Blueprint for Inclusive Research and Development in Education (BIRD-E)

Comprehensive Report: Researcher Working Group

May 2022

A project by InnovateEDU

Support provided by:

Table of Contents

Overview & background

In the field of education, there is a gap between research and practice. While research is produced with the intent to provide meaningful improvements in the classroom, educators are often unaware such evidence exists. The lack of accessible and relevant evidence leads to suboptimal student outcomes, especially in historically underserved communities. Additionally, the language practitioners and researchers use differs, hindering the research's ability to be translated and implemented. The lack of replication and validation in the field can lead to misleading conclusions and practices.

Additionally, teachers generally do not trust reports created by solution providers, as providers have a stake in the product's success. Educators rely on other educators for suggestions on best practices and evidence-based interventions. Penuel et al. (2016) implemented a survey-based study on research use among K-8 instructional policymakers sampled from urban schools and central offices across the nation in the field of education. The study revealed several important findings about how educational research is used by school and district leaders. Notably, most respondents cited using research for decision-making but are most likely to access research through professional associations and conferences rather than individual researchers or U.S. Department of Education databases, including What Works Clearinghouse. Education research and practitioner surveys demonstrate that peer network reliance is a strong practice in education. Thus, it is important to make peer networks knowledgeable about interventions that work for whom and under what conditions. The research study reported positive attitudes about research usage. Still, respondents felt they had limited access to the most impactful research for their classroom needs and largely doubted their abilities to interpret the research results.

A potential solution to the challenges faced by educators is the adoption of a conceptual framework that validates research and evidence to allow enhanced comparisons between educational interventions. This helps to create systems for replication to ensure the legitimacy of conclusions. This solution has already been successfully implemented in the field of health care. For example, Carpenter et al. (2012) developed a framework for articulating cancer comparative effectiveness for research data needs. This framework has inspired the creation of additional frameworks that can help accelerate the pace of comparative effectiveness research and enhance the adoption of research findings by multiple stakeholders interested in improving patient outcomes. An example of a framework building off of Carpenter's model is a data schema for clinical research needs developed by Hruby et al. (2016), which incorporates the usage of the PICO (Population, Intervention, Control or Comparison, and Outcomes) framework.

Another critical lever in improving research use is the expansion of existing evidence repositories like What Works Clearinghouse (WWC) or Education Resource Information Center (ERIC) for educational interventions. To understand the quality of research that exists in the ecosystem, it is important to understand and differentiate between high-quality and low-quality studies.

The Blueprint for Inclusive Research and Development in Education (BIRD-E) project is a multi-stakeholder initiative to develop a universal framework that researchers and practitioners can use to understand, interpret, and synthesize data in education R&D. Modeled on the PICO framework used in healthcare, this is a research and development project that contributes a key piece of research infrastructure for the sector.

This project aims to develop an open-source framework that researchers and practitioners can use to understand, interpret, synthesize, and organize data within education research and development. The framework leverages existing data standards that will create pathways to share information, thus increasing efficiency and effectiveness by lowering barriers to entry for monitoring and evaluation design and increasing productivity by reducing the number of customized data variables in every study. The framework allows for a wide set of resources and studies to be meta-tagged for consideration in the research practice and for data to be captured into digestible data profiles. Ultimately, it helps to democratize research and makes it consumable for researchers and practitioners.

With support from the Bill & Melinda Gates Foundation, InnovateEDU is stewarding the creation of a shared infrastructure through regional data capacity and infrastructure investments, focusing on data privacy and security. Given the long-standing challenges of data in education, it was important to create a collaborative ecosystem of stakeholders who are deeply engaged in determining the data types most useful. The design process should attempt to create consistency through open standards and open-source approaches to be extremely attentive to the challenges faced by historically underserved and poorly served communities -- Black, Latinx, and low-income and disabled students, as well as their teachers. It was also important that this process focuses on the inherent tensions of creating common approaches in a decentralized education world thus, careful examination of scalability, adoption, and ecosystem considerations were needed from the outset.

Following a series of design workshops, InnovateEDU created an ecosystem of collaborators and stakeholders to build an R&D infrastructure that will be:

- **Pilotable:** There are ways to test the prototype infrastructure in the real-world ecosystem.

- **Impactful:** The prototype and the pilot should produce preliminary evidence of potential impact within the first few years of implementation. Evidence might include reduced costs to run validation studies, more efficient testing, increased school engagement with research, enhanced shareability of results of studies, improving the quality of user engagement in the R&D process, etc.

- **Advances Diversity, Equity, and Inclusion:** The infrastructure must operate within target communities. Furthermore, the leadership of the infrastructure and those participating in the design process should be representative of the target populations.

The proposed prototype's objectives are to:

- Articulate an open-source, data needs framework that emerges from an inclusive, multi-stakeholder engagement approach and data-driven methods for evaluation.
- Test the framework's ability to generalize beyond a specific research project through a rapid cycle evaluation in the Bill & Melinda Gates Foundation's (BMGF) R&D portfolio with four or more field pilots.
- Create natural links among other collaborations currently examining the intersection of data in education, such as T3 from the U.S. Chamber of Commerce, ADL, EdMatrix, IEEE, and other efforts underway to organize data and interoperability in the K-12 education data space.
- Engage other projects within the BMGF education portfolio, especially the efforts with Mathematica and Northwestern University, to enhance engagement and collaboration between projects that handle meta-tagging infrastructure within K-12 education. In particular, we should leverage the work with the Education Endowment Fund to create international cohesion for this work.

