Evaluation of the Alabama Numeracy Act
Year 1 Annual Report

Alabama Numeracy Act Evaluation: Year 1 Annual Report

Table of Contents

List of Tables

Evaluation of the Alabama Numeracy Act: Year 1 Annual Report

Background

Recent reports from the National Assessment of Educational Progress (NAEP), the Nation's Report Card, paint a bleak picture of students' mathematics performance in the United States. Nationally, students' math scores in Grades 4 and 8 declined significantly between 2019 and 2022. Much of this decline is attributed to the COVID-19 pandemic, but math scores on NAEP were essentially static for several years before the pandemic. While Alabama students in Grade 4 did not show a significant decline from 2019–2022, Grade 8 students declined substantially. The percentage of Alabama students scoring at or above Proficient on NAEP math in Grades 4 and 8 is significantly lower than the national average.

Poor math performance in school leads to poor math ability in adulthood. The Program for International Assessment of Adult Competencies (PIAAC) showed that working-age adults in the United States performed better than the international average in literacy but well below the international average in numeracy. Roughly 30% of Americans performed in the lowest two of the five levels on the PIAAC for numeracy, meaning they were only comfortable with basic arithmetic, counting, sorting, and similar tasks. Literacy and numeracy are foundational to acquiring and maintaining complex skills and abilities. American students will compete in a workforce that is becoming more global every day. If their preparation in mathematics does not improve dramatically, they will enter that workforce at a sizeable disadvantage.

The Alabama Numeracy Act (ANA) addresses the urgent need to improve the math proficiency of Grade K–5 students and ensure those students are proficient in math at or above grade level at the end of Grade 5. The ANA represents a comprehensive system of improvements designed to support educators in all aspects of instructing students in math. At the outset of the implementation of the ANA, schools among the lowest performing 5% were identified as "full-support" schools, and those performing in the bottom 6% to 25% were identified as "limited-support" schools. While both sets of schools receive support under the ANA, full-support schools receive more intensive support.

ANA will create a pool of mathematics coaches to support schools in improving numeracy in Grades K-5. To be successful, it is imperative that effective coaches are identified, those coaches are provided with the tools to help teachers improve their math instruction, curricular supports are provided that enable effective instructional practices, teachers implement those practices with fidelity, student performance outcomes are monitored, and aspects of the system are adjusted based on clear, actionable evaluation data that reflects every step of this process. It is an ambitious but vital system that Alabama students depend on to succeed.

The Human Resources Research Organization (HumRRO), along with its partner Mathematica, was awarded a contract in fall 2023 to conduct an evaluation of the ANA. This 5-year contract focuses on key ANA aspects implemented by full- and limited-support schools. The overall ANA evaluation, which includes process and outcome components and supplemental studies,[1] addresses the following 17[2] research questions:

---

[1] Appendix A presents our approach to conducting the process and outcome evaluation components and eight supplemental studies.
[2] Although the January 2024 report listed the 17 research questions, it incorrectly indicated 12; this number has been corrected.

A. Were all processes and activities required by the ANA implemented by stakeholders? What factors facilitated or impeded the implementation? How were barriers overcome?

B. To what extent did the implementation of the ANA improve mathematics proficiency of students in Grades K–5? To what extent was the improvement consistent for all subgroups? What are the characteristics of full- and limited-support schools that make the greatest progress improving proficiency scores?

C. To what extent do full- and limited-support schools that are assigned a math coach yield better performance than such schools that do not have a coach?

D. To what extent is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school?

E. To what extent do performance evaluations of math coaches by principals and regional coordinators in full- and limited-support schools relate to differences in math achievement?

F. To what extent is the Alabama Framework for MTSS (Multi-tiered Systems of Support) being implemented in Grades K–5?

G. To what extent do ratings of implementation of MTSS (reported in F above) within schools relate to the distribution of students within tiered placements?

H. What are the status and gains in math knowledge and skills of K–5 teachers (e.g., as perceived by the math coach and/or principal)?

I. To what extent do principals' and regional coordinators' ratings of coaches explain variance in principal and coach evaluations of teachers?

J. To what extent do ratings of the math knowledge and skills of K–5 teachers within full- and limited-support schools (e.g., as made by coaches or principals) account for differences in student performance on formative and summative assessments in math?

K. To what extent do required screening and diagnostic assessments identify students who are subsequently identified as needing tiered services and/or receive diagnosis relating to math (e.g., specific learning disability or dyscalculia)?

L. What positive and negative outcomes emerged within schools, LEAs, ALSDE, and other stakeholder groups that were not anticipated as a result of the implementation of any component of the ANA?

M. What were the impacts of the School Turnaround Academy?

N. What were the impacts of the Instructional Leadership Framework?

O. To what extent were the relationships between process and outcomes achieved as expected based on logic models? What external factors impacted the anticipated accomplishments and relationships? Is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school?

P. To what extent are stakeholders aware of and satisfied with the implementation of the ANA?

Q. What are the overall costs and actual or anticipated financial benefits of the ANA?

Each research question will be addressed directly and indirectly across multiple evaluation components, allowing HumRRO to triangulate findings to inform a rich evaluation and provide substantive recommendations. Table 1 summarizes how—through the process and evaluation components and supplemental studies—each of the 12 research questions across all years of the evaluation study will be addressed. Appendix B provides descriptive information about the research questions and general timing for conducting the various evaluation activities throughout the 5-year contract.

Table 1. Research Question by Evaluation Component and Study

Evaluation Component	Research Question																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Process Evaluation	**✓**	✓	✓	**✓**	✓	✓		✓		✓		✓	✓	✓	**✓**	✓	
Outcome Evaluation	✓	**✓**			✓		✓	✓	✓	✓			**✓**		**✓**		
Math Coach Comparison Study	✓		**✓**		✓												
Math Coaches and Student Achievement	✓	✓	✓	✓	**✓**			✓	**✓**	✓				✓			
MTSS	✓			✓		**✓**	**✓**				✓			✓			
Teacher Math Pedagogy	✓				✓		**✓**	✓	**✓**					**✓**			
Screening Assessments	✓					✓					**✓**						
Unintended Consequences	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	**✓**	✓	✓	✓	✓	✓
Stakeholders	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**✓**	✓
Cost Effectiveness	✓						✓					✓	✓				**✓**

Note. Green check marks indicate a direct focus; grey check marks indicate an indirect focus.

Year 1 Evaluation Activities

The first year of the ANA evaluation is devoted to building the foundation for the overall evaluation. HumRRO will begin monitoring and documenting key aspects of the ANA implementation during Year 1, with an increased focus on the quality and effectiveness of ANA implementation in subsequent years. Our approach involves working with key stakeholders to develop an efficient system that supports and sustains long-term data collection, information tracking, and management processes to serve the dual purpose of monitoring the fidelity of ANA implementation and the quality and effectiveness of the services provided.

The first part of Year 1 focuses on outlining all the data requirements, including the criteria, indicators, and metrics, along with their current availability to address the research questions. Because a vast amount of data is required to support the ANA evaluation, Year 1 will involve establishing a data sharing agreement (DSA), determining procedures for the controlled and organized transfer of data, and developing protocols to manage and maintain data across all study years. Once the foundation for the overall evaluation is established in Year 1, the focus of

subsequent years shifts from a development cycle to system implementation, monitoring and evaluation, and feedback.

Completed Year 1 Evaluation Activities

Some of the first activities we completed involved introductory and kickoff meetings. Since contract award in fall 2023, we scheduled and conducted an introductory meeting with the Executive Director of the Alabama STEM Council, a technical kickoff meeting with the HumRRO evaluation project team and the Executive Director of the Alabama STEM Council, regularly scheduled internal project team meetings, and regularly scheduled working meetings with the Director of the Office of Math Improvement (OMI) and Alabama State Department of Education (ALSDE) staff.

Other initial activities we have completed include reviewing ANA-related documents to build understanding and inform our evaluation activities. While we have reviewed many materials, the key documents we have reviewed to date include:

- Alabama Achieves: A Strategic Plan for a New Decade (ALSDE, 2019)

- Bridging Policy to Practice: A Strategic Data Management Framework for the Alabama Numeracy Act (Associated Research & Knowledge Specialists, LLC., 2023)

- SB171 – ACT #2022-249 (Orr & Melson, 2022)

- SB300 – ACT #2023-340 (Orr, 2023)

- Teacher Observation Tool (Learning Lab – Cognia, 2021)

- The Alabama Coaching Framework (ALSDE, 2020)

As input for establishing the requisite DSAs, we prepared a matrix with data requirements to support the process and outcome evaluations, as well as the various supplemental studies. This matrix outlines the sources and elements required for each data collection method. This information will be integrated into the DSAs.

We developed a draft logic model to serve as a roadmap for the ANA evaluation. We carefully reviewed the ANA statute to develop the draft logic model. This model outlines the resources and inputs; activities; outputs; and short-, mid-, and long-term outcomes. The draft logic model includes a schedule for when the various entities are expected to begin implementing their assigned tasks. We shared the draft logic model with the OMI Director for input and integrated appropriate revisions.

We also drafted a protocol for conducting informational interviews and focus groups with key stakeholder groups responsible for the implementation of various ANA components (e.g., Office of Mathematics Improvement, including regional coordinators and staff; Office of School Improvement; math coaches; Alabama Math, Science, and Technology Initiative [AMSTI] staff). The purpose of these interviews/focus groups is to gather information about how the various stakeholders are interpreting and implementing the ANA. We will use this information to develop surveys and other data collection tools. We will schedule and conduct these informational interviews and focus groups once relevant DSAs are in place.

Remaining Year 1 Evaluation Activities

We continue to work with OMI and ALSDE to establish a DSA that will enable our evaluation team to conduct Year 1 planned data collection activities. Once the DSA is in place, we will begin gathering Year 1 process and outcome evaluation data from the limited- and full-support schools. The process data will focus on the schools' implementation of the ANA. Evaluation efforts in subsequent years will focus more on evaluating the quality and impact of the implementation processes. The outcome data will focus on state assessment scores and performance ratings (e.g., math coaches). We present in Appendices C and D the remaining Year 1 activities we will complete (January through September 2024).

