaps for those working in subject areas that are often tested in large-scale assessments (e.g., PISA). Meanwhile, several researchers in mathematics didactics, science didactics, and the didactics of German language and literature engage on the mentioned levels of transdisciplinarity on a regular basis. In other subject-matter didactics like music, history, or geography, however, currently only a small number of researchers make use of this shifting paradigm. One reason for this discrepancy is that the latter subjects have not been in the focus of national or international large-scale assessments and, therefore, have never experienced the pressure to take on empirical research methods (see below). Furthermore, subject-matter didactics researchers traditionally follow a career path, where they are educated first in the respective subject matter (including its didactics) for being a future teacher but usually receive little training in empirical research methods.

Pedagogical Content Knowledge (PCK) as a Cornerstone in Subject-Matter Didactics

The field of subject-matter didactics increasingly sees teachers' professional competence (e.g., upper part of **Figure 1**) as the central hub for developing and maintaining quality in teaching and learning. Within his prominent taxonomy of teacher knowledge (also see lower part of **Figure 2**), Shulman (1986), in addition to the categories of *content knowledge* (CK) and *pedagogical knowledge* (PK), conceptualizes the concept of *pedagogical content knowledge* (PCK) as one decisive aspect of a teacher's professional knowledge:

> *Within the category of pedagogical content knowledge I include, for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others. Since there are no single most powerful forms of representation, the teacher must have at hand a veritable armamentarium of alternative forms of representation, some of which derive from research whereas others originate in the wisdom of practice. Pedagogical content knowledge also includes an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons. If those preconceptions are misconceptions, which they so often are, teachers need knowledge of the strategies most likely to be fruitful in reorganizing the understanding of learners, because those learners are unlikely to appear before them as blank slates.* (p. 9–10)

In a theoretical review, Rothgangel and Vollmer (2020) remark that "Lee Shulman's notion of 'Pedagogical Content Knowledge' (PCK) comes closest to the meaning of subject-matter didactics" (p. 129). According to Shulman, PCK can be considered an "amalgam" of CK and PK (Shulman, 1987; p. 8; also see **Figure 2**). Thus, a teacher's PCK draws on knowledge repositories of subject-matter and pedagogy as well as psychology and transforms them into classroom performance. In German classes, for example, teachers need to combine their knowledge of youth literature and textual genres with insights into the reading process and their own diagnostic knowledge of individual children's competencies. They should then use this basis to develop an instructional design for the effective teaching of reading, interpreting literary texts, and developing and sustaining reading motivation (Schilcher and Wild, 2018). Of course, by focusing on the concept of PCK, subject-matter didactics does not lose sight of other areas of teachers' professional competence, like teachers' beliefs and enthusiasm as well as their continuous professional development in communities of practice.

In the past years, Shulman's idea of teachers' PCK has been taken up as a central concept in empirical studies in subject-matter didactics. In the COACTIV study on mathematics teachers' competencies (**Figure 1**), for instance, PCK tests were constructed including several items on how to explain mathematical content and how to deal with typical student difficulties. These scenarios were implemented in a paper-and-pencil format as well as in a test format based on short video vignettes (Krauss et al., 2020). It could be shown that the PCK of secondary mathematics teachers, especially as measured by the paper-and-pencil instrument, was—among many other modeled teacher competencies—by far the highest predictor for student achievement (Kunter et al., 2013). For an overview on corresponding psychometric knowledge test constructions on PCK in various other subjects than mathematics, for instance, Krauss et al. (2017, 2020) can be consulted. In the following, we focus on some aspects of PCK specific to 1) teacher education and 2) subject-matter didactics research.

PCK in Teacher Education

In 2000, the mediocre PISA results in mathematical literacy, science literacy and reading literacy of German 9th graders (Baumert et al., 2001) were a "shock," not only for teachers and educational administrators but also for the general German society. Since these results were interpreted as an indication of a potential lack of quality in teacher education in many public and scientific debates, a broad discussion on a reform of teacher education followed—including the role of subject-matter didactics. Later this was fueled by Hattie's (2009) slogan, "what teachers do matters." To set compulsory standards, German educational policy makers established new standards of teacher education with an underlying model of teacher competencies (for further development see, e.g., KMK, 2019).

Ideally, teacher education should be regarded as a process of professionalization that integrates knowledge repositories rather than teaching them as isolated content. But the curricular structure of teacher education in various countries shows that CK and PK are most often taught separately even though within the same study program. Following, for instance, Kirschner et al.

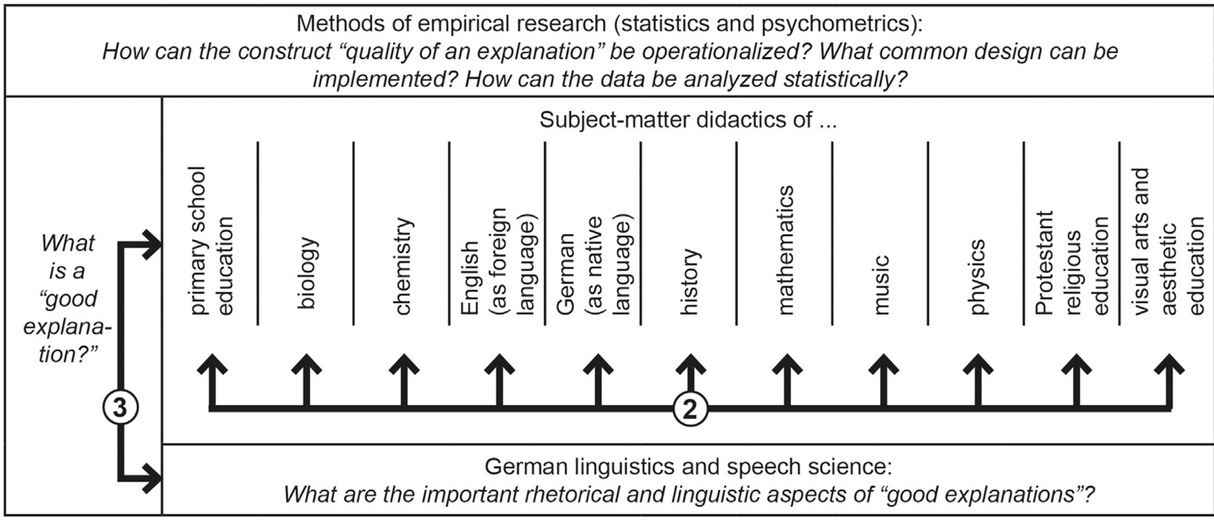

FIGURE 3 | Transdisciplinarity, multidisciplinarity, and interdisciplinarity in the FALKE I project.
①: Transdisciplinarity: Experts in empirical research supported construct operationalizations, design development and data analyses (for both shared and subject-specific research questions).
②: Multidisciplinarity: The research question ("What determines a 'good explanation' in the respective school subject?") was analyzed domain-specifically (but with a common theoretical, conceptual and administrative framework) in eleven disciplines in parallel.
③: Interdisciplinarity: Both speech science and German linguistics contributed by considering rhetorical and linguistic aspects of explanations in all eleven subjects.

(2017), the underlying idea seems to be that this parallel teaching practice facilitates the implicit development of the "amalgam" of PCK in some miraculous way. Although subject-matter didactics in German teacher education programs includes pedagogy and psychology among its reference sciences—in addition to the respective content-related disciplines—corresponding teaching collaborations remain sparse. Even given the existence of institutionalized subject-matter didactics, the three columns (**Figure 2**) only rarely communicate with regard to teacher education. Worldwide, the subject-matter didactics disciplines have dedicated themselves to teaching subject-specific PCK in university teacher education (for other areas of subject-matter didactics see, for instance, Rothgangel and Vollmer, 2020).

The German Federal Ministry of Education and Research (Bundesministerium für Bildung and Forschung, 2014) announced in 2014 a program called *Qualitätsoffensive Lehrerbildung* (Engl.: Teacher Education Initiative) in order to promote collaboration in German teacher education among different areas of expertise. From 2015 to 2023, German federal and state governments provide funding for different university-based projects intended to improve the process of teacher education in a sustainable manner along three slightly varying funding lines. A key criterion for the allocation of funding is a better coordination of teacher education specialists across disciplinary boundaries (i.e., the three columns in **Figure 2**) that is also ideally evidence-based (for research issues, see next section). Each of the three subprojects of the FALKE research program (**Figure 1**) was funded by one of the three BMBF funding lines (altogether funding for 26 doctoral positions could be acquired across all three FALKE projects). In this paper we especially discuss experiences in the first subproject of FALKE (**Figure 3**, **Table 1**).

Research on PCK

In the history of educational research on subject-matter teaching and learning, two pathways for theoretically and empirically investigating PCK, including its determinants and consequences, have unfolded. On the one pathway, educational psychologists, predominantly in Anglo-American countries, have become experts on subject-specific learning processes. So, for instance, psychologists like Stanovich (1991) and Schiefele et al. (2012) became experts on the development of reading, Graham and Harris (2005) and Hayes and Flower (1980) on the development of writing, and Hill and colleagues on mathematics education (Hill et al., 2005; Hill et al., 2008). As a result, some psychologists have contributed research that specializes particularly in the core literacies mentioned.