The anticipated long-term success outcomes are two-pronged:

1. **Efficiency and effectiveness:** Reduced cost to run research studies, increased school engagement with research, enhanced shareability of studies results, and improved user engagement in the R&D process.
2. **Replicability and scalability:** Sector-wide replicability and scalability of frameworks for different types of research for practitioners and other stakeholders.

The Researcher Working Group was formed in June 2020 to officially initiate the design, development, and implementation of the proposed R&D infrastructure. The collective contribution of this group is to spearhead the work under the leadership of InnovateEDU to

- Understand the types of data that need to be collected to inform the development of the framework.
- Review the methodology deployment under the Steering Committee's guidance.

This contribution would be aimed at leading the development of a data framework for education R&D and guiding and supervising the prototype implementation of the framework.

This work was organized by InnovateEDU, which chaired the Steering Committee and Working Groups and coordinated a convening and workshop sequence to analyze existing methodologies, identify needs and gaps within the researcher and practitioner communities, and build a data framework that could be tested in four pilots within the Bill & Melinda Gates Foundation R&D portfolio.

Constitution: Researcher Working Group

Through a clear and deliberate due diligence process, InnovateEDU selected 16 researchers representing a diverse group of stakeholders to serve on the Researcher Working Group. The group began meeting in

July 2020 and is expected to continue until July 2022. The members and organizations are listed below, in alphabetical order.

★ Brian Wright, the University of Virginia
★ Cathryn Cook, Saga Education
★ Christina Cipriano, Yale School of Medicine
★ Cindy Tipper, Carnegie Mellon University
★ Dave Paunesku, Project for Education Research That Scales (PERTS)
★ Erin Huebert, LEANLAB
★ Erin Pollard, Institute of Education Sciences
★ Gabriela Lopez, Chan Zuckerberg Initiative
★ Jessica Heppen, American Institutes for Research (AIR)
★ Neil Heffernan, Worcester Polytechnic Institute
★ Paula Arce-Trigatti, Rice University
★ Ryan Baker, the University of Pennsylvania
★ Sean Talamas, Character Lab
★ Temple Lovelace, Advanced Education Research and Development Fund (AERDF)
★ Vivian Wong, the University of Virginia
★ Virginia Knechtel, Mathematica

Members were invited based on their areas of expertise representing the research community, including research-practice partnerships, research intermediaries, and evaluation agencies providing technical assistance directly to the school district or associated partners. There were several drivers and reasons for members to join this large-scale stakeholder engagement process and spearhead the work with InnovateEDU. Many of the members are already focused on research and evidence's role in improving student outcomes, especially in marginalized populations across the country. By being a part of the working group, the members agreed to be part of an initiative where they could support and co-create a movement that has a greater impact on student outcomes.

Norms and expectations of the Researcher Working Group were established with a clear outcome for this engagement: to create a unified system that allows understanding of what can happen in schools, that allows translation of complex information to facilitate teachers and researchers to make improvements beyond what is possible - to know what works truly, where, for who and under what circumstances in education.

Purpose: Researcher Working Group

The working group participated in all the key responsibilities and engagements as identified by InnovateEDU in collaboration with the community members, especially the Steering Committee. These included:

1. **Gap analysis:** Identify the gaps and challenges of the education R&D ecosystem.

2. **Models of success:** Identify models of success from other industries to inform the design.
3. **Design and composition:** Be a part of the design and determine the composition and engagement process.
4. **Decisions and approval process:** Establish work-flows and decision processes of the Steering Committee and Working Group members.
5. **Strategic roadmap and methodology:** Establish the strategic roadmap of the project and design the comprehensive methodology of the project.
6. **Pilot the prototype:** Select and identify the pilot sites with clear learning objectives, participate in the supervision of the pilot and advise on the integration of the learnings in the framework development.
7. **Finalize the framework:** Finalize the framework draft, naming, and mission and vision documents for external release.
8. **Communication and dissemination strategy:** Finalize the framework and co-design the dissemination strategy for different stakeholders.

However, the members of the Researcher Working Group were specifically tasked with the following responsibilities and engagement to support the development of the framework:

1. **Guide the mapping and review of the current evidence base:** Guide the development of the methodology to explore semantic structures of the evidence base as well as the existing thematic gaps to influence the development of the framework. This was implemented through Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques.
2. **Develop the framework's Social and Emotional Learning (SEL) module**: Guide the methodology and review social-emotional elements to be incorporated into the framework.

InnovateEDU, in collaboration with the Researcher Working Group and the Steering Committee members, mapped these responsibilities and established formal processes and protocols for each. The specific details of each of these are explained in the section below.

Key responsibilities & engagements

1. Map and review evidence repositories

The Researcher Working Group was responsible for the mapping and review of a representative sample of the current evidence repositories to support the development of the framework. InnovateEDU and the University of Virginia School of Data Sciences established a year-long collaboration with the Researcher Working Group to design and implement the methodology using natural language processing techniques.

The design process and methodology included identifying high-impact publications by developing citation networks, similar to traditional bibliometric analyses that surface widely cited and influential research publications. This approach was conducted to produce a summary of concepts that represent key research data needs, which were then organized to reflect how researchers and practitioners think about evidence generation and synthesis. Once a collection of publications had been generated, topic modeling was applied to identify core themes to be used to validate and inform the creation of a data framework for educational research. Topic modeling is an NLP approach that uses unstructured text to generate hidden semantic structures of related documents.

The main objectives of mapping and analyzing the evidence repositories were to find the following:

- Different intervention topics and themes that are widely researched in educational interventions.
- Different variables (concepts) being used by researchers and practitioners that can be categorized into classes for further framework design.
- Identify themes and topics that are less researched and how those concepts need to be included and categorized.
- Take stock of existing biases and gaps in the current data needs and define the missing concepts.
- Design connections and cycles of the concepts and elements within them to generate a generalizable framework that can be widely used in the education space.