Collaboration with OMI

Given the important role it plays in interpreting and implementing the ANA, OMI will play a crucial support role in the ANA evaluation. We have already established a productive working relationship with the OMI Director through weekly working meetings when we discuss immediate and longer-term evaluation activities and data needs. For example, some recent conversations have included discussions about how best to notify limited- and full-support schools about their needed support for the ANA evaluation, procedures for administering surveys to ensure links are not blocked, conducting a case study on a smaller sample of limited- and full-support schools, to name a few topics. During our regularly scheduled meetings, the OMI Director has helped us to refine our understanding of the ANA and the roles of the various stakeholder groups implementing different ANA components. The OMI Director has also been accessible outside the scheduled meeting times to answer questions as needed.

Potential Concerns

HumRRO has established collegial, cooperative relationships with key OMI and ALSDE staff and are working closely with them to complete Year 1 evaluation activities. However, we note several potential concerns that may impact our evaluation of the ANA:

- **Data infrastructure**. We agree with the concern expressed by Associated Research and Knowledge Specialists, LLC in their report addressing the data management requirements of the ANA: "[…] the ALSDE and OMI are constrained by the limited resources available to them. The new systems, tools, and databases required to support ANA will need to be developed, integrated, and supported as an addition to the Department's existing responsibilities" (ARK, 2023, pp.6–7). Our evaluation approach incorporates both the collection of new data and the use of data collected as part of the ANA's implementation. The outcome evaluation task particularly relies on complete and accurate data on educator and student performance. Data infrastructure limitations threaten the quality of data and, by extension, the reliability and validity of evaluation findings.

- **Vague statute language**. The ANA establishes several short- and long-term goals related to statewide math proficiency, and it generally indicates the resources and procedures that will lead to accomplishing those goals. While the statute outlines the required tasks that the various stakeholders must implement, guidance and options for implementation are not clearly articulated. As an example, the ANA states that math coaches will be required to *work with the principal on family engagement* and *engage in ongoing learning*. Another example involves K–5 teachers—the ANA states that each teacher shall *build fluency with procedures on a foundation of conceptual understanding,*

strategic reasoning, and problem solving over time and *provide students access to tools, including any available technology, that support mathematical thinking*. There are many ways that these tasks can be implemented, with some methods more effective than others. Because the conditions and environment are likely unique to each school, we understand and agree that the statute should allow for some flexibility in implementing the required ANA-related tasks. However, the extent to which math coaches, teachers, and others responsible for implementing the ANA share a common understanding of how these requirements are operationalized will impact the ANA evaluation. As indicated earlier, much of the Year 1 ANA evaluation will be dedicated to identifying and documenting the specific procedures and processes each stakeholder is putting in place to implement the ANA. We also will work with OMI and other stakeholders during Year 1 to determine the metrics for measuring the quality of the ANA implementation procedures and processes in subsequent years. All stakeholders must be clear on how the ANA requirements will be operationalized to determine what is appropriate and obtain the most accurate measurement of effectiveness or quality.

References

Alabama State Department of Education (ALSDE) (2019). *Alabama Achieves: A Strategic Plan for a New Decade*. https://alabamaachieves.org/wp-content/uploads/2020/12/ACHIEVES2020-V20.pdf

Alabama State Department of Education (ALSDE). (2020). *The Alabama Coaching Framework: Ensuring a Single Approach to Coaching*. RMC Research Corporation. The Alabama Coaching Framework: Ensuring a Single Approach to Coaching (ed.gov)

Associated Research & Knowledge Specialists, LLC. (2023). *Bridging Policy to Practice: A Strategic Data Management Framework for the Alabama Numeracy Act*.

Learning Lab (2021). Teacher Observation Tool Resources. Cognia.

Orr (2023). Alabama Senate Bill 300, ACT #2023-300, ACT #2023-300. https://legiscan.com/AL/sponsors/SB300/2023

Orr, A., & Melson, T. (2022). Alabama Senate Bill 171, ACT #2022-249. https://legiscan.com/AL/bill/SB171/2022

Appendix A: Conducting ANA Evaluation Process and Outcome Evaluation Components and Supplemental Studies

Process Evaluation

The primary goals of the process evaluation are to determine (a) the extent to which designated schools implement all ANA requirements with fidelity, (b) facilitators of and barriers that impede schools implementing the ANA, and (c) strategies that schools can use to overcome challenges and barriers to implementing the ANA. Our approach involves gathering process evaluation data each contract year to assess the extent to which designated schools implemented the required ANA processes, with a focus in Years 2–5 on monitoring the quality of the schools' implementation of those processes. The process evaluation activities we complete will address the following research questions:

- A. Were all processes and activities required by the ANA implemented by stakeholders? What factors facilitated or impeded the implementation? How were barriers overcome?

- D. To what extent is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school?

- O. To what extent were the relationships between process and outcomes achieved as expected based on logic models? What external factors impacted the anticipated accomplishments and relationships? Is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school?

Although the process evaluation will focus on limited- and full-support schools, the first contract year will focus only on those ANA components for which guidance has been established. We intend to collaborate and work closely with AIDT and other stakeholders to (a) identify which ANA activities should be rigorously monitored, (b) establish guidance for implementing the ANA activities with fidelity, (c) develop instruments and tools to collect process data and monitor the fidelity with which they are implemented, (d) determine annual criterion expectations, and (e) determine how the Alabama Teacher Observational Tool (ATOT) and the Alabama Teacher Growth Program should be used within the ANA.

Develop Annual Criteria

During Year 1, we will work closely with AIDT to develop criteria to evaluate the quality of implementation of the required processes. We will schedule a virtual meeting with relevant staff to discuss the program logic model and the interim milestones that must be achieved to realize the intended outcomes. We will also discuss the indicators of key program milestones. HumRRO will then draft a series of evaluation criteria for each required process. We will identify criteria for each implementation year to allow us to monitor the extent to which limited- and full-support schools are on track to realize the intended outcomes. At the end of each year, we will review the next year's criteria in conjunction with AIDT to determine if any adjustments are needed. Because our evaluation of the criteria will rely primarily on an annual survey administered to all limited- and full-support schools, we plan to evaluate the full population of schools served by ANA. This will require that virtually all schools participate in the survey, and we will work with AIDT to build a strategy to ensure adequate participation rates.

Identify or Develop Instruments

We will develop several instruments to collect the process evaluation data, including a survey, focus group and interview protocols, and an observational tool. As noted earlier, we will collaborate closely with AIDT and other stakeholders to develop the process evaluation data collection and monitoring instruments. Our focus during the first year will be on monitoring the ANA processes implemented by the designated schools across the state consistent with ANA expectations set for August 2023 through July 2024. In Years 2–5, instruments will be revised to focus on measuring the quality with which the designated schools implemented the ANA processes according to the established criteria.

We will develop a survey to measure the implementation of ANA processes and activities and the Alabama Coaching Framework. The survey will include close-ended and open-ended questions (e.g., yes/no questions, Likert scale questions). We will develop multiple parallel survey versions tailored to specific stakeholder groups. These parallel versions will include items tailored to the group. For example, district-level stakeholders may be asked to report on process implementation at multiple schools within their district. In contrast, principals will be asked to report only on process implementation within their school. Surveys will employ branching logic to ensure participants see only relevant questions, creating a more efficient and less onerous survey experience. We will develop parallel surveys to collect data from OMI regional coordinators, district staff, school principals, math coaches, and teachers.

We will develop separate focus group and interview protocols for each stakeholder group. For example, Office of Mathematics Improvement (OMI) regional coordinators may be asked about the district and school characteristics that facilitate or impede ANA process implementation. In contrast, district staff may be asked to focus solely on school-level characteristics. These protocols will include scripted introductory comments and background information (e.g., purpose, how the data will be used), targeted questions, suggested probing questions, and scripted closing comments. We will develop focus group and interview protocols to collect data from Office of School Improvement (OSI) staff, OMI regional coordinators, district staff, school principals, math coaches, teachers, families, and students.

We will develop an observational tool to structure the process evaluation data gathered during site visits. The tool will include observable indicators (e.g., ANA activities or procedures, behaviors) of required processes and a rating scale to evaluate each indicator's implementation level. The tool will include a checklist of indicators and sections for notes to allow observers to provide context for their ratings. For example, if an indicator for a required process was identified as *Program goals are clearly posted*, observers may rate the indicator as *fully met* (e.g., implemented in all classrooms), *partially met* (e.g., implemented in some classrooms), or *not met,* e.g., implemented in no classrooms). As needed, we will develop separate observational tools for limited- and full-support schools to reflect any differences in the required processes in each school type.

Process Evaluation Data

Our process evaluation activities will include an annual survey, focus groups or interviews, and site visit observations. To the extent possible, we intend to elicit information or probe participants about specific facilitators for or barriers to implementing the ANA as we administer or conduct each data collection method. If given appropriate clearance, we understand that we may access quantitative school data to help us identify factors that impact ANA implementation. We will complete and report on these process evaluation activities annually. For the final

contract year, we will report on that year's activities and provide summative conclusions and recommendations targeted across the entire evaluation.

Survey

We will administer all surveys via HumRRO's survey platform. We will endeavor to survey all relevant stakeholders from the stakeholder groups listed above for the entire population of full- and limited-support schools. We will work with the AIDT and the Office of School Improvement to obtain contact information for potential survey participants. Because potential respondents will likely not be familiar with HumRRO, we will ask AIDT (or another appropriate designee) to send an introductory email informing them of the study, the survey's purpose, and the importance of their participation. We will follow this with an email containing a survey link to each potential survey participant. We will follow up on the initial emails by sending two reminders to encourage participation.