Yet, this trend has led to an increasing particularization of subject-matter domains, whereas "the capacity to think in disciplinarities, that is, in larger units of science, (is) decreasing" (Mittelstrass, 2011; p. 33). And while further particularization might work for highly domain-specific research, it may in fact be detrimental in teacher education (see previous section), where a general overview of the subject matter is just as important as in-depth knowledge. What is more, psychologists usually have neither a deep subject-matter knowledge that covers all fields of a certain school discipline (for instance, language teaching with a focus on literary history and youth literature, textual genres, film, media, linguistics, orthography, etc.) nor an understanding of their interdependencies (e.g., for promoting reading literacy).

TABLE 1 | Study design of FALKE.

FALKE I: 11 parallel online questionnaires (evaluation of 6 videos of teacher explanations per subject)			
Evaluation of the same videos	Six videos at approx. three minutes	Three selected and shortened video clips at approx. 30 seconds	Six videos at approx. three minutes
Participants *(status groups)*:	Holistic assessment *(school grades from 1 to 6)*	Criterium-based assessment *(rating scale from 1 to 6)*	Criterium-based assessment *(rating scale from 1 to 6)*
School Students **Teacher Students** **Teachers** **Subject-matter didactics specialists**	Global judgement on the quality of the explanations shown in the videos	Items on speech and body expression as well as on personality of the explaining teacher	Items on structuredness, addressee orientation, linguistic comprehensibility and subject-specific quality aspects of the explanations

Moreover, the scope of pedagogical and psychological research often does not go beyond what is regarded as a key competency in education, namely reading, writing, mathematics, and foreign-language acquisition (mostly English as a second language) at a basic level. Consequently, classroom-based learning processes in music, art, religious education, and geography, for example, but also in advanced mathematics like integral calculus in the upper grades, have not received an appropriate share of research. In Kansanen's (2002) view, psychological research has not been able to develop the full scope of research on all school subjects and for real-life teaching and learning in all grades (see Kansanen on withdrawal, fractionation, and even irrelevance of research in educational psychology). Most importantly, he also emphasizes that educational psychology has focused on empirically examining learning rather than teaching which may explain the lack of research on teacher's professional competence at this time, especially regarding different school subjects.

In Germany, the "PISA shock" (see above) was a wake-up call for the subject-matter didactics disciplines to reconsider not only the content and quality of teacher education, but also their own research and publication habits. Around the turn of the millennium, there was too little empirically sound knowledge about subject-specific learning and teaching—despite a long-lasting, lively (but mostly only theoretical) discourse on subject-matter didactics. Since then, subject-matter didactics like German (as a first language), mathematics, the first foreign language (English or French), biology, chemistry, and physics that were repeatedly subject to rigorous standardized testing procedures (e.g., in large scale studies such as PISA, TIMSS, DESI, PIRLS, etc.) have managed to use external pressure to shift their research paradigms toward competence- and output-orientation, both based on empirically gathered evidence. In addition, we can observe a sharp increase in publications and international conference contributions at a competitive level, while empirical research in other subject-matter didactics has been much slower to take off (e.g., with regard to researching instructional quality, cf. Praetorius et al., 2018).

Furthermore, the PISA 2000 shock was the driving force behind the modelling of teachers' competencies and the empirical investigation of the impact of specific competence aspects on student learning ('predictive validity'). Thus, the COACTIV study on German mathematics teachers' competencies was undertaken as a satellite study of PISA 2003. One of its main findings that PCK is by far the strongest predictor of students' learning success (e.g., Kunter et al., 2013) was a particularly interesting result for researchers of subject-matter didactics. In the following, in Germany, a second pathway for examining teachers' professional competencies, specifically the concept of PCK, has developed in subject-matter didactics. For this purpose, PCK tests (each of these accompanied by corresponding CK and/or PK tests) were constructed in the following by many other subject-matter didactics (e.g., the German projects FALKO, ProwiN, TEDS or KiL / KeiLa). Comparatively little research has been published in the Anglo-American world on the construction and validation of psychometric tests of teacher knowledge categories such as PCK, CK or PK (cf. Krauss et al., 2020). In the next section we outline the first FALKE-study that focuses on subject-specific explaining, which is–according to Shulman (1986)–a crucial facet of PCK.

THE FALKE STUDY

Development and Outline

The FALKE I research group (**F**achspezifische **L**ehrer**k**ompetenzen im **E**rklären; English: Subject-specific teacher competency in explaining) is, to our knowledge, the only educational research project that integrates 14 heterogeneous scientific disciplines (**Figure 3**). In this group, trans-, multi-, and interdisciplinary perspectives are coordinated and orchestrated in order to gain a broad understanding of the act of explaining, its corresponding characteristics, and the effect of (oral) teacher explanations given to school students in the classroom.

The project is positioned in a line of research that started with the German COACTIV study on multiple teacher competencies (**Figure 1** for the history of the FALKE research program). COACTIV was followed by the FALKO project (beginning in 2010), in which six subject-matter didactics disciplines at the University of Regensburg constructed and validated domain-specific knowledge tests on PCK and CK in line with the corresponding tests for mathematics teachers in the COACTIV study. In FALKO the subject-matter didactics of English (as a foreign language), German (as a native language), Latin, physics, Protestant religion, music, and history were involved (Krauss et al., 2017). Finally, the three FALKE projects at the University of Regensburg (FALKE I, FALKE II and FALKE digital, conducted under the three funding lines of the BMBF as mentioned above) followed the overarching concepts of the previously mentioned studies.

In the remainder of the paper, the rationale for, the administration of, and the experiences surrounding the first FALKE project are reported (the authors were researchers under the first funding line, which is why this project is also called "FALKE I").

The aim of FALKE was to empirically examine the didactic quality of teacher explanations in eleven school subjects in parallel. Among the 14 participating disciplines at the University of Regensburg were 11 subject-matter didactics, namely of biology, chemistry, German as a native language, English as a foreign language (TEFL), Protestant religious education, history, mathematics, physics, primary school education, music education, and visual arts and aesthetic education.

Two other relevant disciplines participated with their expertise, speech science and German linguistics. In addition, specialists on research methodology in educational sciences made a substantial contribution to the project (**Figure 3**). One senior and one junior researcher from each discipline were active members of the group. In all, 13 out of the 14 junior researchers were funded by the BMBF (for details, see above).

At the very start, a common conceptual, methodological, and administrative research framework was developed to create the opportunity to generalize results across the eleven school disciplines (see **Table 1** for the design of FALKE). This design also allowed for identifying commonalities and differences of teacher explanations among the different subjects.

Within this framework, each of the 11 subject-matter didactics produced six video vignettes. Each of those vignettes shows a short, classroom-situated explanation by a teacher to a class that is topically salient for the respective subject. For example, the vignettes for English as a foreign language focused on explicit explanations of vocabulary meaning and morphology. In the music education videos, the teacher concentrated on the use of visual or acoustic forms of representation by explaining elementary issues of music theory.

The video vignettes were embedded in an online questionnaire (resulting in eleven instruments differing with respect to the specific videos) that asked for the perceived structuredness, addressee orientation, linguistic comprehensibility, and speech and body expression in each of the explanations (cf. **Table 1**). These constructs were operationalized—for all participating subjects in parallel—by several closed items each. In addition, each video was followed by some subject-specific items (which, of course, also differed between subjects).

In the empirical study, participants from four different relevant "status" groups (i.e., students from school, pre-service teachers at university, in-service teachers, and subject-matter didactics researchers) rated the didactical quality in the filmed explanations, holistically first (by giving an overall rating using school grades, i.e., without any suggestions by listed criteria) and then—after seeing the video vignettes again—based on closed items representing the implemented criteria.

The uniform research design (**Table 1**) makes it possible to use classical quantitative analysis methods such as variance analyses or linear regressions to examine group mean differences and relationships between features in each individual subject (e.g., to find out which of the criteria implemented have a particularly strong influence on the perceived overall quality of the explanation; for first results, see Lindl et al., 2019). As this study is based on an extensive overall sample consisting of four subsamples for each school subject (**Table 1**, altogether N = 3.116 particpants evaluated the videos), it is necessary to consider the individual school subject as a higher level variable in multilevel models and (latent) structural equation models. Only such meta-analytical transdisciplinary approaches allow for the estimation of commonalities and differences between the individual subjects (e.g., via random effects) that can be checked for significance. In a final step, these approaches enable a transdisciplinary generalization of subject-specific findings. Further statistical methods that are especially appropriate for inter- and transdisciplinary educational research (with an exemplary focus on FALKE) are presented and discussed in the same issue of *Frontiers in Education* in Lindl et al. (2020).

Trans-, Multi-, and Interdisciplinary Research in Action

What makes the FALKE project unique is the orchestration of research approaches in trans-, multi-, and interdisciplinary fashion under a common conceptual, methodological, and administrative umbrella that has clearly defined processes, instruments, and procedures of analysis.

The cooperation of the 11 subject-matter didactics with the department of statistics and educational measurement was transdisciplinary in nature (①: first row in **Figure 3**). According to Mittelstrass (2011), this collaboration reorients the participating subject-matter didactics toward an evidence-based positioning that will probably remain in place after FALKE concludes. Underlying this cooperation was the original motivation of addressing a real-world problem: explaining subject-matter.