The main assumptions and dependencies of this exploratory research included:

- The learning domains selected are an area where there is a gap between research and practice.
- The corpora of documents collected from clearinghouses and journals represent all research papers of the selected teaching and learning domains.
- The corpora have a sufficient number of documents or studies for the evaluation to be meaningful.
- Each research paper in the corpora has the minimum, necessary information that can be used in the NLP model.

Some of the main limitations of the evaluation included the actual number of research studies collections related to the selected teaching and learning domains that. Another limitation was the analysis only gave access to the title and abstract of the research studies, and the limitations of the unsupervised machine learning methodologies may not have been the most suitable for the corpora generated.

Methodology

The methodology of the evaluation focused on using text mining and topic modeling approaches to find the underlying language used in educational research papers. Text data is naturally more unstructured and unstandardized than many data types as it is meant to be human thought on paper. Two common methods - term frequency-inverse document frequency (TF-IDF) and Latent Dirichlet Allocation, or LDA -

were used to process such data. These methods are examples of unsupervised learning, meaning that instead of looking for a definitive answer to what each paper is about, the methods focus on understanding the underlying structure of a paper and group it with other papers similar to it.

TF-IDF measures how important a word is to a document in the corpus by multiplying the raw term frequency by the inverse document frequency. A high TF-IDF score indicates that a word is most likely to be relevant to the meaning of a particular document in the corpus. To compare each word's TF-IDF scores across the corpus, these scores were converted to normalized Z-scores by subtracting the mean and dividing by the standard deviation for each word's TF-IDF score averaged across all documents in the corpus. TF-IDF facilitates the identification of a list of the most important words and word pairs within the corpus to discover what words currently make up the vocabulary of education research.

The LDA model reads through a large collection of documents and uncovers the hidden semantic structures of the corpus. The model finds the probabilities of allocating each paper to each topic, depending on the words that appear in each paper. Namely, each document in the corpus is a mixture of topics, and each topic, in this case, is an abstract concept representing the co-occurring patterns of the words. The LDA model generated a set of topics that could inform the broad themes to be included in the development of the framework.

The methodology was first implemented using a small representative sample from WWC and then expanded to a much larger dataset from ERIC and selected journals. The data compilation process included identifying all unique papers from the entire WWC (including all reviews) and the ERIC database, as well as a set of established journals and creating a dataset for evaluation.

The corpora examined by the two models were a combination of titles and abstracts of research publications from What Works Clearinghouse (WWC), Education Resources Information Center (ERIC), and selected journals. The titles and abstracts of 2000 unique papers in WWC were collected using the ERIC API (Application Programming Interface) spanning from 1981 to 2020. From the ERIC database, close to 300,000 papers were identified, and a representative sample of 100,000 titles and abstracts ranging from 2013-2021 were selected using stratified sampling based on the year of publication. The final set of titles and abstracts came from a set of established journals such as the Journal of Research on Educational Effectiveness, American Educational Research Journal, and Review of Educational Research. 700 papers from 1980-2020 were identified across all three journals, sourced from the University of Virginia Library Virgo API. The main reason for the inclusion of only titles and abstracts in the corpora was the copyright regulations that prohibit full-text mining using the model.

A small exploratory analysis was first performed by running a two-topic LDA model on the titles and abstracts of papers that met WWC standards without reservation against papers that did not, separated by math and reading subdomains. The analysis demonstrated allocation distributions to these two topics and suggested structural differences between WWC and non-WWC literature. TF-IDF scoring was implemented to study the vocabulary within WWC titles and abstracts and determine the most relevant unigrams and bigrams currently present in WWC. Finally, the LDA model was utilized to cluster WWC titles and abstracts into topics or sets of words grouped by underlying semantic similarities. Through an optimization process, it was found within the WWC, that there was an average coherence score of 0.4096 among the 39 out of 50 models that returned 11 optimal subtopics.

Findings

This section focuses on the findings of the TF-IDF, and LDA approaches for each dataset and the integration of words and topics across different datasets.

Term frequency-inverse document frequency (TF-IDF)

The What Works Clearinghouse (WWC) is a vetted database and includes rigorous papers focused on research designs, including randomized control trials, use of control groups, and statistical significance. Table 1 shows the TF-IDF scores values for bigrams in WWC and shared words with other corpora. Bigrams are two-word phrases that are most commonly cited in the representative sample. Initially, unigrams (one word) and bigrams (two words) were analyzed for this exploration. It was determined that bigrams were much more relevant for this process and used for further evaluation.

Overall, bigrams provided more information in the corpus. For example, “high school” is by far the most significant bigram, followed by “student achievement,”“national board,” “control group,” and “professional development.” Between math and reading papers, demographic information, such as “school district” and “high/middle/elementary school,” are the most common bigrams. On the other hand, important bigrams that are unique to reading include “reading recovery,” “significant difference,” and “statistically significant.” Meanwhile, important bigrams that are unique to mathematics include “problem-solving,”“student achievement,” and “grade [of] student.”

As highlighted in the table below, the bigrams unique to WWC are “control group,” “statistically significant,” and “results indicated.” This is indicative of the WWC’s intensive focus on study design, sample and statistical analysis, and the focus on the results from these analyses.