We will also work with AIDT to determine the timing of the survey; however, we anticipate administering the survey during the late fall/winter of the school year. This will allow time for annual processes and activities to ramp up yet be early enough in the year to use the survey results to identify districts and schools for focus groups/interviews and site visits.

For Year 1, the survey will focus on whether the required processes were implemented as expected in full- and limited-support schools. Subsequent years of the survey will be designed to evaluate progress toward meeting established evaluation criteria.

Focus Groups or Interviews

We will identify stakeholders to participate in virtual focus groups or interviews based on preliminary survey results. We will not conduct focus groups or interviews in all full- and limited-support schools, but rather we will identify one limited- and one full-support school in every OMI region. We will use available school-level demographic and performance data to identify schools for this sample, ensuring that we include a range of student demographics, enrollment sizes, geographic locations, and overall and subgroup performance. We will also interview the Office of School Improvement (OSI) and OMI regional coordinators to reflect on and provide context to observed survey data patterns.

These focus groups and interviews will allow us to explore response patterns or themes that emerge from the survey data. Focus groups and interviews will be conducted each year. During the first year, focus groups and interviews will provide context for the initial implementation of required processes and inform the development of the implementation quality criteria. In subsequent years, focus groups and interviews will focus on identifying barriers and facilitators to the implementation of the required processes and determining what efforts were made to reduce barriers and the extent to which they were effective.

For each sampled school, we will conduct five Microsoft Teams-based focus groups/interviews: one with the OMI coordinator for that region, one with staff from the district in which the school is located, one with the school principal, one with math coaches serving the school, and one with math teachers from the school. This will yield 110 focus groups/interviews with schools, plus an interview with OSI, for a total of 111 focus groups/interviews per year. The timing of the focus groups/interviews will depend on the timing of the implementation survey administration, though we anticipate they will occur each spring.

Site Visit Observations

Based on our preliminary survey results, we will identify a sample of limited- and full-support schools to visit and conduct observations. These visits will gather information to cross-validate patterns observed in survey results and provide additional information about implementation of the required ANA processes. We will not conduct site visit observations in all full- and limited-support schools, but rather we will identify and visit select limited- and/or full-support schools in every school district. In addition to the survey results, we will work with AIDT and use available school-level demographic and performance data to identify these select schools. We will include a range of student demographics, enrollment sizes, geographic locations, and overall and subgroup performance. We will request AIDT's assistance in recruiting the identified districts and schools. Specifically, we will ask AIDT to contact the identified districts/schools and confirm their participation agreement. Our proposed partner, Mathematica, will work with districts/schools to schedule each site visit.

Site visit observations will be conducted each year. During Year 1, site visit observations will provide context for the initial implementation of required processes and inform the development of the implementation quality criteria. In subsequent years, site visit observations will focus on identifying barriers and facilitators to implementing the required processes and determining what efforts were made to reduce barriers and whether they were effective.

We will conduct six site visit observations yearly, three at limited-support schools and three at full-support schools. When scheduling the site visits, our team will work with each district/school to coordinate two separate on-site focus groups with families and students, respectively. Two staff will be scheduled to conduct each observational site visit. All site visit staff will be trained on the observational tool and best practices for conducting school observations.

Data Analysis

We will analyze the data collected during the process evaluation activities throughout the year. First, we will analyze the survey data to determine the extent to which schools implement required processes and meet established progress criteria. We will compute descriptive statistics (e.g., frequencies, means, standard deviations) for all close-ended items and analysis of variance (ANOVA) and regression techniques, as appropriate. Results from these analyses will be compared to established evaluation criteria. Open-ended items will be evaluated for themes and patterns within and across data from limited- and full-support schools. As noted, we will use survey results to identify 22 schools for focus groups/interviews and site visit observations.

Focus group and interview data will be largely qualitative in nature. We will complete a content analysis of participants' responses to identify themes and patterns within and across limited- and full-support schools. Site visit data will be both quantitative and qualitative. We will analyze these data similarly to the survey data analyses (described above). Results from our analysis of focus group/interview and site visit observational data will inform the annual review of evaluation criteria and updates to the annual survey.

Outcome Evaluation

The outcome evaluation activities we complete will address the following two research questions:

B. To what extent did the implementation of the ANA improve the mathematics proficiency of students in grades K–5? To what extent was the improvement consistent for all subgroups? What are the characteristics of full- and limited-support schools that make the greatest progress in improving proficiency scores?

M. What were the impacts of the School Turnaround Academy?

O. To what extent were the relationships between process and outcomes achieved as expected based on logic models? What external factors impacted the anticipated accomplishments and relationships? Is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school?

The most important evaluation criterion for ANA is student math performance. While numerous math assessments are available, interpreting students' scores and attributing changes in scoring patterns are both complex and vital to the evaluation. HumRRO will approach the outcome evaluation in Year 1 by cataloging the available outcome data, establishing metrics that can be used to indicate progress, and developing clear and concise methods of presenting those data for multiple audiences. Our recent evaluation of Colorado's accountability system and national evaluation of the impact of the No Child Left Behind (NCLB) Act demonstrate our experience and expertise with outcome evaluations of similar scope and complexity.

Outcome Evaluation Metrics

Year 1 Outcome Evaluation Data

We expect the data needed to address the two outcome evaluation research questions will be provided in a readable format and at the grade-by-school level. During Year 1, we will establish baseline data for all the metrics described above and create data visualizations to monitor trends over time. Draft data visualizations will be shared and revised in an iterative process to ensure that the presentation of data is clear and meets the needs of the evaluation. Our goal is to generate clear and interpretable data that can be used to monitor the effectiveness of the ANA, guide interim policy to improve implementation of the ANA, and help identify LEAs and schools where additional attention might be needed.

Once the data displays are agreed upon, we will populate them with the initial baseline data from 2022–23. All data used to generate the data displays will be provided in electronic format. This will allow the state to verify all findings independently and provide a resource to conduct post hoc analyses if additional research questions arise during the evaluation. HumRRO will provide these data files annually.

We expect AIDT to provide all outcome data. While additional data will be collected for the supplemental studies (see those sections of the proposal), the outcome measures focus on student performance. Student performance measures at multiple levels will be used as variables for investigations conducted for the supplemental studies.

Years 2–5 Outcome Evaluation Data

The same data elements will be collected for Years 2–5 so trendlines can be created based on changes in the data elements established in Year 1. For each data element, we will determine if the data are supportive of the effectiveness of the ANA, not supportive, or inconclusive. These data will be displayed to form the overall metrics for the outcome evaluation in a simple tabular format. HumRRO has prepared these types of data tables for state and federal clients, and they provide a clear, interpretable, easily understood matrix for understanding the impact of the ANA program.

On its surface, outcome data, especially test score data, are easy to interpret. The major challenges for this study will be to (a) associate score changes with ANA and (b) address any conflicting or inconclusive data. The first issue can be addressed through supplemental studies. Schools and LEAs will report data related to educator knowledge and skills, and the performance of math coaches, and data will be collected to evaluate multi-tier support system (MTSS) programs. These data should correlate positively with student academic performance. If ANA functions as intended, the efforts of educators should be reflected in students' performance. The second issue is more complex. For example, if the ACAP data show marked improvements in student performance, but that performance is not reflected in NAEP results, it could call the effectiveness of the ANA and the validity of the evaluation metrics into question. This issue will be addressed using multiple metrics, and our report will focus on conclusions based on the preponderance of the evidence. There may be several reasons the ACAP and NAEP results might not be parallel. For example, they may address sufficiently different math constructs that students' achievement might not be equally reflected on both assessments. It also may be possible that one metric might be more instructionally sensitive than the other. No evaluation should rely on a single outcome metric but introducing multiple metrics can obfuscate results. Responsible evaluations report conflicting data and address those when drawing conclusions. The ANA evaluation has the advantage of using multiple metrics across several years, so early conflicting results can be investigated and verified before the evaluation concludes.

We describe below how we will characterize and use metrics from the available data sources to form an overall evaluation of the ANA program for Years 1–5.

> Improvement over 2022–2023 baseline performance on ACAP math performance in grades 2–5

It is straightforward to examine statewide mean scores in math from one year to the next to determine if overall improvements have occurred. HumRRO will compute a statewide mean difference and an effect size difference for overall math performance by grade level. Using a common metric, we can then track overall performance from one year to the next. These data will be used as a comparison point for analyses of data from NAEP (described below). However, it will be more important to monitor trends in performance across years for schools in full- and limited-support status compared to other schools within the state. Means and effect size differences for math performance by grade must be computed at the school level. From those data, we can determine if the full- and limited-support schools improve more or faster than other schools and we can examine the data for consistency across schools. Because implementation of the ANA strategies, especially coaching and educators' response to coaching may vary from one setting to the next, it will be important to identify outliers before drawing broad conclusions about the program's effectiveness.

Improvements in the percentage of students in full- and limited-support schools scoring at or above grade level in mathematics on the Alabama Comprehensive Assessment Program (ACAP) in grades 2–5 by grade level and by cohort (e.g., third graders in spring 2024 who are fourth graders in spring 2025 and fifth graders in spring 2026)

These analyses will also use the ACAP data (or similar summative math assessment data) but will focus on achievement of "grade-level performance." This kind of analysis allows us to determine if the state is making progress toward ensuring that more students start their next grade ready for the expected content. Focusing on the proportion of students at grade level represents a different lens than examining changes in mean scores and it is an important metric to include in the evaluation. We will examine statewide and school-level performance to track overall trends and monitor differences between full- and limited-support schools and the rest of the schools in the state.

In addition to examining grade level performance, it will be prudent to examine the proportions of students performing at Level 1 on the ACAP. The ANA targets the lower-performing schools in the state for the highest levels of assistance. It is reasonable to expect those schools to have more students scoring at Level 1 than other schools. Early successes in the program may be heralded by changes in the performance of the lowest-performing students, so we may see changes in the proportions scoring at Level 1 before we see significant changes in the proportions of students scoring at grade level. It is prudent to monitor both these metrics and mean scores to provide a full context for describing changes in performance resulting from the implementation of ANA. Differences in mean scores can be judged statistically (e.g., significantly improved). Still, these data give us a more practical and interpretable metric (e.g., How many more kids are reaching proficiency by grade 5?).

