

## Mathematica and the Blueprint:

### Using the Blueprint to understand the effectiveness of automated essay scoring interventions

---

#### OVERVIEW

Mathematica illuminates the path to progress for public- and private-sector changemakers. Mathematica operates at the intersection of data, methods, policy, and practice, with the mission to improve public well-being. It supports their partners to make evidence-based decisions and to use evidence to guide practice in a variety of sectors, including education, healthcare, child welfare, and family support. They work alongside their partners to tackle social challenges in real time.

Mathematica grantees design and conduct research to inform program improvements, and to measure program costs and the extent to which the programs improve outcomes for students. Mathematica has partnered with grantees in a few different programmatic areas, including automated essay scoring tools designed to create more opportunities for students to practice writing.

The BIRD-E Blueprint is a free, open-source framework that aims to modernize education research through a common, research-based 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 design, collection, and reporting of research to answer the most pressing question of what works, for whom, and under what conditions. It contains a list of critical data elements that summarizes and represents key education data needs in the K-12 spectrum including early childhood.

---

#### APPROACH

The evaluation process entailed two distinct approaches:

1. Assess the applicability and coverage of the Blueprint by cross-mapping the modules and submodules of the Blueprint against the established evaluation framework of different pilot partners.
2. Instrument a research question with the partners using the Population, Intervention, Comparison and Outcome framework and use the Blueprint to identify and classify the elements needed for evaluation in the specific context of the research question.

The success of the pilots was measured by 3 major criteria:

1. **Coverage:** The coverage is defined as the total number of elements mapped as a percentage of the total number of elements identified to represent the education data needs in the evaluation frameworks of the pilot partners.
2. **Understandability:** The understandability is defined as the assessment of clarity in definitions and understanding of the elements by other stakeholders. The precision, organization and ambiguity was assessed using qualitative and quantitative data.
3. **Generalizability:** The generalizability is defined as the extent to which the elements, modules and sub-modules are aligned to real-world application and experiences. The generalizability was calculated as a median of the coverage for all pilots and aggregated the scores of all pilots to determine the overall generalizability of the Blueprint.

Mathematica partners with grantees in a few different programmatic areas, but the pilot focused on the evaluation of automated essay scoring tools designed to create more opportunities for students to practice writing. The pilot tested the Blueprint against a set of research questions from the current cohort of grantees using the essay scoring tools.

---

## APPROACH CONT.

Mathematica and BIRD-E examined the questions across the grantees and consolidated them to four common research questions. An exhaustive list of elements were identified from the current research plans of the grantees in a collaborative engagement. 57 elements were found across the evaluation plans to conduct the mapping process.

This highlighted the ability of the Blueprint to act as a translational layer between researchers and practitioners and substantiated the researchers' use case. By making research inputs and outcomes more comparable, Mathematica would be better positioned to understand their grantees' impact and what works for students.

---

## RESULTS & RECOMMENDATIONS

- All 57 elements were mapped to the Blueprint, though not all elements had a unique match.
  - This was primarily due to identified elements from the evaluation plans having broad definitions which led to the aggregation of these elements into few elements in the Blueprint.
- Elements that were more difficult to match were those specific to writing, like student argumentative writing proficiency and essay revisions.
  - Despite the bespoke and narrow nature of these elements, it was matched with the elements in the Blueprint.
- Overall coverage for this pilot was 88%, with 50 Blueprint elements mapped to the 57 compiled Mathematica elements.
- Qualitative methods were used to calculate understandability, with an understandability score of 62.16%.