Table 1. TF-IDF scores for bigrams in WWC

Category	Bigrams
Unique to WWC	'national board', **'control group',** 'contains table', 'head start', 'learning disability', 'table figure', '**statistically significant'**, 'learning community', 'reading intervention', 'reading instruction', 'test score', **'result indicated'**
Shared words with ERIC Corpus	'public school', 'study examined', 'student learning', 'school student', 'significant difference', 'school district', 'professional development', 'community college'
Shared words	'study examined', 'student learning', 'randomly assigned', 'student

with journal corpus	achievement', 'middle school', 'professional development', 'high school', 'elementary school', 'grade student'
Shared words: all three corpus	'study examined', 'student learning', 'school student', 'middle school', 'professional development', 'high school'

Overall, TF-IDF analysis confirms the importance of three types of words in WWC corpus: 1) words that are intrinsic to education papers (“student,” “school,” “teacher”), 2) words that refer to the methological approach of study being presented (“randomly assigned,” “control group,” “statistically significant”), and 3) words that suggest the sample and subgroup being studied (“grade,” “gender,” “state,” “school district,” “high school,” “learning disability,” etc.). Based on these findings, it was recommended that the framework needed to incorporate elements related to the students, the schools, and the teachers involved in educational studies. It should also incorporate elements related to the type of educational study being presented and elements related to the group of students being examined.

The ERIC corpus, being the most general, presented a large range and types of bigrams. TF-IDF scores showed the most relevant words that appear in the ERIC corpus. Table 2 below shows the unique and shared words with each corpus.

Table 2. TF-IDF scores for bigrams in ERIC

Unique to ERIC	'preservice teacher', 'teaching learning', 'purpose study', 'data collected', 'learning environment', **'autism spectrum'**, 'english language', 'young people', 'present study', **'spectrum disorder',** 'primary school', '**special education',** 'student teacher'
Shared words with WWC Corpus	'public school', 'study examined', 'student learning', 'school student', 'significant difference', 'school district', 'professional development', 'community college'
Shared words with journal corpus	'teacher education', 'higher education', 'study examined', 'student learning', 'united state', 'school student', 'case study', 'middle school', 'professional development'
Shared words: all three corpus	'study examined', 'student learning', 'school student', 'middle school', 'professional development', 'high school'

It is interesting to note that analysis of the ERIC corpus revealed a much larger set of words about disabilities compared to the WWC corpus. In ERIC and the corpus of journals, “case study” was a dominant topic, whereas, in ERIC and the WWC, the term “public school” seemed to be a dominant topic across both datasets.

Three major journals - namely Journal of Research on Educational Effectiveness, American Educational Research Journal, and Review of Educational Research - were chosen to construct the model for analysis of titles and abstracts for each of the papers.

Table 3. Journals used in analysis

Journal	Number of titles and abstracts used
Journal of Research on Educational Effectiveness	84
American Educational Research Journal	369
Review of Educational Research	259

Reviewing the TF-IDF scores, many words emerged in the journal corpus that were not common with ERIC and WWC, especially related to the effect of gender, race, or income status on a student's learning. Additionally, according to TF-IDF, elementary, middle, and high school all showed up as high-ranking words. Table 4 below shows the unique and shared words within each corpus.

Table 4. TF-IDF scores for bigrams in journals

Unique to educational journals	'meta analysis', 'effect size', 'academic achievement', 'self concept', 'problem solving', **'african american',** 'social emotional', 'future research', **'sex difference',** 'treatment effect', 'review research', 'small group', **'low income',** 'teacher efficacy', 'review literature'
Shared words with WWC Corpus	'study examined', 'student learning', 'randomly assigned', 'student achievement', 'middle school', 'professional development', 'high school', 'elementary school', 'grade student'
Shared words with ERIC corpus	'teacher education', 'higher education', 'study examined', 'student learning', 'united state', 'school student', 'case study', 'middle school', 'professional development
Shared words: all three corpus	'study examined', 'student learning', 'school student', 'middle school', 'professional development', 'high school'

Between all three corpora, there is an emphasis on middle and high school and on the grade of the student (characterized by "school student") and on learning and professional development.

Latent Dirichlet Allocation (LDA)

The LDA model was used to find the probabilities of allocating each paper to a topic, or a set of words, depending on the words that appear in each paper. Given some set number of topics "n," the LDA model approximates the probability distribution of words in each topic and the probability distribution of topics in each paper. Therefore, the emphasis is on finding the best number of topics “n” should be, and then passing that parameter to the LDA model. An optimization procedure found that eleven topics generally returned the highest model coherence on the WWC corpus. An eleven-topic LDA model was run on the corpora to find a percentage of documents allocated to each topic and the bigrams with a high proportion of allocation in each topic.

Each topic had an informed interpretation based on the bigrams that emerged. It is important to note that each title and abstract of the paper was assigned to one topic (the topic with the highest probability in the document-topic probability distribution), thus, all the percentages shown in Table 5 sum up to 100%. Based on the percentage of documents that are allocated to each topic, the analysis found which topics were most well-represented in the corpus.

Table 5 below illustrates the percent of documents, frequent bigrams for each topic, and the topic interpretations for the WWC and ERIC corpora. Based on the informed interpretation of the topics, the most prevalent topics in the WWC corpus were related to study characteristics, higher education, school characteristics, and reading development. It was observed that the distribution of topics in the WWC corpus was not uniform, as there were a small number of topics with a large number of papers allocated to them, meanwhile, all other topics had a small number of papers allocated to them.

The same eleven-topic model was deployed on the ERIC corpus to compare to the eleven-topic LDA model fit on the WWC corpus. Based on the percentage of papers that are allocated to each topic, the model concluded the topics were most well-represented in the ERIC corpus. Based on the interpretation of these topics, it could be concluded that the most prevalent topics in the ERIC corpus were related to the learning environment, education research, school characteristics, and education system.