Standard scores and state ranking on NAEP math tests

Because NAEP scores are released every 2 years and the results are only presented at the state level, these scores are of limited value for evaluating the impacts of the ANA. However, if the ANA functions as intended, we may see increases in the proportions of students scoring *Basic* and *Proficient* and reductions in the proportions scoring *Below Basic*. If the NAEP scoring trends are parallel to ACAP results, NAEP scores would be considered corroborative evidence supporting the program (NAEP indicates significant trends in its reports).

Percentage of students in Grades K and 5 in all schools scoring at or above grade level on approved state formative math and reading assessments during spring assessment

Much like NAEP, formative assessments can provide context for understanding the results of the ACAP. If the formative assessments address similar constructs, we expect high correlations with ACAP and parallel performance regarding overall gains at the school level and the proportions of students scoring proficient or on grade level. We also expect that a change in scores at the school level from one year to the next would parallel ACAP. Differences between ACAP and formative assessments could signal differences in constructs, instructional sensitivity of the assessments, or expectations related to the definition of on-grade level. We will monitor the formative assessments from year to year in the same ways we monitor ACAP performance, and we will present those findings together.

Number and percentage of students in all schools who started third grade with a math deficiency and completed fifth grade on-grade level based on the ACAP

Students who overcome a math deficiency detected in third grade by fifth grade demonstrate substantial math improvement and establish their readiness for future math success. Monitoring the number of students who accomplish this feat will be an important indicator of the effectiveness of the ANA. However, it is also important to monitor the proportions of students identified in third grade across years. Using SY2022–23 as a baseline, HumRRO will establish two metrics that will be tracked over time. First, the proportion of students identified in third grade will be monitored from one year to the next. Second, the number of students identified in third grade who score on-grade level in fifth grade will also be monitored. By simple subtraction, we can monitor the number of students who overcome their math deficiency in two years (math deficient in SY2022–23 and on-grade level in SY2024–25). If the ANA functions as intended, we should see the proportions of math deficient students in third grade decline over the years, and we should also see the number who overcome their math deficiency increase.

> Within full- and limited-support schools: the number and percentage of incoming students in grades 1 and 2 identified as having a math deficiency

HumRRO will monitor and report the number and percentage of incoming students in grades 1 and 2 identified as having a math deficiency. As described above, it is important to track these numbers by school and overall to determine if there are differences among school types and if there are notable outliers among the data. Understanding these data as they change across grades and implementation years for ANA is vital to the evaluation.

> Within full- and limited-support schools: the number and percentage of incoming students in grades 4 and 5 identified as having a fractional reasoning deficiency

Deficiencies in fractional reasoning can signal long-term challenges in higher-level math, especially algebra, trigonometry, higher-level geometry, and calculus. If ANA can reduce the number of students in grades 4 and 5 with fractional reasoning deficiencies, it could mean substantive improvements for students as they move to middle and high school grades, as well as in postsecondary education. HumRRO will track the number of identified students and the proportion by the school as well as for the overall state population across all years of the evaluation.

> Number and percentage of all students retained in grades K–5 based on math deficiencies by grade level within a local education agency (LEA)

Tracking the number of students retained in grades due to math deficiencies represents a complex metric for establishing the effectiveness of the ANA. Early in the program, more students may be retained because of better monitoring and more accurate determinations of math deficiencies. However, this issue should be resolved in the first year or two after ANA implementation and we should see declines from that point forward. HumRRO will monitor retention of K–5 students for math deficiencies overall, and within LEA by grade level across each year of the evaluation.

> Number of schools reporting a decline from the previous year in the number and percentage of students retained in grades K–5 based on math deficiency by grade level

In addition to monitoring at the state and LEA levels, HumRRO will also compute retention by grade at the school level. Once the baseline is established, we can monitor trends at all levels and identify outliers. It may be useful to identify LEAs or schools that have shown substantive decreases in retentions for math deficiencies so those schools and LEAs can be targeted for

more in-depth study. Retentions (of change in retentions across years) may also be a useful variable to compare to other indicators identified during the evaluation.

School Turnaround Academy Impact

The AIDT and Alabama Department of Education (ADOE) will determine and pilot data collection methods and data systems to evaluate the implementation and impact of the School Turnaround Academy in SY2023–24. HumRRO will provide consulting and advice to help ensure that these systems effectively and efficiently gather the data necessary to draw conclusions about the impact of the academy during Year 1. Minimally, the impact evaluation should include metrics to ensure (a) participation by educators in activities associated with the academy; (b) measurable changes in the instruction, curriculum, schedule, or priorities of turnaround schools; (c) changes in the experiences of students attending turnaround schools; and (d) improvements in student perceptions and performance attributed to turnaround activities. We recognize that the School Turnaround Academy may not immediately impact student math achievement, but the four listed components above will allow Alabama to monitor more immediate impacts that must be met for the academy to succeed.

Once the data collection systems are in place and data are gathered, our evaluation will take two primary approaches. First, we will determine whether there is substantive variance in the implementation measures (e.g., participation, changes in adult behaviors, changes in student perceptions) among turnaround schools. If so, we can generate metrics to indicate stronger versus weaker implementers of turnaround practices. For example, if a smaller proportion of educators from a school attend related professional development, are there fewer or less noticeable changes in instructional behaviors? If educator behaviors change substantially, do those changes coincide with improvements in student perceptions regarding their educational experience? If the turnaround practices vary qualitatively, it may also be possible to determine which practices are more effective for improving the performance metrics that will be developed. This work will begin during Year 2 and continue throughout the evaluation.

The second approach will use student performance as an outcome variable and correlate metrics associated with implementation as predictors of aggregated test scores. This work can also begin in Year 2 but may not yield significant results until the practices have been fully implemented in turnaround schools. HumRRO will establish baselines and trends to track turnaround school performance by grade and across years. If the School Turnaround Academy positively impacts school performance, we expect overall student academic performance to improve as well.

Supplemental Studies

Comparison Study

The comparison study will address the following research question:

C. To what extent do full- and limited-support schools that are assigned a math coach yield better performance than such schools that do not have a coach?

Study Design

This study will use a quasi-experimental design (QED) to assess the impact that math coaches have on school math performance in full- and limited-support schools. The ANA indicates the lowest 5 percent performing public elementary schools will be identified as full-support schools and the lowest 6–25 percent will be identified as limited-support schools in Year 1.[3] As such, we predict identifying approximately 45 full- and 180 limited-support schools (totaling 225 schools) in Year 1 as the intervention schools. Full-support schools will retain their identification for 3 years (at which point they will advance out of the lowest 10 percent performing schools or begin receiving intensive support; we assume limited-support schools will similarly not be re-classified for at least 3 years. Because ANA implementation involves tailored supports based on needs and resource capacity, the comparison group might change over several academic years, making it challenging to find a sufficient number of schools to serve as the comparison group; thus, we propose conducting the following three separate QEDs:

- The Year 2 QED study will be a smaller-scale interim impact study conducted at the end of Year 2 to evaluate the preliminary impact of math coaches to help inform potential changes in subsequent years. We will retrospectively identify intervention and comparison groups during SY2024–25 and examine the short-term impacts the math coaches have on school math performance after two years of implementation (SY2023–24 through SY2024–25).

- The Year 3 QED study will serve as our main impact analysis, spanning three years of math coaching implementation, and allow for the greatest observed impact as math coaches will have had more opportunity to provide additional resources and support to teachers and students. We will retrospectively identify intervention and comparison groups during SY2025–26 and examine the 3-year impact that math coaches have on school math performance (SY2023–24 through SY2025–26)

- The Year 4 QED study will serve as a continued evaluation of the impact of math coaches. We propose separating the Years 3 and 4 studies due to the challenges in defining a sufficient comparison group related to the complexity of the ANA implementation. Because full-support schools will begin receiving intensive support if they do not demonstrate positive gains, it will be challenging to differentiate Year 4 impacts due to coaching versus the intensive support and additional resources that schools might receive. Additionally, we expect that over the 5-year ANA implementation, it will become more challenging to control for confounding variables as schools receive increased support. Despite these challenges, we believe examining any continued impacts coaches have after 4 years of implementation is prudent. While it might be challenging to attribute positive gains to coaches, doing so will give us insight into

---

[3] US Department of Education, National Center for Education Statistics, Common Core of Data (CCD, 2023), *Public Elementary/Secondary School Universe Survey*, 2021-22 v.1a.

whether schools that received coaching support for 4 years look different from those that did not.

Matching

We understand that all schools will receive a math coach by SY2027–28 as capacity builds. For each of the three QEDs, the baseline year will be SY2023–24. For the Year 2 and Year 3 studies, we assume that approximately 200 schools will have yet to receive a coach in each respective year (SY2024–25 for Year 2 and SY2025–26 for Year 3). We will retrospectively identify approximately 200 intervention schools for each study during SY2024–25 for the Year 2 study and during SY2025–26 for the Year 3 study.

We intend to use propensity score matching (PSM) with the nearest neighbor without replacement to find a matched comparison group of schools that were never classified as full- or limited-support schools and have not yet received a math coach between their respective study periods. As appropriate, we may stratify on students' demographic and school demographic variables to achieve closer matches on student achievement when matching. We will examine the number of students assigned preliminarily to each comparison group based on our matching strategy and note if this results in unbalanced group sizes.