We call the collaboration of the 11 subject-matter didactics disciplines *multidisciplinary* (②: middle row in **Figure 3**) because all of the subjects implemented the same research paradigm and tried to answer the same questions in parallel. The conceptual framework, study design, and research questions had to be inclusive enough to integrate the characteristics of the individual subject-matter didactics, at least to a certain extent, while at the same time maintaining a minimum level of standardization across subjects in order to arrive at comparable results. This parallel procedure of 11 disciplines guarantees a higher validity in generalizing the results across a range of school subjects.

The collaboration of the subject-matter didactics disciplines with speech science and German linguistics was interdisciplinary by nature (③: last row in **Figure 3**) because no discipline transformed itself. Just for the FALKE project, the aspects of adequate language, voice, speech, and body expression were added for short-term cooperation.

Lessons Learned: How can more than 20 Scientists Solve a Problem Together?

The final integration of the 11 obtained subject-specific data sets into one comprehensive data set allows for drawing overarching conclusions about which explanations are perceived as good across school subjects and pertinent status groups. In addition to the forthcoming publications of dissertations and journal articles by the junior researchers from each of the subjects involved, the results of the individual subject-matter didactics as well as overall meta-analyses will be summarized in a compendium (Schilcher et al., 2021). Managing researchers in eleven closely collaborating subject-matter didactics disciplines including the fact that all had to gain an understanding of the research traditions, salient questions, and approaches coming from the other research domains was at the same time a challenge and an achievement.

Over the course of the project, each participating discipline had to follow the research plan that had been agreed upon. Sometimes this meant that cherished and certainly valuable subject-dependent presuppositions had to be suspended (or even ultimately questioned) during the study. For example, as far as teaching English as a foreign language is concerned, the strong focus on teacher-centered explanations runs contrary to the central methodological paradigm of communicative language teaching. In other subject-matter didactics, the predominant theoretical paradigm is based on constructivist learning theory, which is itself based on student-centered discovery learning. In practical teaching, however, teacher explanations play a central role (Wragg and Wood, 1984; Scheffel, 2019). Thus, for FALKE, it was first necessary to work out what place teacher explanations on, for example, concepts, experiments or arguments, would find in theories on student-centered instruction.

In such a large project, however, issues other than answering the research questions can arise. Bergmann et al. (2005) define a number of problems that have to be mastered in any transdisciplinary project on three interwoven levels: the organizational level, the cognitive-epistemic level, and the communicative-psychological level. Finally, we will briefly address these issues with respect to the FALKE project.

Issues of Organization

The biggest challenge of large collaborative projects is to establish and maintain a culture of participation within an organizational infrastructure that channels trans-, multi-, and interdisciplinary development. Such a reliable network of communication should, at the same time, inspire and focus the development of the research project without losing track of the original objectives, as well as the ever-present restraints of time and funding. Naturally, there is a high danger of missing valuable contributions along the way.

A fixed structure for meetings, information exchange, and development of new ideas is a necessary precondition when working in large transdisciplinary groups. The larger the project, the more important a transparent organization of the project processes and agreements is. One of the most difficult tasks in such a project is informing all of the researchers at all times about all processes and involving all of them in the important decision-making processes. Whenever a task is distributed among several people, there is a high risk that information will not necessarily reach all of those involved. In the FALKE project, there was a clear structure of different group meetings: monthly meetings with the entire group and fortnightly meetings between project management and junior researchers.

The objective of the meetings involving the whole group was to set a decisive course, for example with regard to theoretical aspects (e.g., which theories are shared by all 11 subject-matter didactics?), the joint research questions, or the experimental design (①: first row in **Figure 3**). Additional meetings of smaller groups (mainly of subgroups of the doctoral students) were aimed at making progress in terms of content, such as achieving a common understanding of central concepts or discussing the definition and operationalization of the various facets of an explanation following a literature review (e.g., structuredness, addressee orientation, linguistic comprehensibility and subject-specific quality aspects of the explanations, etc.) (②: middle row in **Figure 3**). In addition, a common exchange platform for collecting secondary literature or recording work results or agreements was established. When selecting and constructing the video vignettes, the junior researchers cooperated closely with their respective supervisors (mainly working in pairs), since professional expertise in the subject was of decisive importance here (and thus a fourth kind of cooperation existed within each subject-matter didactics group between the doctoral student and his or her advisor).

On the organizational level, the common analytical framework, identifying relevant predictors (including agreement on their operationalizations at item level), and the (centralized) statistical analyses turned out to be most critical for the progress of the project. With those in place, the methodologists could guarantee the basis for common analyses and interpretation of the data for all disciplines while taking the commonalities and specificities of all of the subjects into account. Simultaneously, the junior researchers engaged in extensive training on empirical research methodology given in centrally organized lectures and workshops. Additionally, the project's experts in research methodology participated in whole-group presentations and discussions during the phase of analyzing the entire data set.

The last phase of the project was dedicated to producing a joint volume of the results to be published in addition to the individual dissertations (Schilcher et al., 2021). In this compendium, the conceptual framework will be explicated by the project management team in an introductory chapter, and then the individual subject-specific chapters written in cooperation between senior and junior researchers will follow and be uniformly structured in order to allow easy comparisons between subjects. The book will close with a chapter on the overall results (including the meta-analyses) presented by FALKE's experts on statistical analysis.

To sum up, transdisciplinary projects often deal with complex issues where many different levels and problems have to be mastered. The exchange of information between the different working units is a central challenge. Even if minutes and information on results of discussions are provided reliably and on a regular basis, these cannot completely reflect the discussion processes. What is more, working in large groups can be cumbersome at times, and it is an ongoing challenge to keep up the momentum.

Cognitive-Epistemic Issues

On a cognitive-epistemic level, the focus of FALKE was the linking of different types of knowledge and competence repositories, from different disciplines as well as between academic and non-academic stakeholders (Bergmann et al., 2005). In FALKE, research domains that had embraced different epistemic traditions were involved in order to conduct joint research (**Figure 3**). First, each subject-matter didactics discipline had to clarify its position toward explicit teacher-centered explanations. While some publications and empirical research on explanations had already existed (e.g., in the natural sciences and mathematics), explicit teacher-centered explanations seem to play less of a role in other subject areas, both with respect to research concepts and in daily teaching and learning practices. The apparently universal formula, "Explanation leads to understanding," is only partially true with processes studied in a wider sense, for example in argumentation or regarding aesthetic as well as spiritual concepts and practices (see Baumert et al., 2001, for different modes of encountering the world—in German: "Modi der Weltbegegnung"—that are also differently reflected in the respective school subjects).

In FALKE, knowledge generated by pedagogy and psychology about learning and understanding in general (e.g., "cognitive activation," but also methodological concepts such as "operationalization of constructs" or psychometrical quality criteria) had to be discussed with regard to particular subjects and their respective concepts and had to be transferred to the research traditions of the individual subjects. Speech science presented their findings on the performative side of explanations, which in turn influenced the production of the video vignettes. The same is true regarding German linguistics (e.g., with respect to the length of sentences or the avoidance of complex, non-frequent words, etc.). With regard to research methodology, the measurability and operationalization of all general and subject-specific constructs had to be overseen. The cross-subject discourse, however, revealed an extremely fruitful effect of the project in the sense that subject-matter didactics disciplines with a longer history of empirical research helped those from fields newer to evidence-based research practices, which in turn stimulated the former with fresh ideas. And teachers participating in pilot studies also functioned as collaborators by assessing the face validity of the selected contents regarding their relevance to daily teaching and learning practices. The same applies to the students from various schools who also commented on the videos during pilot studies of the different subjects.

Obviously, sharing expertise and adapting concepts is fundamental for a trans-, multi-, and interdisciplinary research project like FALKE. It has become impossible for any individual researcher or any academic discipline to apply and combine all of the research perspectives and knowledge repositories of varied subject-matter didactics, subject-matter knowledge, pedagogical and psychological knowledge, and the methodology of empirical educational research, as well as a practical understanding of teaching.

Even though the processes of teaching and learning come together in a complex event, that occurrence has often only been investigated through the lenses of a limited number of academic disciplines. But working teachers have always strived to combine these different repositories of knowledge in their practical work. As can be seen from a single component of the teaching process such as explanation, these individual perspectives of researchers from different disciplines already lead to a condensation of knowledge about teaching processes. Such amalgamated knowledge can be brought into teacher education more easily, a process that is further facilitated when that knowledge is based on empirical evidence accepted by all the participating disciplines.

Communicative-Psychological Issues

It is no surprise that project groups who work on the basis of shared interests and respect, mutual acceptance, openness and transparency, sympathy, commitment, equality, and a willingness to compromise have a good chance for success (Boehm, 2006). Boehm actually concludes that the quality of the personal relationship has a stronger influence on the strength of the group than do the structures or organization in place. She argues that difficulties of cooperation in interdisciplinary projects are therefore more likely to be rooted in problematic emotional relationships than in the differences between the disciplines (Boehm, 2006).