Table 5. Percent of documents and interpretation of topics found in corpora

WWC Topics

	Percent of Documents	Our Interpretation
Topic 1	**19.10%**	**Study Characteristics**
Topic 2	0.10%	Youth Intervention
Topic 3	2.50%	Child Autism Intervention
Topic 4	0.10%	Child Literacy Development
Topic 5	16.26%	Higher Education
Topic 6	**22.75%**	**School Characteristics**
Topic 7	**4.32%**	**Student Achievement**
Topic 8	**3.96%**	**Language Development**
Topic 9	5.35%	Learning Disabilities
Topic 10	18.27%	Reading Development
Topic 11	7.31%	Early Childhood

ERIC Topics

	Percent of Documents	Our Interpretation
Topic 1	2.56%	Child Development
Topic 2	**8.69%**	**Student Achievement**
Topic 3	**7.60%**	**Language Development**
Topic 4	9.19%	Research Discipline
Topic 5	6.45%	Student Health Factors
Topic 6	17.88%	Learning Environment
Topic 7	11.15%	Education Research
Topic 8	**10.76%**	**School Characteristics**
Topic 9	8.29%	Educator Characteristics
Topic 10	10.72%	Education System
Topic 11	**6.86%**	**Study Characteristics**

It is also important to note that this distribution of topics in the ERIC corpus is much more uniform than that in the WWC corpus. For instance, the smallest percentage of documents in one topic was just 0.10% for the WWC corpus, but 2.56% for the ERIC corpus; and the highest percentage of documents in one topic was 22.75% for WWC, but just 17.88% for ERIC. This illustrates that the larger corpus encompasses a broader range of topics.

Based on the interpretation of the topics, it was found that four topics were present in both the WWC corpus and the ERIC corpus: **study characteristics, school characteristics, student achievement, and language development**. Topics such as higher education, early childhood, and reading development emerged only in WWC. Meanwhile, topics such as learning environment, research discipline, and education research emerged only in ERIC.

When the LDA model was deployed on the selected journals, twelve topics generally returned the highest model coherence in the journal corpus. Through the twelve-topic LDA model, a percentage of documents allocated to each topic and the bigrams with high proportions of allocation in each topic were analyzed. Similar to other corpora, interpretation of what each topic represented based on bigrams from each topic was established.

The most prevalent topics in the journals were related to the learning environment, study outcomes, education inclusivity, and study methodology. The distribution of topics in this corpus is even more uniform than in the ERIC corpus. Since this corpus is much smaller than the other two corpora, each topic's bigrams' frequencies were on a much smaller scale than the frequencies in the WWC and ERIC corpora. Table 6 below highlights the percentage of documents and the topic interpretation for journals in the corpus.

Table 6. Percent of documents and interpretation of topics found in journals

Percent of Documents		Frequent Bigrams in Topic	Theme of Topic
Topic 1	4.10%	"ability student", "ability level", "success failure", "low ability"	Student Ability
Topic 2	**16.72%**	"student learn", "public pedagogy", "learning environment"	**Learning Environment**
Topic 3	**10.64%**	"treatment effect", "effect size", "regression analysis", "case study"	**Study Outcomes**
Topic 4	**11.55%**	"gender difference", "inclusive education", "equal educational"	**Education Inclusivity**

Topic 5	3.95%	“science education, “immigrant education”, “computer simulation”	Miscellaneous
Topic 6	5.78%	“parent involvement”, “victim bully”, “bully cycle”, “child family”	Student Personal Issues
Topic 7	5.17%	“academic achievement”, “student achievement”, “test score”	Student Achievement
Topic 8	5.62%	“child relationship”, “intervention child”, “teacher child”	Child Intervention
Topic 9	8.36%	“academic deficit”, “classroom management”, “staffing problem”	School Performance
Topic 10	**15.96%**	“effect size”, “treatment group”, “social study”, “control condition”	**Study Methodology**
Topic 11	6.84%	“teacher efficiency”, “novice teacher”, “expert teacher”	Teacher Performance
Topic 12	5.32%	“school violence”, “teacher color”, “young black”, “school safety”	Urban School Traits

In conclusion, it was found that the smaller WWC corpus encompassed a smaller range of topics, with only a few topics capturing most of the documents in this corpus. The topics of the WWC corpus were concentrated mostly on study characteristics, and school characteristics, as most papers reviewed by WWC are about educational interventions and randomized control trials.

The larger ERIC corpus encompassed a broader range of topics, with each topic capturing about an even number of documents in this corpus. The topics of the ERIC corpus were represented evenly from learning environment to language development, as ERIC houses a large and more general pool of educational research papers. Lastly, a small corpus of journals encompassed a range of new topics compared to the two other corpora. The corpus of journals introduced new topics like education inclusivity and urban school traits. This might suggest that each journal focuses on specific, diverse themes compared to other educational research papers.

The LDA preliminary analysis shows that for mathematics studies, there are strong structural differences between math papers approved by WWC and math papers not included by WWC. Meanwhile, differences are not that strong in reading studies - between reading papers approved by WWC and reading papers not reviewed by WWC. These results imply that the language used in high-quality papers on mathematics studies differs from that used in papers on standard-quality mathematics studies. Alternatively, the lexicon used in high-quality reading papers is not that different from that used in standard-quality reading papers. This also shows that there could be some intrinsic differences between math and reading papers.

NLP techniques like the ones employed are two-fold: one part requires quantitative exploration, while the other requires qualitative interpretation. The scope of this exploration and review was limited to the quantitative side, while the work of the groups of researchers, practitioners, and educators is limited to the qualitative side. Multiple interpretations of topics can emerge based on the bigrams studied and will be a reference point for future review strategies.

Recommendations

Through the TF-IDF and LDA model, topics can be identified that are most well-represented in the corpus, based on the number of documents allocated to each topic. The most prevalent topics in the corpora are school characteristics, study characteristics, reading development, and higher education in descending order of the number of documents in each topic. As a result, it is recommended that these four themes should be represented in the data framework.