The comparison groups will be matched on prior ACAP math achievement (SY2022–23) and student- and school-level demographics (e.g., socioeconomic status, percentage of English learner students). Additionally, we will attempt to match on relevant ANA-related resources in consultation with AIDT to reduce the influence of confounding intervention-related variables. For example, schools might have varying percentages of teachers who have earned the K–5 math coach endorsement and it will be important to control for that when creating groups to help ensure that any perceived impacts are due to math coaches rather than some other intervention-related factor. Once the comparison groups have been created, we will establish baseline equivalence between the intervention and the comparison groups at the student level for each grade content area. We will examine the school-level demographic characteristics of the intervention and reference groups. We will also identify the number of schools with (a) students only in the intervention group, (b) students only in the reference group, and (c) students in both the intervention and the reference groups.

The Year 4 study intervention and comparison groups will be retrospectively identified during SY2026–27 and use a similar matching process as that used in Years 2 and 3. Because ANA will be in its later stages of maturity, we might be unable to identify 200 intervention and comparison schools. If that is the case, depending on the sampling frame and in consultation with AIDT, we will determine an appropriate number of schools to include, preferably between 75 and 100 schools in each group.

Data and Analyses

Our preliminary plan is to perform separate hierarchical linear models (HLM) for each grade 3 through 5 and as the overall school. Using HLM will better account for any variance associated with each school and individual students.[4] We will use student- and school-level ACAP math

---

[4] Studies have used a similar design and approach to evaluate intervention impact: Swain, M., Randel, B., & Dvorak, R. (2019). *Impact Evaluation of Mathematics i-Ready Instruction for Middle School Grades using 2018 – 19 Data*. (2019-109). HumRRO; Wendt, S., Rice, J., & Nakamoto, J. (2019). *A cross-state evaluation of MIND Research Institute's ST Math program and math performance*. WestEd; Randel, B. (2018). *Impacts of Edmentum's Exact Path on Student Reading Achievement.* Century Analytics.

achievement data from SY2022–23 through SY2026–27. While the intervention occurs at the school level, using student-level data will reduce our standard errors and result in more precise model estimates. This will also allow us to examine differential subgroup impacts. Because kindergarten through grade 2 will not have complete ACAP data, we propose to run this analysis as a grade band using the state-approved formative assessment data.

Other covariates include student- and school-level demographics (e.g., free/reduced lunch status, disability status, race/ethnicity, school size, urbanicity) and we will leverage data informed by other parts of the overall evaluation, largely from the process evaluation, to monitor the main resources and components implemented at each school to address potential confounds (e.g., the number of math coaches, percentage of teachers who have earned the K–5 math coach endorsement). We will consult with AIDT regarding the final list of covariates.

Math Coach Evaluation and Student Math Achievement

The math coach evaluation and student math achievement study will address the following two research questions:

E. To what extent do performance evaluations of math coaches by principals and regional coordinators in full- and limited-support schools relate to differences in math achievement?

I. To what extent do principals' and regional coordinators' ratings of coaches explain variance in principal and coach evaluations of teachers?

Math Coach Performance and Student Math Achievement

We understand AIDT and the ADOE will develop or adopt measures during Year 1, which will be used by principals and regional coordinators serving limited- and full-support schools to evaluate their math coaches' behaviors in Years 2 and beyond. HumRRO will provide consulting expertise to help ensure that the tools used to create the ratings address important coaching behaviors. Our preference is for behaviorally anchored rating scales, with clear and descriptive exemplars, that allow raters to rate coaches accurately. Specific rated items or behaviors can be aggregated into scales verified using confirmatory factor analyses. We anticipate that the coaches will be rated across as many as four distinct scales, with an overarching coaching quality indicator generated by aggregating across scales.

Ideally, multiple ratings will be collected for each coach. The multiple ratings can be generated by having multiple raters (e.g., three principals and a regional coordinator) and across time (e.g., fall and spring ratings). This will provide evidence that the scales can be used to make accurate and consistent ratings as well as allow for coaches to demonstrate improvement in their performance over time.

The next step will be to determine if the coaches' ratings relate to changes in student performance in full- and limited-support schools. We will begin by using simple correlations to determine if there are relations between student performance and scale ratings (any identified scale and the overall coaches rating). For this purpose, student performance will be defined as "changes from baseline performance." By using a change metric rather than a simple aggregated score we will eliminate much of the covariance between student scores and school-level demographic variables. The change metric will also be the more important data when considering ANA's goals. These analyses will be conducted by school and by grade to determine if there are coaching behaviors that may differentially impact specific grade levels (e.g., grade 2 versus grade 5). These analyses will use multiple outcome variables (e.g., ACAP and formative math assessment scores).

Math Coach Ratings and Teacher Evaluations

The ANA will establish performance metrics for multiple levels of educators, including teachers and math coaches. In this system, teachers will be evaluated by principals and coaches and coaches will be evaluated by principals and regional coordinators. Because both principals and coaches will evaluate teachers, their ratings provide an important intersection for evidence of the validity of the ratings. Assuming the teachers are rated based on the same construct (e.g., math knowledge or math pedagogy), the correlation between coaches' and principals' ratings could be interpreted as a convergent validity coefficient. A strong positive correlation would be one source of evidence supporting the teacher evaluation component of ANA. We consider this ANA component to be a vital source of data for determining the overall program impact and effectiveness. If the ratings were uncorrelated, it would signal that the teacher evaluations were not consistently applied or that they did not adequately represent the intended construct. Inconsistent ratings can occur when the instruments do not function as intended or when the raters do not adhere to behavioral benchmarks when making their ratings.

It is also important to compare principals' and regional coordinators' evaluations of coaches to the coaches' evaluations of teachers. These comparisons can inform whether the system is functioning coherently, and it can help us monitor potential drift in terms of how educator performance is defined across all academic years. First, we will determine if high coach ratings by principals and regional coordinators are associated with higher educator ratings by coaches. If so, it could signal that principals and regional coordinators encouraged teachers to adopt the practices that the coaches promoted before the coaches were assigned. In this scenario, the coach promotes even more of the practices supported by the principals and coordinators and teacher ratings are high from all parties. If the opposite is true (i.e., low coach ratings are associated with low teacher ratings), it could signal that coaches may be encouraging practices that principals and coordinators did not promote. This could represent incoherence in the system, where leaders support differing practices for teachers whereby teachers are not provided consistent guidance. It is also possible that principals' and coordinators' ratings of coaches may not be associated with coaches' ratings of teachers, in which case follow-up investigations may be necessary to determine the root cause of the discrepancies. For example, administrators who may struggle to encourage teachers to adopt more effective practices may rate coaches highly, even as the coaches rate their teachers poorly.

The relationship between the ratings of coaches and teachers will be thoroughly explored. If the results are not conclusive, HumRRO will conduct brief focus group interviews with a sampling of coaches, principals, and regional coordinators to obtain context that helps explain the results. We anticipate conducting as many as six virtual focus groups per year of the evaluation, beginning in Year 3. The results of these analyses will provide support to draw conclusions regarding the effectiveness of the ANA, particularly as the ANA functions as an evaluation of teacher knowledge and pedagogy.

MTSS and Student Math Achievement

The study to examine the effects of MTSS on student math achievement will address the following two research questions:

  F. To what extent is the Alabama Framework for MTSS (Multi-tiered Systems of Support) being implemented in Grades K–5?

  G. To what extent do ratings of implementation of MTSS (reported in F above) within schools relate to the distribution of students within tiered placements?

We will work closely with AIDT and state-level stakeholders to (a) determine the extent to which MTSS implementation fidelity at the school level relates to student academic achievement in grades K–5 by performance level and grade and (b) assess the relationship between student placement in MTSS tiers and their math performance on the state summative assessment and how that relationship varies with the level of MTSS implementation fidelity. We will address both questions for the full- and limited-support schools separately.

The potential data sources we will use for this study include MTSS Fidelity of Implementation results; student, school, and LEA characteristics; screening assessment data; formative math assessment data; ACAP scores; and tiered placements. At this time, we anticipate our analyses of these data will include descriptive statistics to describe characteristics and key variables; comparative analysis to determine progression by cohort; regression analysis to determine relations between implementation and outcomes; comparative analysis by cohort, level of support, grade level, student performance level, and student subgroups; regression analysis to determine relationships between implementation and student placement in MTSS tiers; and regression analysis to determine relationships between typical tier placement and outcomes.

Alabama's MTSS aims to support the whole child through a proactive, team-driven approach that engages all stakeholders (state, regional, LEA, school community, family, and students).[5] The MTSS aligns with the 2020 Alabama Achieves Strategic Plan[6] and uses a Problem-Solving Team (PST) model to guide general education intervention services for all students with academic or behavioral difficulties. Alabama considers the PST model a central factor in schools' successful framework implementation.[7] Each LEA develops its own context-specific MTSS plan,[8] which is based on the MTSS framework and guidance provided by the Alabama State Department of Education (ALSDE).[9] This supplemental study will evaluate the relationship between student math achievement and implementation fidelity to the state framework.

We will draw on the strengths and perspectives of partners and participants. We will gather their feedback on the proposed design before finalizing our evaluation plan. We will be flexible and adaptable to the specific context of the schools included in the study and maintain regular communication with stakeholders throughout the process to enhance the study's credibility and usefulness. To assess differences by support strategy, we will compare findings for full- and limited-support schools.

Year 1 Key Activities

As noted, we will work closely with AIDT and other key stakeholders in Year 1 to build consensus around the objectives of the MTSS supplemental study, the measures to use for evaluation, and the mechanisms for collecting data. We will then work to identify indicators within the state's data collection and monitoring of school-level MTSS implementation and consult with stakeholders on developing tools that assess implementation fidelity with the state framework. We launch the study with a fully articulated evaluation plan that stakeholders and

---

[5] ALSDE. Alabama multi-tier system of supports AL-MTSS.

[6] ALSDE. Alabama achieves: A strategic plan for a new decade. 2020.

[7] Alabama Administrative Code, Chapter 290-3-102(19) (b) 7 and Chapter 290-3-1-02(20) (a).