As already mentioned, the FALKE research program developed from the smaller FALKO group (**Figure 1**), whose members cooperated for many years and could, therefore, look back on a number of joint conference contributions and articles, and on a compendium jointly edited by all participating senior researchers in subject-matter didactics (Krauss et al., 2017). Last but not least, the FALKO group had many meetings both in formal and informal settings. The spirit of this group spread to most of the new members so that cooperation was mostly experienced as an enrichment for both the senior and the junior researchers. It only makes sense to work in a research network if you enjoy attending the meetings and respect the contributions of your colleagues. Overall, trans-, multi-, and interdisciplinary projects require a high degree of personal commitment and mutual tolerance. When individual researchers who have not previously worked together join forces, those projects can entail risks because there is no relationship in place. Another advantage of FALKE was that the junior researchers were not completing doctorates in the same subject area and thus were not in direct competition with each other. Even though there was occasional friction in FALKE, the group remained stable until the end of the project, and most of its members will continue working together in a spirit of trust in the years to come, as is reflected in the ongoing projects FALKE II and FALKE digital (**Figure 1**, above).

CONCLUSION

A key learning outcome of FALKE I is that trans-, multi-, and interdisciplinary projects, in particular, are largely shaped by the nature of the problem, the scientists and stakeholders involved, and the institutional setting (Thompson Klein, 2008). As discussed, explaining is a complex process. It is also an essential component of a teacher's overall educational expertise (i.e., of his or her PCK). There is still little research being done on explaining, partly because different perspectives have to be considered in order to understand this process.

For FALKE, classroom-based teaching had to be investigated, and stakeholders (e.g., experienced teachers and subject-matter didactics specialists like teacher educators) had to be consulted to include their perspectives in an initial step. In the next step, key aspects of explaining needed to be conceptualized in a way in which both domain-specific and general constructs were addressed as a basis for operationalization (i.e., formulation of items that specify the construct). Then videos had to be produced (six per subject) that could be implemented in a computer-based online questionnaire (with items asking for overall and for criteria-based judgments on the didactical quality of the explanations shown). Next, pertinent populations had to be identified whose respective judgments would be of relevance in this context. Corresponding samples had to be recruited, and the study had to be administered. Finally, the data obtained had to be managed, analyzed, and discussed.

Each step of the research process was dependent on the group having reached the required level of knowledge in each field, but also on the group's mutual respect for each other's perspectives. Therefore, it must be considered that working in a transdisciplinary group puts the junior researchers under considerable pressure. This aspect of the work needs to be permanently on the minds of project leaders and subject-specific senior researchers with responsibility for the well-being of academic novices. Hence, to provide a collegial and non-competitive working atmosphere seems to be an essential criterion for long-term successful cooperation. To achieve this, roles must be clearly assigned and the focus of PhD dissertations should also allow for individual pathways to academic qualification.

On a practical, organizational level, project coordination is indispensable for moderating, bundling, and preparing the various decision-making processes for everyone. However, in FALKE, a flat hierarchy was established; for instance, the junior researchers could decide for themselves on the predictors that they wished to operationalize. An alternative would have been a more hierarchical organization with fixed functional roles. It might have actually saved time and energy if more functional roles had been specified and the junior researchers had been less intensively involved in the research design process.

While for senior researchers project management is only one of many tasks, a project coordinator should be at least available to the project most of the time. The same applies to statistical analyses: even if a (small) number of researchers on subject-matter didactics worked with empirical methods already beforehand, the actual data management and the analysis of the overall data set is nevertheless a task that should be handled by one person.

Another lesson learned is that previous cooperation among the researchers on smaller projects leads to a basis of trust that minimizes organizational difficulties because direct communication channels and routines (and ideally even friendships) have already been established. In the follow-up projects FALKE II and FALKE digital (**Figure 1**), a number of the group's members opted to continue this type of research approach in related educational contexts.

In Germany, the establishment of university-based subject-matter didactics disciplines was a first important step toward integrating perspectives on classroom-based teaching and learning. Here, researchers have already built a networked repository of knowledge and research practices for providing evidence-based teaching and teacher education. Not least because the sheer number of international publications and novel insights has expanded enormously but also because we have gained a better grasp of the complexity of educational problems, we now need an overarching trans-, multi-, and interdisciplinary approach to researching subject-matter education. In well-established disciplines, transdisciplinary research projects are often common practice. In educational research and, what is more, in subject-matter didactics, we are only now seeing the beginning of this innovative research and novel opportunities to compete for the necessary funding.

The main advantages of this transdisciplinary approach are the development of a common theoretical framework and the extensive comparability of the results from each subject. We are convinced that the FALKE research program can serve as a noteworthy example for promoting this kind of transdisciplinary educational research. We feel that we were able to prove that it is possible for a group of researchers from eleven different subject-matter didactics—with the addition of researchers from German linguistics and speech science on the one hand and educational research methodology on the other—to meet at a common research starting point and thus contribute to our individual disciplines.

Looking at criteria to evaluate multi-, inter-, or transdisciplinary work (e.g., the degree to which new insights relate to prior disciplinary knowledge in the multiple disciplines involved, the sensible balance reached in weaving disciplinary perspectives together, or the effectiveness with which the integration of disciplines advances understanding and inquiry; Boix-Mansilla, 2006), we made substantial progress (Schilcher et al., 2021). Finally, transdisciplinary (educational) projects allow all researchers to experience the search for knowledge as the guiding and connecting principle of universities.

AUTHOR CONTRIBUTIONS

All authors have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This publication is a result of the KOLEG project (Kooperative Lehrerbildung Gestalten) at the University of Regensburg, which was funded by the German Federal Ministry of Education and Research as part of the joint quality offensive for teacher training by the federal and state governments (Grant No.: 01JA1512).

ACKNOWLEDGMENTS

We would like to thank the numerous students, student teachers, teachers, and teacher trainers of the different school subjects very much for their voluntary participation in the study. We would also like to thank Frances Lorié very much for her helpful comments on the manuscript and its careful proofreading.

REFERENCES

Abraham, U. (2019). Die Germanistik und das Schulfach Deutsch, oder: keine einfache Beschreibung eines komplexen Verhältnisses! *Didaktik Deutsch* 46, 6–12.

Arnold, K.-H. (2012). “Didactics, didactic models and learning,” in *Encyclopedia of the sciences of learning*. Editor N. M. Seel (Heidelberg, Germany: Springer), Vol. 4, 986–990.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., et al. (2001). *PISA 2000 Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen, Germany: Leske + Budrich.

Bergmann, M., Brohmann, B., Hoffmann, E., Schramm, E., and Voß, J. (2005). *Qualitätskriterien transdisziplinärer Forschung. Ein Leitfaden für die formative Evaluation von Forschungsprojekten*. Frankfurt, Germany: Institut für Sozial-Ökologische Forschung (ISOE).

Boehm, B. (2006). *Vertrauensvolle Verständigung—Basis interdisziplinärer Projektarbeit*. Stuttgart, Germany: Steiner.

Boix-Mansilla, V. (2006). Symptoms of quality. Assessing expert interdisciplinary work at the frontier: an empirical exploration. *Res. Eval.* 15 (1), 17–29. doi:10.3152/147154406781776075

Bundesministerium für Bildung und Forschung (BMBF) *Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung der “Qualitätsoffensive Lehrerbildung”*. Available at: https://www.bmbf.de/foerderungen/bekanntmachung-951.html.

Comenius, J. A. (1967). *The great didactic*. Editor M. W. Keatinge (New York, NY: Russel & Russel).

Findeisen, S. *Fachdidaktische Kompetenzen angehender Lehrpersonen*. Wiesbaden: Springer.

Gage, N. L. (1968). Technical Report. “The microcriterion of effectiveness in explaining,” in *Explorations of the teacher’s effectiveness in explaining*. Editors N. L. Gage, M. Belgard, D. Dell, J. E. Hiller, B. Rosenshine, and W. R. Unruh, Vol. 4, 1–8.

Graham, S., and Harris, K. (2005). *Writing better effective strategies for teaching students with learning difficulties*. Baltimore: Brookes.

Hattie, J. A. C. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hayes, J., and Flower, L. (1980). “Identifying the organization of writing processes,” in *Cognitive processes in writing*. Editors L. Gregg and E. Steinberg (Hillsdale: Lawrence Erlbaum), 3–30.

Hill, H. C., Blunck, M. L., Charalambos, Y. C., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. *Cognit. Instruct.* 26 (4), 430–511. doi:10.1080/07370000802177235

Hill, H. C., Rowan, B., and Loewenberg Ball, D. (2005). Effects of teachers’ mathematical knowledge for teaching on student achievement. *Am. Educ. Res. J.* 42 (2), 371–406. doi:10.3102/00028312042002371

Kansanen, P. (2002). Didactics and its relation to educational psychology: problems in translating a key concept across research communities. *Int. Rev. Educ.* 48 (6), 427–441. doi:10.1023/A:1021388816547

Kirschner, S., Sczudlek, M., Tepner, O., Borowski, A., Fischer, H. E., Lenske, G., et al. (2017). “Professionswissen in den Naturwissenschaften (ProwiN),” in *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven*. Editors C. Gräsel and K. Trempler (Wiesbaden, Germany: Springer VS), 113–130.

KMK (2019). Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Available at: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf.