More work may be needed to understand the evolution of education research, and this can be accomplished by tracking how the needs of the data framework change after it has been deployed. In collaboration with the data science team, the Researcher working group had two-fold recommendations for future investigation and exploration.

1. A large and inclusive feedback mechanism from different sets of stakeholders is required for the interpretation of generated topics. Currently, the interpretation is limited to the working groups and the data science team and could inhibit a more expanded interpretation of the emerging topics.

2. The three corpora should be broken down further by publication year to find the trends in how the topics of each corpus change over time. Due to the limitation of the scope of exploration and available computation power of the data science team, the evaluation analyzed only one-third of the full Eric corpus. Further investigation is needed to run and fine-tune the LDA model on all three hundred thousand ERIC papers. Additionally, other machine learning algorithms should be used to explore each corpus further and find integrations between all three datasets to investigate emerging trends further.

This universal framework should help resolve the issue of inconsistent research terminology by documenting the definition of each element in a data dictionary. By standardizing research terminology with the validation of multiple working groups, the resulting framework has the potential to move the field towards increased data interoperability and inclusivity, thus alleviating the issues of inconsistent data collection and non-inclusive data distribution.

2. Social and Emotional Learning (SEL): Module development

Social and Emotional Learning (SEL) is a critical component in understanding the factors that impact student outcomes in the short and long term. Social and Emotional Learning (SEL) refers to an interrelated set of cognitive, affective, and behavioral skills and strategies that underscore learning and development. SEL interventions support the development of inter and intrapersonal skills and promote health and outcomes for all students. Implementing SEL practices leads to better emotional intelligence, behavior regulations, and other skills necessary for the overall development of a student.

The field of SEL research is relatively young and therefore does not have a well-defined large body of research or common language. Several attempts have been made to develop a common and comprehensive language for SEL. The BIRD-E team and the Researcher Working Group were entrusted to conduct a landscape analysis of the SEL domain and identify the most relevant, evidence-driven SEL frameworks to create a list of common elements that would support the development of the SEL module in the framework. While there is no dearth of SEL frameworks in the ecosystem, it was important to identify frameworks that are representative of a wide range of disciplines and are widely adopted. It was also critical that these frameworks included descriptive skills, traits, competencies, strengths, mindsets, and/or attributes that are defined and can be coded, especially focused on K-12 education.

There is a need for greater precision in the field of SEL and other related non-academic areas, a point emphasized by practitioners, researchers, and funders. However, it is difficult to do this work and assess which skills to focus on and how to build them out and measure them. Many organizations in the last few years have initiated a comprehensive compilation of key competencies in SEL and have experienced limited success. This is further exacerbated by the unorganized information and options available in this field. Most of the non-academic domains lack clarity about what it means and are crowded with a large number of systems or frameworks that often use different or even conflicting terminology to talk about a similar set of skills. We need a system to address this issue that compares different frameworks, connecting them back to scientific evidence and making informed decisions about school standards and strategies.

To start the compilation of the SEL module and elements within it, the Researcher Working Group identified the Explore SEL initiative established by Harvard University, a foundational platform that reviews the current SEL frameworks in the K-12 ecosystem. Explore SEL is a product of the Taxonomy Project, which is an ongoing project designed to create a scientifically-grounded system for organizing, describing, and connecting frameworks and skills across the non-academic domain. The Taxonomy project uses a coding system grounded in a foundational body of knowledge, including human development and psychology, cognitive and behavioral neuroscience, and the intervention and prevention sciences. The system maps frameworks and terms onto one another to illustrate whether and how non-academic constructs and terms are related to one another. The system is designed to preserve the integrity of each framework without obscuring nuances in meaning or links to evidence. The resulting database of coding frameworks and terms – or the taxonomy – serves as the foundation for this project. SEL Explore contained 40 frameworks. Out of the 40 frameworks, 22 frameworks were selected that had a particular focus on K-12 grades for further analysis. Table 7 depicts the frameworks that were included in the landscape analysis and Table 8 includes frameworks that were excluded from the landscape analysis.

Table 7. Explore SEL frameworks included

Framework Included	Age Range
21st Century Learning	Not Specified
ACT Holistic Framework	K-Career
Building Blocks for Learning	K-12
CASEL	PreK - 12
Character Lab	K-12
Developmental Assets	Age 12-18
EU NESET Framework for Social and Emotional Education	Early Childhood
Head Start	0-5
Habits of Mind	Not Specified
Hilton & Pellegrino Clusters of 21st Century Competencies	K-12
K-12 SEL Standards (Anchorage)	K-12
K-3 SEL Standards (Connecticut)	K-3
KIPP	K-12
LEGO’s Skills for Holistic Development	Early Childhood
MELQO Model Framework	Early childhood
MESH	K-12
OECD	Primary and Secondary
Social, Emotional, and Ethical (SEE) Learning Framework	K-12
The Five C's Model of Positive Youth Development	Adolescent
The Practice Model	0-29
Unicef India Comprehensive Life Skills Framework	Early Childhood
Vision of the Haitian Child in Society: Social Emotional Framework	pre-primary to post-secondary