[8] LEAs (that is, LEAs) draft guidelines, handbooks, or toolkits to describe MTSS implementation in their context, examples include (1) Elmore County Public Schools 2022-23 Multi-Tiered Systems of Support Framework, (2) Birmingham City Schools Multi-Tiered Systems of Support Guidelines and Toolkit, and (3) DeKalb County Schools MTSS Reference Manual.

[9] The PST Guidance Manual described here is an example of guidance provided by ALSDE.

evaluation participants understand, support, and buy into. The plan will specify data collection elements, procedures, timelines, and responsibilities. Finalizing this plan in collaboration with affected partners and interested stakeholders is a critical precursor to successfully conducting the work in Year 2. Our next activity will involve selecting high-quality MTSS implementation indicators that apply to all schools included in the study: for example, indicators from a common Implementation Fidelity Rubric or School Implementation Assessment Tool that includes specific criteria for the essential components of the MTSS framework and school ratings along a continuum. We will rely on data collected or provided by the state. We anticipate that pilot efforts for Cohorts 1 and 2, updated guidance from state and regional MTSS team members,[10] and implementation partnerships with external organizations such as the University of Alabama will inform the implementation indicators.[11] As we complete Year 1 activities, we anticipate a potential challenge given that standards for implementation fidelity and/or implementation assessment tools will not yet be approved or widely used, resulting in limited data collection. If encountered, we will mitigate this challenge by identifying proxy indicators of implementation fidelity for early years' analyses as well as consult with AIDT and other stakeholders on a plan for collecting the proxy indicators.

Years 2–5 Activities

During Years 2–5, we will collect and analyze the relevant study. The quantitative data might include descriptive statistics, correlational analyses, and regression analyses to identify patterns and relationships between MTSS implementation and student math performance. As we interpret and report findings, we will collaborate with core partners to adjust our evaluation plan as needed and present the results in ways designed to help inform policy decisions.

Teacher Math Pedagogy and Student Math Achievement

The teacher math pedagogy and student math achievement study will address the following three research questions:

> H. What are the status and gains in math knowledge and skills of K–5 teachers (e.g., as perceived by the math coach and/or principal)?

> J. To what extent do ratings of the math knowledge and skills of K–5 teachers within full- and limited-support schools (e.g., as made by coaches or principals) account for differences in student performance on formative and summative assessments in math?

> N. What were the impacts of the Instructional Leadership Framework?

Previous research suggests teachers' math knowledge relates to elementary students' achievement gains in math.[12] Across Alabama, the ALSDE has hired math coaches to strengthen teachers' math instruction and knowledge in full- and limited-support schools to improve math proficiency of grades K–5 students. Working in close partnership with AIDT and other key stakeholders, we will (a) document the *status* and *improvements* in mathematics

---

[10] Shankles, K. (2022). AL-MTSS Educational Specialist and 12 Regional MTSS Specialists. Alabama Board of Education meeting notes, 2022. https://aplusala.org/blog/2022/08/11/across-the-board-august-2022-key-takeaways-from-the-alboe-meeting/

[11] An example of a partnership that can inform selecting implementation indicators is the grant ALSDE awarded to the University of Alabama's School of Education Department. https://www.uab.edu/education/home/soe-news/ennis-to-lead-multi-tiered-training-grant-with-alabama-state-department-of-education

[12] Hill, H. C., Rowan, B., & Ball, D.L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42(2),* pp. 371–406.

knowledge and skills of K–5 teachers within full- and limited-support schools and (b) determine the extent to which the differences in students' math performance in grades K–5 assessments are due to changes in teachers' math knowledge and skills.

The potential data sources we will use for this study include math coaches' and principals' ratings; growth metrics; Alabama Teacher Observation Tool data; Learning Mathematics for Teaching data; ACAP scores; data on teachers' knowledge and skills; student, school, teacher, classroom, and LEA characteristics; Alabama Standards for Instructional Leaders' ratings; and select summative data (e.g., i-Ready, STAR). At this time, we anticipate our analyses of these data will include descriptive statistics to describe teacher knowledge and skills and numerous regression analyses to determine various (e.g., the relation between teacher knowledge and skills and student outcomes, the relation between principals; leadership and teachers' performance).

Our approach will align research questions and data collected with appropriate descriptive and regression analyses. When appropriate, we will explore how results vary with baseline characteristics of students, teachers, classrooms, and LEAs using data from the student information system or collected by the ALSDE on schools and teachers. This information will help us contextualize findings and understand why components of the Alabama Numeracy Act (ANA) did or did not lead to expected outcomes.

Year 1 Key Activities

We will work closely with AIDT and other stakeholders to identify a research plan and select specific measures and methods of monitoring teachers' knowledge and skills that the state can feasibly collect and track. After reviewing the metrics that are already being collected, including ratings from math coaches and principals, growth metrics, or data from the Alabama Teacher Observation tool, we will recommend additional validated measures of teachers' knowledge, such as the Learning Mathematics for Teaching tool.[13] We will determine the exact measure and frequency of data collection in consultation with the state-level stakeholders by the end of Year 1. We will draw on our familiarity with these measures and other existing measures to determine the most cost-effective and accessible measures.

Years 2–5 Key Activities

We understand that the state will collect data on teachers' knowledge during Years 2–5. We will analyze the information collected from K–5 teachers in full- and limited-support schools to calculate the proportion of teachers who show grade-appropriate skills and math knowledge. We will analyze how teacher knowledge and skills are associated with student outcomes after controlling for background characteristics, including student, classroom, teacher, school, and LEA characteristics, especially lagged student outcomes. We will conduct these analyses without controls for prior teacher knowledge to capture the overall relationship between teacher knowledge and student outcomes. We will also conduct these analyses with controls for prior teacher knowledge to capture how gains in teacher knowledge affect student outcomes. We will examine the extent to which the relationships between teacher knowledge and student outcomes vary with years of coaching received by the teachers. Evaluations will begin in Year 2 and continue during subsequent years. Finally, we will examine the associations of implementation of the Instructional Leadership Framework with principals' leadership, teachers' knowledge and skills, and student outcomes, controlling for background characteristics,

---

[13] Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education, 35(5)* pp. 330–351.

particularly the lagged values of each of these outcomes. We are sensitive to the challenges that might occur due to attrition in the teacher sample for this study. We will mitigate this challenge by working with AIDT to consider alternate evaluation approaches, including Bayesian interpretation of estimates, which work well even with small sample sizes.

Effectiveness of Screening Assessments

The study to examine the effectiveness of the screening assessments will address the following research question:

> K. To what extent do required screening and diagnostic assessments identify students who are subsequently identified as needing tiered services and/or receive diagnosis relating to math (e.g., specific learning disability or dyscalculia)?

This study will require available score data from required math screening and diagnostic assessments administered in limited- and full-support schools. We will also require data on subsequent student classifications into needing tiered services or having a math-related diagnosis. Using these data, we will calculate classification accuracy rates, sensitivity, and specificity of the required assessments.

Based on the results from these analyses, we will make recommendations to AIDT regarding the screening and diagnostic assessments that are most effective in accurately identifying students who will need future math-related support. This will inform improvements to required processes related to student identification for intervention.

We will request AIDT's assistance in identifying all math screening and diagnostic assessments used across the various districts serving limited- and full-support schools. We will also request assistance with collecting any district data that are not maintained by the state.

Unintended Consequences of the ANA

The unintended consequences of the ANA study will address the following research question:

> L. What positive and negative outcomes emerged within schools, LEAs, ALSDE, and other stakeholder groups that were not anticipated as a result of the implementation of any component of the ANA?

Investigating the unintended consequences of implementing any program can be challenging. Programs have a life cycle and different consequences can occur at varying stages of that life cycle. Some of those consequences have been observed frequently enough that they are no longer unanticipated, such as the performance scallop that often occurs when a new assessment is implemented. Educators do not initially know what to expect, so student performance often drops when a new test is administered. Then, in the second year, student performance increases substantially as educators adapt their practices to accommodate the new assessment. The performance then tends to level out for Years 3 and beyond. Other consequences can be much more difficult to anticipate. For example, new programs that target teacher performance may increase the requirements for professional development that teachers are expected to attend each year. These programs may improve teacher performance, but they may also decrease teacher morale and job satisfaction if the professional development is perceived as burdensome or punitive (e.g., only required for teachers in low-performing schools). Similarly, programs designed to promote higher-order thinking among students, when coupled with strong accountability consequences can have the opposite of the intended effect.

In some prior studies, HumRRO has found that educators reduce the demonstration of highly complex problem-solving to simple memorized algorithms in their efforts to promote higher student scores. This can be especially problematic if the state uses highly memorable or readily available assessment items.

There are many ANA components, and it is certainly possible that unintended consequences could occur at all levels of the education system. These consequences could impact students, parents, teachers, school and district administrators, regional coordinators, the ALSDE, and potentially many others. The best method for discovering these consequences is to discuss the ANA with members of stakeholder groups.

HumRRO will conduct a series of site visits to investigate the unintended consequences, positive and negative, of the ANA. Site visits will begin in Year 2 and occur each year. Site visits will be scheduled to cause the least disruption for schools and districts, while also gathering the most relevant information. Ideally, the site visits would occur in the fall, with a strong chance of score reports being accessed. Score reports represent one key lever that can instigate unintended consequences, so referencing them while they are relevant will increase our likelihood of gathering accurate information. HumRRO will coordinate with AIDT to schedule site visits. These site visits will occur alongside those proposed for the process evaluation to maximize information gathering and minimize cost and burden to schools.

HumRRO will develop semi-structured interview/focus group protocols to facilitate site visits. A separate protocol will be developed for each stakeholder group, but with parallel questions to allow for triangulation and verification of interview/focus group results. For example, both teachers and students will be asked about instructional pedagogy, but students will be asked in age-appropriate ways. Questions will be semi-structured to allow interviewers to pursue topics that arise naturally through conversation. We will specifically ask each stakeholder about the benefits of the ANA and any concerns they might have or challenges they've encountered regarding implementation or impact.