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., et al. (2020). Competence as a continuum in the COACTIV study: the “cascade model.” *ZDM Math. Edu.* 52, 311–327. doi:10.1007/s11858-020-01151-z

Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Goehring, A., Hofmann, B., et al. (2017). *FALKO—Fachspezifische Lehrerkompetenzen: Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Münster, Germany: Waxmann.

Kulgemeyer, C., and Peters, C. H. (2016). Exploring the explaining quality of physics online explanatory videos. *Eur. J. Phys.* 37 (6), 1–14. doi:10.1088/0143-0807/37/6/065705

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., and Neubrand, M. (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers*. New York, NY: Springer.

Leutner, D., Fleischer, J., Grünkorn, J., and Klieme, E. (2017). “Competence assessment in education: an inroduction,” in *Competence assessment in education (methodology of educational measurement and assessment)*, Editors D. Leutner, J. Fleischer, J. Grünkorn, and E. Klieme (Cham, Germany: Springer International Publishing), 1–6.

Lindl, A., Gaier, L., Weich, M., Frei, M., Ehras, C., Gastl-Pischetsrieder, M., et al. (2019). “Eine ‘gute‘ Erklärung für alle?! Gruppenspezifische Unterschiede in der Beurteilung von Erklärqualität—erste Ergebnisse aus dem interdisziplinären Forschungsprojekt FALKE,” in *Lehrer. Bildung. Gestalten. Beiträge zur empirischen Forschung in der Lehrerbildung*. Editors T. Ehmke, P. Kuhl, and M. Pietsch (Weinheim, Germany: Beltz), 128–141.

Lindl, A., Krauss, S., Schilcher, A., and Hilbert, S. (2020). Statistical methods in transdisciplinary educational research. *Front. Edu.* 5 (97). doi:10.3389/feduc.2020.00097

Mittelstrass, J. (2011). On transdisciplinarity. *TRAMES* 15 (65/60), 329–338. doi:10.3176/tr.2011.4.01

Odora, R. J. (2014). Using explanation as a teaching method: how prepared are high school technology teachers in free state province, South Africa? *J. Soc. Sci.* 38 (1), 71–81. doi:10.1080/09718923.2014.11893238

OECD (2019). *PISA 2018 results. What students know and can do*. Paris, France: OECD Publishing.

Pohl, C., van Kerkhoff, L., Hirsch Hadorn, G., and Bammer, G. (2008). “Integration,” in *Handbook of transdisciplinary research.* Editors G. Hirsch Hadorn, S. Hoffmann-Riem, W. Biber-Klemm, D. Grossenbacher-Mansuy, and C. Joye (Berlin, Germany: Springer), 411–424.

Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., and Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum—zwischen Generik und Fachspezifik. *Unterrichtswissenschaft* 48, 409–446. doi:10.1007/s42010-020-00082-8

Praetorius, A-K., Klieme, E., Herbert, B., and Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM Math. Edu.* 50, 407–426. doi:10.1007/S11858-018-0918-4

Rosenfield, P. L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Soc. Sci. Med.* 35 (11), 1343–1357. doi:10.1016/0277-9536(92)90038-R

Rothgangel, M., and Vollmer, J. (2020). Towards a theory of subject-matter didactics. *Res. Subject-Matter Teach. Learn* 3, 126–151. doi:10.23770/rt1838

Scheffel, L. (2019). *Erklären im Unterricht: Taktiken und Strategien*. Weinheim, Germany: Beltz.

Schiefele, U., Schaffner, E., Möller, J., and Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Read. Res. Q* 47 (4), 427–463. doi:10.1002/RRQ.030

Schilcher, A., Krauss, S., Lindl, A., and Hilbert, S. (2021). FALKE – Fachspezifische Lehrerkompetenzen im Erklären: Untersuchungen zur Beurteilung und zu Kriterien unterrichtlicher Erklärqualität aus der Perspektive von 13 Fachbereichen. *Weinheim/Basel, Germany: Beltz Juventa*.

Schilcher, A., and Wild, J. (2018). "Lesen. Eine Schlüsselkompetenz im interdisziplinären Forschungsdiskurs," in *Empirische Forschung in der Deutschdidaktik*. Editor J. M. Boelmann (Baltmannsweiler, Germany: Schneider), Vol. 3, 163–184.

Schneidewind, U. (2010). Ein institutionelles Reformprogramm zur Förderung transdisziplinärer Nachhaltigkeitsforschung. *GAIA* 19 (2), 122–128.

Seidel, T., and Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Rev. Educ. Res*. 77 (4), 454–499. doi:10.3102/0034654307310317

Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harv. Educ. Rev*. 57 (1), 1–22. doi:10.17763/haer.57.1.j463w79r56455411

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educ. Res* 15 (2), 4–14. doi:10.2307/1175860

Stanovich, K. (1991). "Word recognition: changing perspectives," in *Handbook of reading research*. Editors R. Barr, M. Kamil, P. Mosenthal, and P. D. Pearson (New York, NY: Longman), Vol. 2, 418–452.

Thompson Klein, J. (2008). "Integration in der inter- und transdisziplinären Forschung," in *Transdisziplinäre Forschung. Integrative Forschungsprozesse verstehen und bewerten*. Editors M. Bergmann and E. Schramm (Frankfurt, Germany: Campus Verlag), 93–116.

Turnbull, E. R., Pineo, H., and Aldridge, R. W. (2019). Improving the health of the public: a transdisciplinary research study. *Lancet* 394, S93. doi:10.1016/S0140-6736(19)32890-9

V. Hoffmann, A. Thomas, and A. Gerber (Editors) (2009). *Transdisziplinäre Umweltforschung. Methodenhandbuch.*München, Germany: Oekom.

Woessmann, L. (2016). The economic case for education. *Educ. Econ.* 24 (1), 3–32. doi:10.1080/09645292.2015.1059801

Wragg, E. C., and Wood, E. K. (1984). "Teachers' first encounters with their classes," in *Classroom teaching skills*. Editor E. C. Wragg (London, United Kingdom: Croom Helm), 47–78.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors, SH.

frontiers in Psychology

OPINION
published: 11 May 2021
doi: 10.3389/fpsyg.2021.562658

Cross-Disciplinary Research on Learning and Instruction – Coming to Terms

Nicole Heitzmann[1]*, Ansgar Opitz[1], Matthias Stadler[1], Daniel Sommerhoff[2], Maximilian C. Fink[3], Andreas Obersteiner[4], Ralf Schmidmaier[3,5], Birgit J. Neuhaus[6], Stefan Ufer[7], Tina Seidel[4], Martin R. Fischer[3] and Frank Fischer[1]

[1] Department of Psychology, LMU Munich, Munich, Germany, [2] Department of Mathematics Education, Leibniz Institute for Science and Mathematics Education, Kiel, Germany, [3] Institute of Medical Education, University Hospital, LMU Munich, Munich, Germany, [4] TUM School of Education, Technical University of Munich, Munich, Germany, [5] LMU University Hospital, Medizinische Klinik und Poliklinik IV, Munich, Germany, [6] Biology Education, Faculty of Biology, LMU Munich, Munich, Germany, [7] Mathematics Education, LMU Munich, Munich, Germany

Keywords: conceptualization, cross-disciplinary research, collaborative problem solving, transdisciplinary research, interdisciplinary research, joint theoretical framework, joint methodological approach

OPEN ACCESS

Edited by:
Bernhard Ertl,
Munich University of the Federal Armed Forces, Germany

Reviewed by:
Paul Leon Van Geert,
University of Groningen, Netherlands
Michael O'Rourke,
Michigan State University, United States

****Correspondence:***
Nicole Heitzmann
nicole.heitzmann@psy.lmu.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *15 May 2020*
Accepted: *08 April 2021*
Published: *11 May 2021*

Citation:
Heitzmann N, Opitz A, Stadler M, Sommerhoff D, Fink MC, Obersteiner A, Schmidmaier R, Neuhaus BJ, Ufer S, Seidel T, Fischer MR and Fischer F (2021) Cross-Disciplinary Research on Learning and Instruction – Coming to Terms. Front. Psychol. 11:562658. doi: 10.3389/fpsyg.2021.562658

CROSS-DISCIPLINARY RESEARCH COLLABORATIONS

Research in universities and other organizations is often conducted within established disciplines that are historically based and highly arbitrary (Campbell, 2014). However, emergent phenomena fail to fit into disciplinary boundaries, making cross-disciplinary research necessary, often involving corresponding collaboration (Hall et al., 2008).

One area of research involving complex phenomena that cannot be well addressed by one discipline alone is learning and instruction in higher education. Higher education programs aim to teach professional knowledge to students as a prerequisite for their later professional activities (Blömeke et al., 2015). For example, in teacher education programs usually focus on content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical-psychological knowledge (PK) (see Shulman, 1987). In order to teach such knowledge, it seems reasonable and is increasingly common that psychologists and educational scientists, in addition to experts in the subject matter domains, are involved in designing study programs. Similarly, it also seems reasonable to involve researchers from these various domains for conducting research on how to facilitate teaching in higher education programs. Thus, cross-disciplinary collaboration is the rule rather than the exception in higher education *practice* and is becoming increasingly common in *research* on higher education. An example for a cross-disciplinary research endeavor in learning and instruction is a research unit on facilitating diagnostic competences in simulation-based learning environments in the university context in which researchers from subject matter domains (biology education, mathematics education, and medical education) are working together with researchers from education and from educational psychology[1].