Table 8. Explore SEL frameworks excluded

Framework Excluded	Age Range
Big Five Personality Traits	All
Clover Model	Infant-adult
EDC Work Ready Now! Framework	15-30
Emotional Intelligence	Infant-Adult
Employability Skills	Not Specified
Head Start	0-5
IB Learner Profile	Early Childhood
IRC Social and Emotional Learning Competencies	School Age Child
Kenya BECF Core Competencies for Basic Education	Age 4-18
Kenya TVET Values and Life Skills (VaLl) Framework	15-25
Pratham Life Skills Framework	Adults
Preparing Youth to Thrive	Age 5-18
Room to Read Life Skills Education Learning Outcomes	Adolescent
Sesame Workshop Global Framework For Learning	Early Childhood
Singapore Framework for 21CC and Student Outcomes	Primary - pre-University
UNICEF MENA Life Skills and Citizenship Education- Conceptual and Programmatic Framework	Children
USAID Youth Power Action Key Soft Skills for cross-Sectoral Youth Outcomes	Age 12-29
WHO Skills for Health	preschool-early adulthood
Young Adult Success	Age 3-22

As a next step, all the elements were laid out from each of the selected frameworks to identify any overlap across the frameworks. Image 1 below shows SEL elements that are either highlighted in green or yellow. The elements that are highlighted in green showed up almost across all frameworks and the elements highlighted in yellow had a slight variation in the semantics but were close to the elements highlighted in green.

	Present across multiple SEL Frameworks
	Closest to the CORE SEL elements but not identical

EU NESET Framework for Social and Emotional Ed	MELQO MODEL	MESH	OECD	Singapore framework	LEGO's Skills for Holistic Development	SEE Learning Framework	Five C's Model of Postive Youth Development	PRACTICE MODEL	UNICEF India Comprehensive Life Skills	Vision of the Haitian Child in Society: Social Emotional	Young Adult Success	Habits of Mind
self-awareness	working memory	self-management	achievement orientation	respect	emotional skills	attending to our body and sensations	competence	problem solving	self-awareness	self concept	agency	Persisting
self-management	inhibitory control	self-efficacy	responsibility	harmony	cognitive skills	attending to emotions and feelings	confidence	resilience	communication	self-awareness	integrated identity	Managing impulsivity
social awareness	self-regulation	growth mindset	self-control	resilience	physical skills	map of the mind	connection	achievement motivation	resilience	emotional understanding	competence	Listening to others with understanding and empathy
social management	social cognition	social awareness	persistence	responsibility	social skills	understanding emotions in context	character	control	decision making	collaboration/konbit	self-regulation	Thinking flexibly
	social competence		stress resistance	integrity	creative skills	self-acceptance	caring/compassion	teamwork	negotiation	relationship skills	knowledge and skills	Thinking about our thinking (metacognition)
	emotional well-being		optimism	care		balancing the body	contribution	initiative	empathy	acceptance	mindsets	Striving for accuracy and precision
			emotional control	self-awareness		cognitive and impulse control		confidence	participation	critical thought	values	Questioning and positing problems
			empathy	self-management		navigating emotions		ethics	creativity	self-management		Applying past knowledge to new situations
			trust	responsible decision making		attending to our social reality			critical thinking	resilience/persistence		Thinking and communicating with clarity and precision
			cooperation	relationship management		attending to our shared reality with others			problem solving			Gathering data through all senses
			curiosity	social awareness		appreciating diversity and difference						Creating, imagining, and innovating
			tolerance	active community life		understanding others' feelings and emotions in context						Responding with wonderment and awe
			creativity	global awareness		appreciating and cultivating kindness and compassion						Taking responsible risks
			sociability	national and cultural identity		appreciating and cultivating other ethical dispositions						Finding humor
			assertiveness	socio-cultural sensitivity and awareness		empathic listening						Thinking interdependently
			energy			skillful communication						Learning continuously
			self-efficacy			helping others						
			critical thinking/independence			conflict transformation						
			self-reflection/meta-cognition			understanding interdependent systems						
						individuals within a systems context						
						appreciating the fundamental equality of all						
						appreciating how systems affect well-being						
						exploring one's potential for effecting positive change in community and world						
						engaging in communal and global solutions						

	Present across multiple SEL Frameworks
	Closest to the CORE SEL elements but not identical

Image 1. Elements across frameworks

After the landscape analysis was conducted, 23 elements emerged as the most prevalent elements across frameworks. Table 9 below shows the 23 elements that were drafted in the Advanced Blueprint.

Table 9. List of social-emotional learning elements in the Advanced Blueprint

Social-Emotional Learning Elements	
Grit	Cooperation
Self-awareness	Optimism
Agency	Sociability
Assertiveness	Problem Solving
Growth mindset	Creativity
Self-control	Social awareness
Competence	Curiosity
Self-management	Relationship skills
Resilience	Zest
Confidence	Responsibility
Integrity	Empathy
Self-regulation	

The 23 elements selected was presented for a thorough review by the members of the Working Group and Steering Committee. There was further discussion on the applicability and relevance of these chosen elements and how they would contribute to the development of consistent measurements of SEL competencies. Based on feedback from the experts on SEL and the Steering Committee and Working Groups, 3 SEL competencies were further removed from the compiled competencies, namely: sociability, social awareness, and zest. It was agreed upon that further research and investigation are needed for these elements to be included in the existing Blueprint and can be included in an SEL extension framework to be developed in the future. Table 10 below presents the final list of 20 SEL elements that were included in the released version of the advanced Blueprint.

Table 10. FINAL List of social-emotional learning elements in the Advanced Blueprint

Social-Emotional Learning Elements	
Grit	Integrity
Self-awareness	Self-regulation
Agency	Cooperation
Assertiveness	Optimism
Growth mindset	Problem-Solving
Self-control	Creativity
Competence	Curiosity
Self-management	Relationship skills
Resilience	Responsibility
Confidence	Empathy

Conclusion

It is increasingly evident that education research is critically needed to improve our education systems and to open up opportunities for enhanced teaching and learning. However, our R&D infrastructure in education is inadequate. Not only do we underinvest in the necessary infrastructure, but education research as a field also is not yet structured optimally to impact educators' decisions. Educators and system leaders do not have access to the bodies of research and findings that could support high-quality teaching and learning, and scholarship is too often disconnected from practice. Equity is too often an afterthought or is measured in simplistic and reductive ways, which prevents effective implementation.