HumRRO will work with AIDT to gain access to stakeholders, and we will schedule our site visits to maximize the information we can collect and minimize any disruptions. We anticipate making two 1-week trips, visiting multiple districts and schools during each. We plan for these trips to occur during the fall, with two HumRRO researchers operating as a team for each trip. We will conduct virtual focus group discussions with AIDT before conducting in-person visits to districts and schools. For each trip, we will strive to interview the following stakeholder groups:

- Regional coordinator(s) – one or more depending on proximity and availability.

- Mathematics coaches – we plan to meet with 3–8 coaches in a single focus group.

- Principals – one-on-one interviews at schools; two to four principals per trip.

- Teachers – all math teachers within a school, two or three schools.

- Parents – we will announce our arrival and schedule an after-school parents' meeting for those who are willing to attend.

- Students – within a school, two to four focus groups, limited to 20–30 minutes maximum.

We recognize the challenges associated with holding parent and student focus groups. Parents will only attend if available and interested, so we recognize it will be unlikely that we will obtain a representative sample of parent voices. However, if there are substantial positive or negative

issues raised by parents during these focus groups, we may recommend more representative metrics to verify and quantify parents' perceptions of ANA. Meeting with students can also be challenging for other reasons. To the extent there are parameters regarding our contact with students, our team is willing to undergo any required background checks that schools may require.

Site visit reports will be created by conducting qualitative analyses of the gathered interview/focus group data. Recordings and transcripts will be used for focus groups with adults. Students will not be recorded. Transcripts will be category analyzed by question and by topic, including topics that arise during a discussion that extends beyond pre-planned questions. Fall site visits will inform spring site visits and protocols will be adjusted to address unintended consequences as they are discovered. A brief memorandum will be created following each site visit cycle. The larger evaluation based on the site visits will be included as a chapter in the annual reports.

Stakeholder Awareness and Satisfaction

The stakeholder awareness and satisfaction study will address the following research question:

P. To what extent are stakeholders aware of, and satisfied with the implementation of the ANA?

Stakeholder awareness of and satisfaction with ANA implementation can best be ascertained in a two-pronged approach. The first aspect is to gather information about stakeholder perceptions of the ANA using a less structured, and less generalizable, approach. Site visits described in the unintended ANA consequences and process evaluation studies provide an excellent means of gathering qualitative data on ANA implementation and its impact on multiple stakeholders. The data gathered during those site visits can inform the development of items for the annual implementation survey administered as part of the process evaluation. Response data from these survey items can then be used to establish quantifiable and generalizable data regarding stakeholder awareness and satisfaction.

Data from smaller groups of stakeholders (e.g., ALSDE staff, regional coordinators) will be gathered more directly through interviews and focus groups conducted as part of the process evaluation. We also propose establishing routine annual surveys for parents and students. Each survey will be tailored to the stakeholders receiving it. Survey items will be developed to allow for triangulation across groups where possible yet allow for specific topics to be addressed by a single group when necessary. For example, questions regarding the interpretation of student-level score reports might be most appropriate for parents, while the use of student performance reports for academic goal setting might be most appropriate for students.

We will administer the parent and student surveys using HumRRO's platform. Recipients will be supplied with a link to the survey via email and the survey will be anonymized to ensure the confidentiality of results. HumRRO will rely on cooperation from AIDT to assist us in accessing email addresses for a representative sample of each group. Paper versions of surveys may be sent on request as an accommodation. Similarly, Spanish translations of surveys will be made available as needed, on paper or online.

HumRRO will coordinate with AIDT to determine the best method to survey students. Ideally, student surveys could be appended to a math assessment the students are already taking online. Student surveys will be written at an appropriate grade level for the students receiving them and will be limited to very few questions (no more than 10 simple Likert-type items).

Survey results will be aggregated for each stakeholder group and summarized in an annual report. If the survey questions support the creation of scales related to satisfaction or awareness, those scales will be tracked by year to determine how perceptions of awareness and satisfaction change over the life of the program. All surveyed samples will be compared to population data to gauge the overall representativeness of the sample and generalizability to the state. This two-pronged approach ensures that we develop the best questions for our survey by talking directly to stakeholders about their experiences with ANA so when we collect and describe stakeholder survey data, those data represent an accurate evaluation of awareness and satisfaction across the state.

Cost Effectiveness Analysis

The cost effectiveness analysis study will address the following research question:

> Q. What are the overall costs and actual or anticipated financial benefits of the ANA?

This study generally requires us to summarize the costs and benefits of the ANA and use that information to estimate its cost effectiveness and to conduct cost-benefit calculations. To the extent possible, we will do this work for specific components of the ANA, including the new math coaches, the School Turnaround Academy, the Instructional Leadership Framework, and the MTSS (which is a component of the Instructional Leadership Framework), while acknowledging it could be difficult to disentangle some of the costs and benefits of those interventions. To conduct this work efficiently, we will consult with AIDT, programs supporting the ANA, LEAs, and schools early in the project's first year. These consultations will help us develop and design appropriate data collection instruments to capture ANA costs, especially for any specific types of cost data we should collect contemporaneously, to reduce the possibility of recall error. We intend to collect a rich set of data, which we will analyze early in 2028 so the summative report produced in spring 2028 can include the results.

We will conduct both a cost-effectiveness and a cost-benefit analysis. One benefit of a cost-effectiveness analysis is that one does not have to monetize all the benefits while an advantage of a cost-benefit approach is that it enables one to summarize the findings with a single number—the internal rate of return (IRR), which is the interest rate that balances the costs and benefits. A larger IRR suggests the program is more effective, and we can compare IRRs for the ANA and its components with IRRs for other types of educational investments, including ones that do not target math. In contrast, we can use cost-effectiveness results only to compare the ANA with programs with similar impacts on similar outcomes.[14]

Our recommended approach will be prospective when possible (gathering data in real-time), rather than retroactive, to reduce inaccuracies when recalling something in the past. We will identify the largest cost drivers so we can focus our efforts on those cost components. We will then develop a template with detailed instructions and space for respondents to document how they calculated the key costs. We propose to capture the cost data in a format that is most useful for AIDT as it considers whether to continue with this policy.

---

[14] Hummel-Rossi, B., & Ashdowd, J. (2002) The state of cost-benefit and cost-effectiveness analyses in education. *Review of Educational Research, 72 (1),* pp. 1–30.
Levin, H., McEwan P., Belfield, C., Bowden, A., & Shand, R. (2018). *Economic Evaluation in Education.* SAGE Publications, Inc. https://doi.org/10.4135/9781483396514.
Washington State Institute for Public Policy. (2019) Benefit-Cost Technical Documentation. Washington State Institute for Public Policy.

Cost Data

When possible, we will rely on data collected during other evaluation activities, especially in Years 1–4. For example, we will include relevant questions in the interviews and focus groups we conduct and survey instruments we field during the process and outcome evaluation activities while recognizing the need to avoid over-burdening respondents. This will give us data on a larger set of schools, LEAs, and programs than the set in which we collect more detailed data, in Year 5. During that final year, we will work closely with AIDT, three to five programs, three to five LEAs, and three to five school partners to efficiently gather a more accurate and complete set of cost data. We will purposefully select LEAs and schools with implementation differences to inform future scaling (for example, to reflect differences in the challenges they faced) and programs that appear to be the most important based on discussions with AIDT.

We will consider a broad set of data types (e.g., math coaches' and principals' personnel hours and costs, volunteer hours, materials and equipment, facilities use, professional development) and cost components (e.g., math coaches, School Turnaround Academy costs, Instructional Leadership framework, task forces, MTSS, summer programs, OMI). We will consider fixed costs, such as those for developing and maintaining the ANA, and the marginal per-student costs. Related to this, we will also consider both the short-term costs needed to get the program up and running, most of which might be fixed, and longer-term costs associated with keeping the program going, many of which might be marginal.

We will take care to avoid double-counting costs, primarily by building up costs of the ANA and its components using the ingredients method, which involves collecting estimates of all resources (ingredients) used for each component of the ANA. The approach involves three steps: (a) identify and collect information on all ingredients or resources used, (b) determine the costs of each ingredient at each point in time, and (c) sum the costs within each time period and then weight appropriately across time periods. Mathematica researchers pioneered this approach decades ago,[15] and many education studies have applied it since then, such as the Institute of Education Sciences-funded evaluation of the Teacher Talent Transfer Initiative.[16]

We will ask respondents who can best provide information on the potential sources of costs and work with them to identify the most efficient methods for them to provide us with those data— ideally, so it would minimize the effort on their end while still providing us with complete information.

Benefit Data

The estimated benefits will come from the analyses associated with the outcome evaluation and several associated supplemental studies. More precisely, we will use the estimated impacts of the ANA and its components on math scores to estimate long-term impacts on the future earnings of the students affected. We will translate changes in test scores to changes in long-term earnings using studies that have made that connection most rigorously, such as Chetty et al. (2014).[17]

---

[15] Long, D. A., Mallar, C. D., & Thornton, C. V. D. (1981). Evaluating the benefits and costs of the job corps. *Journal of Policy Analysis and Management*, *1(1),* pp. 55–76.

[16] Glazerman, S., Protik, A., Teh, B., Bruch, J., & Max, J. (2013). Transfer incentives for high performing teachers: Final results from a multisite experiment. NCEE 2014-4003. U.S. Department of Education, Institute of Education Sciences.

[17] Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104(9),* pp. 2593–2632.

Analysis and Reporting

The costs and benefits accrue at very different points in time and costs experienced today generally receive far more weight than those experienced in the future due to discounting. Hence, to summarize the costs and benefits in the cost effectiveness analysis, we will use an appropriate discount rate. When doing the cost-benefit analysis, we can avoid the need to assume a particular discount rate by using the data to calculate IRRs, which are the interest rates needed to balance the expected benefits and costs.[18]

We will also incorporate the impacts of the ANA on grade retention. When students are held back in school, they are likely to spend additional years in school, which costs taxpayers additional money. Being held back also postpones the time when students are likely to start their careers and thus benefit from improved math knowledge. The cost-effectiveness and cost-benefit calculations will implicitly incorporate the changes in costs associated with reduced retention. The cost-benefit calculation will also incorporate the fact that students can start their careers earlier if they are not held back.