Even though there is a decent amount of research on cross-disciplinarity, for example from the science of team science (Hall et al., 2018, 2019), there is only limited research on cross-disciplinarity in the field of learning and instruction, and especially on collaborative processes. In this opinion article, we claim that ideas and concepts from the field of collaborative problem solving have the potential to yield valuable insights when designing or conducting cross-disciplinary research in learning and instruction.

*All authors are part of various cross-disciplinary large scale projects such as research unit COSIMA (https://www.for2385.lmu.de) or international doctoral school REASON (http://www.en.mcls.lmu.de/study_programs/reason).

[1]COSIMA website: https://www.for2385.lmu.de

CONCEPTUALIZATION OF CROSS-DISCIPLINARY RESEARCH ENDEAVORS

There is substantial evidence on some specific features that positively influence cross-disciplinary research collaborations, such as team formation, team composition, or institutional factors (e.g., Epstein, 2014; O'Donnell and Derry, 2014; Hall et al., 2018, 2019). However, it remains unclear how prerequisites such as the intended form of the cross-disciplinary collaboration influence the collaborative problem-solving process, and second, how the collaborative problem-solving process itself influences and is influenced by other factors such as aspects of the cross-disciplinary team or the production of joint artifacts.

We introduce a conceptualization of how ideas and concepts from the field of collaborative problem solving are useful to address challenges that arise from cross-disciplinary research (see **Figure 1**). The conceptualization is based on existing approaches to cross-disciplinary research (e.g., Epstein, 2014; O'Donnell and Derry, 2014; Hall et al., 2018, 2019) and extends these approaches by introducing processes and skills from collaborative problem solving (Hao and Mislevy, 2019; Hao et al., 2019).

The basis of our conceptualization are the three different forms of cross-disciplinary research that are commonly differentiated: multidisciplinary, interdisciplinary, and transdisciplinary (e.g., Lattuca, 2003; Slatin et al., 2004; Collin, 2009; Hall et al., 2012; Klein, 2017). Which form of cross-disciplinary research is intended, can have an influence on the collaborative problem-solving process in the way that it sets the stage for which collaborative problem-solving skills are of major importance. Collaborative problem solving builds the core of our conceptualization. We discuss how factors of the cross-disciplinary team reciprocally influence the processes of collaborative problem solving and how the collaborative problem-solving process itself and the development of joint artifacts influence each other. The environment, in which a cross-disciplinary research endeavor takes place, surrounds the other elements of the conceptualization building another important factor to consider in cross-disciplinary research in learning and instruction.

Form of Cross-Disciplinary Collaboration

Forms of cross-disciplinary collaboration differ in their collaborative problem-solving process and build thus the basis for the conceptualization. Three forms that are commonly differentiated are *multidisciplinary* research, *interdisciplinary* research, and *transdisciplinary* research (e.g., Lattuca, 2003; Slatin et al., 2004; Collin, 2009; Hall et al., 2012; Klein, 2017). However, so far there is no agreed upon definition for each form (e.g., Hall et al., 2008). For the purpose of our analysis, we use the following differentiations (Klein, 2017): In *multidisciplinary* research, different disciplines work on different aspects of a problem independently within their disciplinary boundaries. Researchers from different disciplines contribute specific knowledge and skills with the goal to address a certain phenomenon or issue from multiple perspectives. In *interdisciplinary* research, existing disciplinary approaches are restructured and integrated in order to address a problem relevant for all participating disciplines. Interdisciplinary research can be seen as a spectrum reaching from researchers borrowing concepts and methods from other disciplines to answer a specific research question up to the development of new frameworks that are valid across disciplines (Pohl et al., 2021). Researchers share their knowledge and then identify which concepts or methods from the other disciplines are necessary for answering research questions within their own discipline or that go beyond their own disciplinary boundaries. In interdisciplinary teams, researchers' still focus on their own disciplines even though disciplinary boundaries are crossed to some degree to make the points of contact between the disciplines compatible (Choi and Pak, 2006). *Transdisciplinary* research also seeks to integrate different lines of work from contributing disciplines (Klein, 2010; Pohl, 2010). A key aspect of transdisciplinary research is the collaborative co-production of knowledge from researchers from different disciplines, and possibly also stakeholders from private or public sectors with the goal to solve societal problems (Pohl et al., 2021). Whereas in interdisciplinary research actions in the collaborative process are described with linking, blending, fusing, and synthesizing, actions in transdisciplinary research are transcending, transgression, and transforming (Klein, 2010). Disciplinary boundaries can be challenged on purpose in the process of transdisciplinary research (Pohl et al., 2021). Whereas the current discourse on cross-disciplinary research distinguishes between three discrete forms (multidisciplinary, interdisciplinary, and transdisciplinary), there are considerations that place them on a continuum (Mennes, 2020).

Collaborative Problem Solving

We want to make the claim that even though cross-disciplinary research in learning and instruction can be considered through the lens of collaborative problem solving, the intended form of cross-disciplinary collaboration can influence the role that collaborative problem solving plays in that process. Main aspects of collaborative problem solving important for cross-disciplinary research are collaborative problem-solving skills and different roles to help stimulate the problem-solving process.

Collaborative problem solving involves cognitive skills, such as defining the problem at hand and social skills, such as establishing a shared understanding (Graesser et al., 2018). Regarding the collaborative problem-solving process, four skills are considered to be of major relevance (Liu et al., 2016; Hao and Mislevy, 2019): (1) *Sharing ideas* refers to how individuals bring divergent ideas into a collaborative process (Liu et al., 2016). (2) *Negotiating ideas* refers to building collaborative knowledge and constructing processes within a group. Negotiating occurs by comparing alternative ideas and their associated evidence. Subprocesses of negotiating ideas include agreeing, disagreeing, requesting clarification, elaborating on each other's ideas, and identifying gaps (Liu et al., 2016). Collaborative team knowledge is produced in this process (Liu et al., 2016). (3) *Regulating problem-solving activities* is a social skill that refers to the coordination of discourse within a team. An example is to highlight the goal of a discussion, such as finding an up-to-date instrument to measure

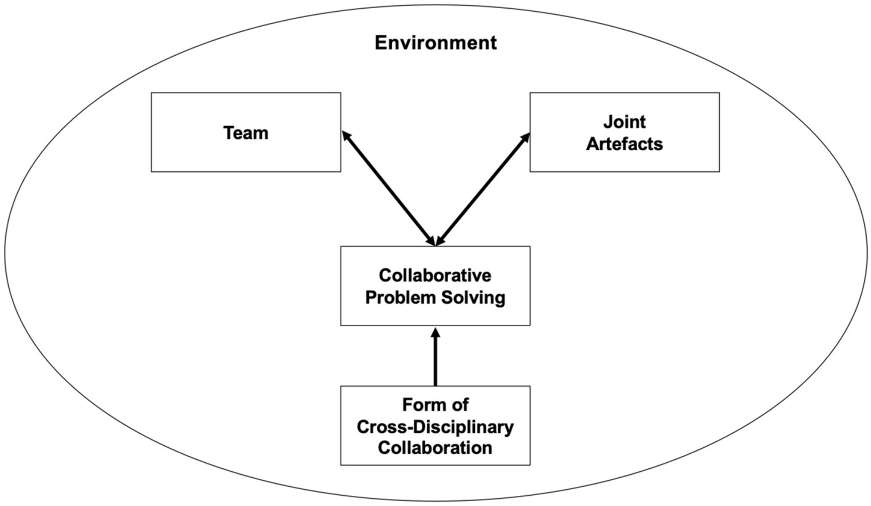

FIGURE 1 | Conceptualization of cross-disciplinary research in learning and instruction.

motivation. An important aspect regarding the regulation of problem-solving activities is that members' individual ideas about what collaboration looks like might differ more in cross-disciplinary projects than in mono-disciplinary projects. External guidance might be needed to ensure successful collaboration (von Wehrden et al., 2019). (4) The social skill of *maintaining conversation* refers to communication that is not directly topic-related but maintains a positive atmosphere (Liu et al., 2016). This kind of non-topic-related communication seems to be of major importance in cross-disciplinary teams in order to support the collective communication competence of the team (Thompson, 2009). Research on cross-disciplinary research collaborations from other fields suggests examining how the involved disciplines differ in their way of collaborative problem solving and communicating and then providing enough guidance while still offering enough possibilities for participation in all collaborative problem-solving processes (König et al., 2013).

Depending on the form of cross-disciplinary collaboration, different collaborative problem-solving skills seem to be central. In a cross-disciplinary research unit in learning and instruction, regulating the problem-solving process is central for multidisciplinary goals. This importance is based on the fundamentally different perspectives on the same problem by researchers from different disciplines, e.g., subject matter didactics, educational psychology, and educational science. In addition to the need to regulate problem-solving processes within the team externally, coordinating resources that exist in the different disciplines and defining interfaces might be necessary. For example, it might be important to organize and moderate meetings in which different disciplinary perspectives on a joint problem can be juxtaposed. For interdisciplinary goals, sharing knowledge across disciplines seems particularly important in addition to regulating the process (see Liu et al., 2016). For interdisciplinary and transdisciplinary goals, negotiating can be considered a specifically important skill for grounding and finding a shared language across disciplines (Bromme, 2000). Based on these examples, we hypothesize that each form of cross-disciplinary collaboration (multidisciplinary, interdisciplinary, and transdisciplinary) requires unique collaborative problem solving and communication skills, because they differ in their main goals as well as in the means of achieving and communicating these goals.