If our common goal is for research to inform more equitable teaching and learning in education systems, then we must both deepen investment in and reimagine R&D infrastructure in education. There is a need for new systems that help us answer questions about what works for whom, how and under what conditions. We need to create systems that facilitate data generation and the sharing of research findings and improve the translation of research syntheses. This will allow for better accessibility and discoverability of research by practitioners, researchers, and policymakers for decision-making.

The BIRD-E project has employed a data-driven approach to developing a conceptual framework - the Blueprint - for defining education research data needs. The resulting Blueprint is an open-source framework that aims to modernize education research through a common, research-based data language to bridge the divide between research and practice in the K-12 data ecosystem. The Blueprint aims to provide a structured, universal, and consistent approach to the design, collection, and reporting of research to answer the most pressing questions facing educators. It serves as a map to modernize current K-12 research so that impactful research cannot only be conducted -- but can meaningfully inform and strengthen practice.

The Blueprint can become a component of the foundational R&D infrastructure needed to create a common comprehensive research framework and create a shared vocabulary to articulate research data needs. Its goal is to facilitate the engagement of all types of stakeholders in inclusive, accessible, and robust generation and use of research. The Blueprint focuses on supporting a learning system within the research and development infrastructure that evolves and considers usability in the practitioner community. It can facilitate communication between researchers and practitioners to ensure improved evidence generation as well as serve as a metadata framework to index and organize research evidence bases for better discoverability in the space of evidence synthesis. Further real-world adoptions and close studies are needed and warranted to test these potentials and serve the education sector.

References

- Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., & Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. the Journal of Machine Learning Research, 3, 993-1022.
- Bill & Melinda Gates Foundation. (n.d.). AI in Education: Imagine the Possibilities [Brochure]
- Campbell, J. and Copa, N. (2019). Standardizing School Innovation Data: Report and Recommendations. Retrieved October 04, 2020, from https://www.aemcorp.com/
- Carpenter, W. R., Meyer, A., Abernethy, A. P., Stürmer, T., & Kosorok, M. R. (2012). A framework for understanding cancer comparative effectiveness research data needs. Journal of Clinical Epidemiology, 65(11), 1150-1158. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3461183/
- Cipriano, C., Naples, L., Wood, M., Sehgal, K., Kilgallon, E., & Zieher, A. (2021, December 21). What is Social and Emotional Learning (SEL)? A supplement to the manuscript, "The State of Evidence for Social and Emotional Learning: A Contemporary Meta-Analysis of Universal School-Based SEL Interventions," Open Science Framework, https://osf.io/6pqx2/
- Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: A systematic review. Journal of the Medical Library Association, 106(4). https://pubmed.ncbi.nlm.nih.gov/30271283/
- Hamilton, L., & Hunter, G. (2020, April 15). Educators Turn to Colleagues for Help with Struggling Students. Retrieved October 04, 2020, from https://www.rand.org/pubs/research_reports/RR2575z9-1.html
- Hruby, G. W., Hoxha, J., Ravichandran, P. C., Mendonça, E. A., Hanauer, D. A., & Weng, C. (2016). A data-driven concept schema for defining clinical research data needs. International journal of medical informatics, 91, 1-9. https://pubmed.ncbi.nlm.nih.gov/27185504/
- Huang, X., Lin, J., & Demner-Fushman, D. (2006). Evaluation of PICO as a knowledge representation for clinical questions. AMIA Annual Symposium proceedings. AMIA Symposium, 2006, 359–363. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839740/
- Menard, A. (2020). What's Wrong with Social Science and How to Fix It: Reflections After Reading 2578 Papers. https://fantasticanachronism.com/2020/09/11/whats-wrong-with-social-science-and-how-to-fix-it/
- New Schools Venture Fund (2019). Making meaning of the NewSchools-Gallup survey of educator and student perceptions of ed tech. Fordham Institute for Advancing Educational Excellent. https://www.newschools.org/wp-content/uploads/2020/03/NewSchools-Gallup-Report.pdf
- Nilsen, P. (2015). Making sense of implementation theories, models, and frameworks. Implementation Science, 10(1). https://implementationscience.biomedcentral.com/articles/10.1186/s13012-015-0242-0
- Penuel, W.R., Briggs, D.C., Davidson, K.L, Herlihy, C., Sherer, D., Hill, H.C., Farrell, C.C., & Allen, A-R. (2016). Findings from a national survey of research use among school and district leaders (Technical Report No. 1). Boulder, CO: National Center for Research in Policy and Practice. https://files.eric.ed.gov/fulltext/ED599966.pdf
- Roder, M., Both, A. & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. Proceedings of the Eighth International Conference on Web Search and Data Mining, Shanghai, February 2-6. https://dl.acm.org/doi/10.1145/2684822.2685324

- Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC medical informatics and decision making, 7, 16. https://doi.org/10.1186/1472-6947-7-16
- Turner, M.A. (July 17, 2013). Evidence-Based Policymaking Requires a Portfolio of Tools. Testimony Submitted for the Record to the Subcommittee on Human Resources Committee on Ways and Means United States House of Representatives. https://www.urban.org/sites/default/files/publication/24451/904598-evidence-based-policymaking-requires-a-portfolio-of-tools_0.pdf
- WWC Process Brief: The Study Review Process. Retrieved April 3, 2021, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_review_052417.pdf