We will complete the analyses of the costs and benefits and reporting in the Year 5 of the evaluation.

---

[18] Orhan, K. (2010). Comparing two approaches to the rate of return to investment in education. *Education Economics, 18(2),* pp. 153–165. https://doi.org/10.1080/09645290802416486

Appendix B:
Evaluation Component and Supplemental Study by Research Question and General Timing for Completion

Evaluation Component and Supplemental Study	Research Question(s) Addressed	General Timing[19]
Process Evaluation	A. Were all processes and activities required by the ANA implemented by stakeholders? What factors facilitated or impeded the implementation? How were barriers overcome? D. To what extent is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school? O. To what extent were the relationships between process and outcomes achieved as expected based on logic models? What external factors impacted the anticipated accomplishments and relationships? Is the Alabama Coaching Framework being implemented with fidelity in each full- and limited-support school?	Years 1–5 (8/14/23–9/30/28)
Outcome Evaluation	B. To what extent did the implementation of the ANA improve the mathematics proficiency of students in Grades K–5? To what extent was the improvement consistent for all subgroups? What are the characteristics of full- and limited-support schools that make the greatest progress in improving proficiency scores? M. What were the impacts of the School Turnaround Academy?	Years 1–5 (8/14/23–9/30/28)
Comparison Study	C. To what extent do full- and limited-support schools that are assigned a math coach yield better performance than such schools that do not have a coach?	Years 2–4 (10/1/24–9/30/27)

---

[19] Note that the comparison and cost effectiveness analysis studies are part of the overall ANA evaluation, but activities related to these two studies do not begin in Year 1.

Evaluation Component and Supplemental Study	Research Question(s) Addressed	General Timing[19]
Math Coach Evaluation and Student Math Achievement Study	E. To what extent do performance evaluations of math coaches by principals and regional coordinators in full- and limited-support schools relate to differences in math achievement? I. To what extent do principals' and regional coordinators' ratings of coaches explain variance in principal and coach evaluations of teachers?	Years 1–5 (8/14/23–9/30/28)
MTSS and Student Math Achievement Study	F. To what extent is the Alabama Framework for MTSS (Multi-tiered Systems of Support) being implemented in Grades K–5? G. To what extent do ratings of implementation of MTSS (reported in F above) within schools relate to the distribution of students within tiered placements?	Years 1–5 (8/14/23–9/30/28)
Teacher Math Pedagogy and Student Math Achievement	H. What are the status and gains in math knowledge and skills of K–5 teachers (e.g., as perceived by the math coach and/or principal)? J. To what extent do ratings of the math knowledge and skills of K–5 teachers within full- and limited-support schools (e.g., as made by coaches or principals) account for differences in student performance on formative and summative assessments in math? N. What were the impacts of the Instructional Leadership Framework?	Years 1–5 (8/14/23–9/30/28)
Effectiveness of Screening Assessments Study	K. To what extent do required screening and diagnostic assessments identify students who are subsequently identified as needing tiered services and/or receive diagnosis relating to math (e.g., specific learning disability or dyscalculia)?	Years 1–5 (8/14/23–9/30/28)
Unintended Consequences of the ANA Study	L. What positive and negative outcomes emerged within schools, LEAs, ALSDE, and other stakeholder groups that were not anticipated as a result of the implementation of any component of the ANA?	Years 1–5 (8/14/23–9/30/28)

Evaluation Component and Supplemental Study	Research Question(s) Addressed	General Timing[19]
Awareness and Satisfaction of Stakeholders Study	P. To what extent are stakeholders aware of and satisfied with the implementation of the ANA?	Years 1–5 (8/14/23–9/30/28)
Cost Effectiveness Analysis Study	Q. What are the overall costs and actual or anticipated financial benefits of the ANA?	Year 5 (10/1/27–9/30/28)

Appendix C: Planned Process and Outcome Evaluation Activities January–September FY2024

Year 1 Project Phase	Process Evaluation	Outcome Evaluation
Data Sharing Agreement *Jan 2024*	Work with OMI/ALSDE to establish data sharing agreement(s)	Work with OMI/ALSDE to establish data sharing agreement(s)
Information Gathering *Jan-Feb 2024*	Conduct information gathering interviews or focus groups (FGs) to build understanding and inform data collection instruments	Obtain reports used by OMI/ALSDE for use as potential templates for reporting ANA outcome data
Planning *Feb-Apr 2024*	Identify the ANA components to be implemented in Year 1 Identify indicators of successful implementation of ANA components Develop criteria/metrics to evaluate the quality of implementation of various ANA components; efforts will focus on Year 1, but also consider implementation criteria for Years 2–5 Identify stakeholders within each full- and limited-support school/district to receive a survey Determine procedures and materials for administering annual surveys and conducting FGs and site visits (SVs)	Identify sources for outcome data (student formative and summative performance data, ranking on NAEP math tests, math coach performance data (including collection of tools used to monitor math coach performance), student percentages [scoring at/above grade level, math deficiency, fractional reasoning deficiency, retained) Determine process and establish procedures for OMI/ALSDE to share outcome data Establish outcome data baseline metrics Determine data visualization templates
Design & Data Collection *Mar-June 2024*	Determine a sample of schools for virtual FGs Determine a sample of schools for in-person SVs Develop an annual survey to measure the implementation of ANA processes and activities; the survey to include parallel versions to accommodate specific stakeholder groups (OMI coordinators, district staff, principals, math coaches, and teachers)	Receive data and data file layouts from OMI/ALSDE Review the quality of data for meeting assumptions of proposed analyses (e.g., normality, linearity).

Year 1 Project Phase	Process Evaluation	Outcome Evaluation
	Develop FG protocols to accommodate specific stakeholder groups (OSI staff, OMI coordinators, district staff, principals, math coaches, teachers, families, and students); these sessions will be held to elaborate on and/or clarify survey findings Develop an observational tool for use during SVs; the tool will include observable indicators of required processes and a rating scale to evaluate the implementation of indicators Administer annual surveys to stakeholders (OMI coordinators, district staff, principals, math coaches, and teachers) Conduct virtual FGs with stakeholders (OSI staff, OMI coordinators, district staff, principals, math coaches, teachers, families, and students) Conduct in-person SVs at the identified sample of limited- and full-support schools	
Data Analysis *July-Sept 2024*	Analyze survey data separately by stakeholder group Analyze FG data separately by stakeholder group	Analyze outcome data separately by metric Prepare draft data visualizations of baseline outcome data

Note. Activities may change based on the availability of information required for study planning and design and implementation status of the ANA.

Appendix D: Planned Supplemental Studies Activities for January–September FY2024

Year 1 Project Phase	Math Coach Evaluation and Student Math Achievement	MTSS and Student Math Achievement	Teacher Math Pedagogy and Student Math Achievement	Effectiveness of Screening Assessments	Unintended Consequences of the ANA	Stakeholder Awareness and Satisfaction
Information Gathering *Jan-Feb 2024*	Piggyback on process evaluation information gathering interviews and FGs	Review existing measures and data collection systems covering MTSS implementation, tiered placements, student math achievement, and other student and teacher characteristics; this information will build understanding and inform data collection instruments	Review existing measures and data collection systems covering measures of teacher math knowledge and skills, measures of student math achievement, and other student and teacher background characteristics, including years of coaching received by the teacher	Identify math screening and diagnostic assessments used across the various districts serving limited- and full-support schools	Piggyback on process evaluation information gathering interviews or FGs	Piggyback on process evaluation information gathering interviews or FGs
Planning *Mar-Apr 2024*	Provide support and consultation to OMI on developing tools for Regional Coordinators and principals to use to measure math coaches' behavior during Years 2 and beyond	Work with OMI/ALDSE to recommend refinements to existing measures, new measures, refine data collection systems, and design study	Work with OMI/ALDSE to recommend refinements to existing measures, add new measures, refine data collection systems, and design study	Develop processes and establish procedures for collecting data not maintained at the state level	Piggyback on determining a sample of schools for process evaluation in-person SVs	Piggyback on determining procedures and materials for administering process evaluation annual surveys to parents and students

Year 1 Project Phase	Math Coach Evaluation and Student Math Achievement	MTSS and Student Math Achievement	Teacher Math Pedagogy and Student Math Achievement	Effectiveness of Screening Assessments	Unintended Consequences of the ANA	Stakeholder Awareness and Satisfaction
Design & Data Collection *May-June 2024*		Finalize measures, data sources, and study design Finalize data collection timeline Draft study design report; submit final study design report	Finalize measures, data sources, and study design Finalize data collection timeline Draft study design report; submit final study design report.	Obtain available score data from math screening and diagnostic assessments Obtain data on subsequent student classifications into needing tiered services or having a math-related diagnosis	Piggyback on developing process evaluation observational tool for use during in-person SVs (note SVs will not be conducted for this study until Year 2)	Piggyback on administering process evaluation annual surveys
Data Analysis *July-Sept 2024*		Process data Analyze data to identify relationships between MTSS implementation, tiered placements, and student math achievement, with and without controls for other student and teacher characteristics	Process data Analyze data to identify relationships between teacher math knowledge and skills and student math achievement, with and without controls for other student and teacher characteristics	Calculate classification accuracy rates, sensitivity, and specificity of required assessments Recommend the screening and diagnostic assessments most effective in accurately identifying students needing math-related support	No Year 1 activities	Analyze quantitative and qualitative survey data separately for parents and students Triangulate quantitative and qualitative findings separately for parents and students

Note. Activities may change based on the availability of information required for study planning and design and implementation status of the ANA.