Possibly, it can be beneficial for the definition of specific working routines, such as for the development of learning environments, to assign different collaborative problem-solving activities to different roles. Roles can be conceptualized with reference to internal collaboration scripts. Internal collaboration scripts are mental schemas that typically include a set of roles and associated activities (Fischer et al., 2013). These internal scripts may differ widely across disciplines. For example, the collaboration script in one discipline can involve that junior researchers first formulate a draft for a manuscript and later senior researchers comment on that draft. In other disciplines, junior researchers might be involved at other stages of the publication process. Therefore, making the task of specific roles explicit during interactions within the team seems important.

The regulation of the problem-solving process should be assigned to the role of a facilitator who mediates between actors from different disciplines (see also Bammer, 2016; Salazar et al., 2019). The facilitator can take over processual leadership tasks to ensure that the interactions between team members are productive (Gray, 2008). In order to support the development of joint artifacts, it seems reasonable to spend resources on a facilitator with their own research experience at least on the post-doc level.

Team

When building a cross-disciplinary research team, the science of team science has already described important aspects for team composition and team formation (e.g., Hall et al., 2018, 2019). We focus on aspects of collaboration that are in close connection to collaborative problem solving. These aspects include overlapping expertise within the team, a strategy for publications, and a clear shared goal.

A deep understanding of more than one discipline is difficult to achieve (Pohl and Hadorn, 2008). Most research teams have to engage in collaborative problem solving between various researchers with deep discipline specific knowledge. Campbell (2014) uses the metaphor of a fish's scales to describe the composition of successful cross-disciplinary teams. In his model, each fish scale symbolizes one individual with a unique set of expertise. In order to build a successful team, each "fish scale" has to overlap to a certain degree with the neighboring fish scales. There are fish scales that are close to each other and others that are further apart. Those further apart from each other are not directly connected but are indirectly connected via the other fish scales. What can be drawn from Campbell's (2014) metaphor is that it is not necessary that researchers from all disciplines collaborate directly in a collaborative problem solving process, which would be highly laborious; rather, they may also be connected via researchers from other disciplines.

In research on learning and instruction it seems likely that the "connecting fish scale" is represented by researchers from the educational sciences or educational psychology because these disciplines are concerned with learning in general. For example, in the research unit on facilitating diagnostic competences in simulation-based learning environments researchers from mathematics education and medical education did not have a direct link at first. These two groups of researchers were only indirectly connected via their collaboration with the field of psychology. It seems possible that researchers from the connecting fish scale can have a major influence on the collaborative problem-solving process because they play a major role in regulating the problem-solving process.

A major challenge of cross-disciplinary teams is the lack of an adequate joint reward system during the collaborative problem solving process (O'Donnell and Derry, 2014). Within disciplinary boundaries it is relatively clear how much a publication in a journal, book, or conference proceedings will benefit a researcher's career. For example, publications in conference proceedings are typically less valued than international journal publications for an educational psychologist. However, the value of a publication becomes less clear when it appears outside of a researcher's disciplinary boundaries or in an interdisciplinary journal. Furthermore, joint publications face additional problems such as over-inclusive authorship (Elliott et al., 2017; Settles et al., 2018) or what disciplines see as reliable epistemic processes or epistemic ideals (Chinn et al., 2011). The entire meaning of collaboration in a team of authors varies across disciplines. An exclusive focus on cross-disciplinary publications may be particularly problematic for young researchers, whose goal is to develop a record and profile of expertise *within* their disciplinary field. It seems even reasonable to suggest that young researchers should be encouraged to submit their first manuscripts primarily to disciplinary journals.

For cross-disciplinary research in learning and instruction, it is a major challenge to identify phenomena and questions that allow for research that is relevant or even cutting edge in all of the participating disciplines (e.g., Epstein, 2014). Examples of participating disciplines in learning and instruction are psychology, education, and various subject matter didactics such as mathematics education or biology education. In order to have interdisciplinary and transdisciplinary goals in a research endeavor in learning and instruction, it seems crucial to identify a phenomenon that makes integration of concepts and methods from different disciplines necessary. A helpful method for defining such goals may be integrating question that bring together different avenues of inquiry (Cosens et al., 2011).

Joint Artifacts

Another major aspect for cross-disciplinary research in relation with collaborative problem solving is the development of joint artifacts. O'Donnell and Derry (2014) stress the importance of artifacts, which they call *tools*. For research on learning and instruction in higher education it seems characteristic that different concepts, methods, and technologies are used in the subject matter domains (e.g., biology or mathematics), in psychology, and in educational science. Therefore, it seems reasonable to suggest the development of three types of joint artifacts early in the collaborative problem-solving process in order to identify possible barriers but also potentials for innovation: a joint conceptual framework, a joint methodological framework, and a joint technological framework. In order to develop such artifacts it seems advisable to include an overarching coordination mechanism that ensures methodological and conceptual standardization and progress (see König et al., 2013). The development of joint artifacts can be of major relevance for collaborative problem-solving processes, such as information sharing and negotiating.

- A joint conceptual framework can identify relevant theoretical ideas and their interconnections. It can ensure that common ground exists and that terms are defined precisely.
- A joint methodological framework refers to methods and more detailed research practices. A precise description of methods is important because methods and best practices vary between disciplines. What is considered a gold standard in one discipline can be seen as less important in another discipline; for example, an empirical-experimental approach is difficult to combine with hermeneutic methods.
- A joint technological framework defines the technology relevant for collaboration and for addressing the research questions. Every discipline in the context of learning and instruction has its own set of preferred research technologies, for example simulations that create extensive logfiles to measure and facilitate learning (Fink et al., 2020). Joint technologies may help to integrate data from different research projects, and later transfer the results into practice. In order to have a suitable technology for learning, it can be necessary for researchers to develop their own software.

Environment

The last aspect in our conceptualization of cross-disciplinary research in learning and instruction is the environment that surrounds the other aspects. In connection with cross-disciplinary collaborations there are various environmental factors such as societal and political factors that influence whether a research endeavor will receive attention and funding. In this section we focus on a factor that researchers can influence to a certain degree: the institutional climate.

The institutional climate refers to the perceptions, attitudes, and expectations of an institution toward cross-disciplinary research. Epstein (2014) argues that the institutional climate can support horizontal, cross-disciplinary structures that allow researchers to cluster around phenomena. As the institutional climate in many academic institution may only change slowly and gradually, it can take years of preparation and the completion of smaller projects to develop a sound environment for a research collaboration. In particular, it may only marginally be susceptible to individual members of the institution, making joint efforts and initiatives necessary. Thus, it seems reasonable to plan enough time for preparing both capacity as well as the environment for the actual research endeavor. It seems advisable to start with a smaller-scale project, such as the joint supervision of a single Ph.D. project or a joint publication. A well prepared institutional climate might also be beneficial for collaborative problem solving and particularly for maintaining conversation.

DISCUSSION

Cross-disciplinary research collaborations in the context of learning and instruction are of critical importance to address the complex problems of 21st century education. However, many promising projects fail beyond the actual research conducted due to avoidable issues (Fam and O'Rourke, 2021). The research reviewed here allows for formulating reasonable hypotheses about favorable processes and conditions with a psychological focus from the perspective of collaborative problem solving. These hypotheses may support scientific achievements such as the use of pilot projects, the early development of joint artifacts, conceptual, methodological, and technical frameworks, or the role of an experienced facilitator supporting the collaborative problem-solving process through intellectual grounding, coordination and negotiation. Whether and under which conditions these hypotheses are valid for cross-disciplinary research collaborations on learning and instruction and beyond remains an open empirical question. In further research the theoretical foundation as well as the relationship between the four aspects of our proposed conceptualization should be further expanded and specified using theories on science and technology studies (e.g., Hackett et al., 2008), actor-network theory (e.g., Latour, 1996), or theories on complex systems (e.g., Stacey, 1995). We believe our proposed conceptualization based on theoretical considerations and on our own experiences in a cross-disciplinary research unit on facilitating diagnostic competence in simulation-based learning environments can provide helpful terminology and some theory-inspired heuristics on how to realize the great potentials and to avoid the stumbling blocks when attempting the challenging task of cross-disciplinary research collaboration in learning and instruction.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research for this article was funded by the German Research Association (Deutsche Forschungsgemeinschaft, DFG) (FOR2385, FI 792/12-2).

REFERENCES

Bammer, G. (2016). Negotiating boundaries, leadership, and integration and implementation sciences (I2S). *J. Environ. Stud. Sci.* 6, 432–436. doi: 10.1007/s13412-015-0325-x

Blömeke, S., Gustafsson, J.-E., and Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift Psychol.* 223, 3–13. doi: 10.1027/2151-2604/a000194

Bromme, R. (2000). "Beyond one's own perspective: the psychology of cognitive interdisciplinarity," in *Practicing Interdisciplinarity*, eds P. Weingart and N. Stehr (Toronto: Toronto University Press), 115–133. doi: 10.3138/9781442678729-008

Campbell, D. T. (2014). "Ethnocentrism of disciplines and the fish-scale model of omniscience," in *Interdisciplinary Collaboration: An Emerging Cognitive Science*, eds S. J. Derry, C. D. Schunn, and M. A. Gernsbacher (Hove: Psychology Press), 3–21.

Chinn, C. A., Buckland, L. A., and Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: arguments from philosophy and psychology. *Educ. Psychol.* 46, 141–167. doi: 10.1080/00461520.2011.587722

Choi, B. C. K., and Pak, A. W. P. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clin. Invest. Med.* 29, 351–364.

Collin, A. (2009). Multidisciplinary, interdisciplinary, and transdisciplinary collaboration: implications for vocational psychology. *Int. J. Educ. Vocat. Guid.* 9, 101–110. doi: 10.1007/s10775-009-9155-2

Cosens, B., Fiedler, F., Boll, J., Higgins, L., Johnson, G., Kennedy, B., et al. (2011). Interdisciplinary methods in water resources. *Issues Integr. Stud.* 29, 118–143.

Elliott, K. C., Settles, I. H., Montgomery, G. M., Brassel, S. T., Cheruvelil, K. S., and Soranno, P. A. (2017). Honorary authorship practices in environmental science teams: structural and cultural factors and solutions. *Account. Res.* 24, 80–98. doi: 10.1080/08989621.2016.1251320

Epstein, S. L. (2014). "Making interdisciplinary collaboration work," in *Interdisciplinary Collaboration: An Emerging Cognitive Science*, eds S. J. Derry, C. D. Schunn, and M. A. Gernsbacher (Hove: Psychology Press), 245–263.

Fam, D., and O'Rourke, M. (eds) (2021). *Interdisciplinary and Transdisciplinary Failures: Lessons Learned from Cautionary Tales*. London: Routledge.

Fink, M. C., Radkowitsch, A., Bauer, E., Sailer, M., Kiesewetter, J., Schmidmaier, R., et al. (2020). Simulation research and design: a dual-level framework for multi-project research programs. *Educ. Technol. Res. Dev.* doi: 10.1007/s11423-020-09876-0

Fischer, F., Kollar, I., Stegmann, K., and Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educ. Psychol.* 48, 56–66. doi: 10.1080/00461520.2012.748005

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., and Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychol. Sci. Publ. Interest* 19, 59–92. doi: 10.1177/1529100618808244

Gray, B. (2008). Enhancing transdisciplinary research through collaborative leadership. *Am. J. Prev. Med.* 35(2 Suppl.), 124–132. doi: 10.1016/j.amepre.2008.03.037

Hackett, E. J., Amsterdamska, O., Lynch, M., and Wajcman, J. (2008). *The Handbook of Science and Technology Studies*. Cambridge, MA: MIT Press.

Hall, K. L., Feng, A. X., Moser, R. P., Stokols, D., and Taylor, B. K. (2008). Moving the science of team science forward: collaboration and creativity. *Am. J. Prev. Med.* 35(2. Suppl.), S243–S249. doi: 10.1016/j.amepre.2008.05.007

Hall, K. L., Vogel, A. L., and Croyle, R. T. (eds) (2019). *Strategies for Team Science Success: Handbook of Evidence-Based Principles for Cross-Disciplinary Science and Practical Lessons Learned from Health Researchers*. Berlin: Springer Nature.

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., et al. (2018). The science of team science: a review of the empirical evidence and research gaps on collaboration in science. *Am. Psychol.* 73, 532–548. doi: 10.1037/amp0000319

Hall, K. L., Vogel, A. L., Stipelman, B. A., Stokols, D., Morgan, G., and Gehlert, S. (2012). A four-phase model of transdisciplinary team-based research: goals, team processes, and strategies. *Transl. Behav. Med.* 2, 415–430. doi: 10.1007/s13142-012-0167-y

Hao, J., Liu, L., Kyllonen, P., Flor, M., Davier, A. A., and von. (2019). Psychometric considerations and a general scoring strategy for assessments of collaborative problem solving. *ETS Res. Rep. Ser.* 2019, 1–17. doi: 10.1002/ets2.12276

Hao, J., and Mislevy, R. J. (2019). Characterizing interactive communications in computer-supported collaborative problem-solving tasks: a conditional transition profile approach. *Front. Psychol.* 10:1011. doi: 10.3389/fpsyg.2019.01011

Klein, J. T. (2010). "A taxonomy of interdisciplinarity," in *The Oxford Handbook of Interdisciplinarity*, eds R. Frodeman, J. T. Klein, and R. C. S. Pacheco (Oxford: Oxford University Press) 15–30. doi: 10.4324/9781315177298-3

Klein, J. T. (2017). "Typologies of interdisciplinarity: the boundary work of definition," in *The Oxford Handbook of Interdisciplinarity*, eds R. Frodeman, J. T. Klein, and R. C. S. Pacheco (Oxford: Oxford University Press), 21–34.

König, B., Diehl, K., Tscherning, K., and Helming, K. (2013). A framework for structuring interdisciplinary research management. *Res. Policy* 42, 261–272. doi: 10.1016/j.respol.2012.05.006

Latour, B. (1996). On actor-network theory: a few clarifications. *Soziale Welt* 47, 369–381.

Lattuca, L. R. (2003). Creating interdisciplinarity: grounded definitions from the College and University Faculty. *Hist. Intell. Cult.* 3. Available online at: https://journalhosting.ucalgary.ca/index.php/hic/article/view/68806 (accessed January 1, 2021).

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., and Zapata-Rivera, J.-D. (2016). "A tough nut to crack: measuring collaborative problem solving," in *Handbook of Research on Technology Tools for Real-World Skill Development*, eds Y. Rosen, S. Ferrara, and M. Mosharraf (Hershey, PA: IGI Global), 344–359. doi: 10.4018/978-1-4666-9441-5.ch013

Mennes, J. (2020). Putting multidisciplinarity (back) on the map. *Eur. J. Philos. Sci.* 10:18. doi: 10.1007/s13194-020-00283-z

O'Donnell, A. M., and Derry, S. J. (2014). "Cognitive processes in interdisciplinary groups: problems and possibilities," in *Interdisciplinary Collaboration: An Emerging Cognitive Science*, eds S. J. Derry, C. D. Schunn, and M. A. Gernsbacher (Hove: Psychology Press), 51–82.

Pohl, C. (2010). From transdisciplinarity to transdisciplinary research. *Transdiscipl. J. Eng. Sci.* 1, 65–73. doi: 10.22545/2010/0006

Pohl, C., and Hadorn, G. H. (2008). ""Core terms in transdisciplinary research,"," in *Handbook of Transdisciplinary Research*, eds G. H. Hadorn, H. Hoffmann-Riem, S. Biber-Klemm, W. Grossenbacher-Mansuy, D. Joye, C. Pohl, et al. (Berlin: Springer Science + Business Media B.V), 427–432. doi: 10.1007/978-1-4020-6699-3_28

Pohl, C., Klein, J. T., Hoffmann, S., Mitchell, C., and Fam, D. (2021). Conceptualising transdisciplinary integration as a multidimensional interactive process. *Environ. Sci. Policy* 118, 18–26. doi: 10.1016/j.envsci.2020.12.005

Salazar, M. R., Widmer, K., Doiron, K., and Lant, T. K. (2019). "Leader integrative capabilities: a catalyst for effective interdisciplinary teams," in *Strategies for Team Science Success: Handbook of Evidence-Based Principles for Cross-Disciplinary Science and Practical Lessons Learned from Health Researchers*, eds K. L. Hall, A. L. Vogel, and R. T. Croyle (Berlin: Springer International Publishing), 313–328. doi: 10.1007/978-3-030-20992-6_24

Settles, I. H., Brassel, S. T., Montgomery, G. M., Elliott, K. C., Soranno, P. A., and Cheruvelil, K. S. (2018). Missing the mark: a new form of honorary authorship motivated by desires for inclusion. *Innov. High. Educ.* 43, 303–319. doi: 10.1007/s10755-018-9429-z

Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educ. Rev.* 57, 1–23. doi: 10.4324/9781351233866-1

Slatin, C., Galizzi, M., Melillo, K. D., and Mawn, B. (2004). Conducting interdisciplinary research to promote healthy and safe employment in health care: promises and pitfalls. *Publ. Health Rep.* 119, 60–72. doi: 10.1177/003335490411900112

Stacey, R. D. (1995). The science of complexity: an alternative perspective for strategic change processes. *Strat. Manag. J.* 16, 477–495. doi: 10.1002/smj.4250160606

Thompson, J. L. (2009). Building collective communication competence in interdisciplinary research teams. *J. Appl. Commun. Res.* 37, 278–297. doi: 10.1080/00909880903025911

von Wehrden, H., Guimarães, M. H., Bina, O., Varanda, M., Lang, D. J., John, B., et al. (2019). Interdisciplinary and transdisciplinary research: finding the common ground of multi-faceted concepts. *Sustain. Sci.* 14, 875–888. doi: 10.1007/s11625-018-0594-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership