McREL INTERNATIONAL

Curiosity for Better Learning

Sirius Research Prospectus: Implementation & Efficacy Study

Prepared by McREL for Sirius Education Solutions

November, 2022

McREL project team

Dylan P. J. Kriescher, Ph.D.

Haley E. Johnson, Ph.D.

Lisa M. Jones, Ph.D.

McREL is a nonprofit, nonpartisan research, development, and service organization that helps schools, districts, and education agencies improve outcomes for all students. We provide:

- Analysis and alignment of curriculum, instruction, assessment, and standards
- Professional development, learning, and coaching for teachers, principals, and system leaders
- Research, data analysis, logic models, and program evaluations
- School improvement recommendations, action planning, implementation support, and technical assistance

4601 DTC Boulevard, Suite 500, Denver, CO 80237

www.mcrel.org 800.858.6830 info@mcrel.org

Introduction

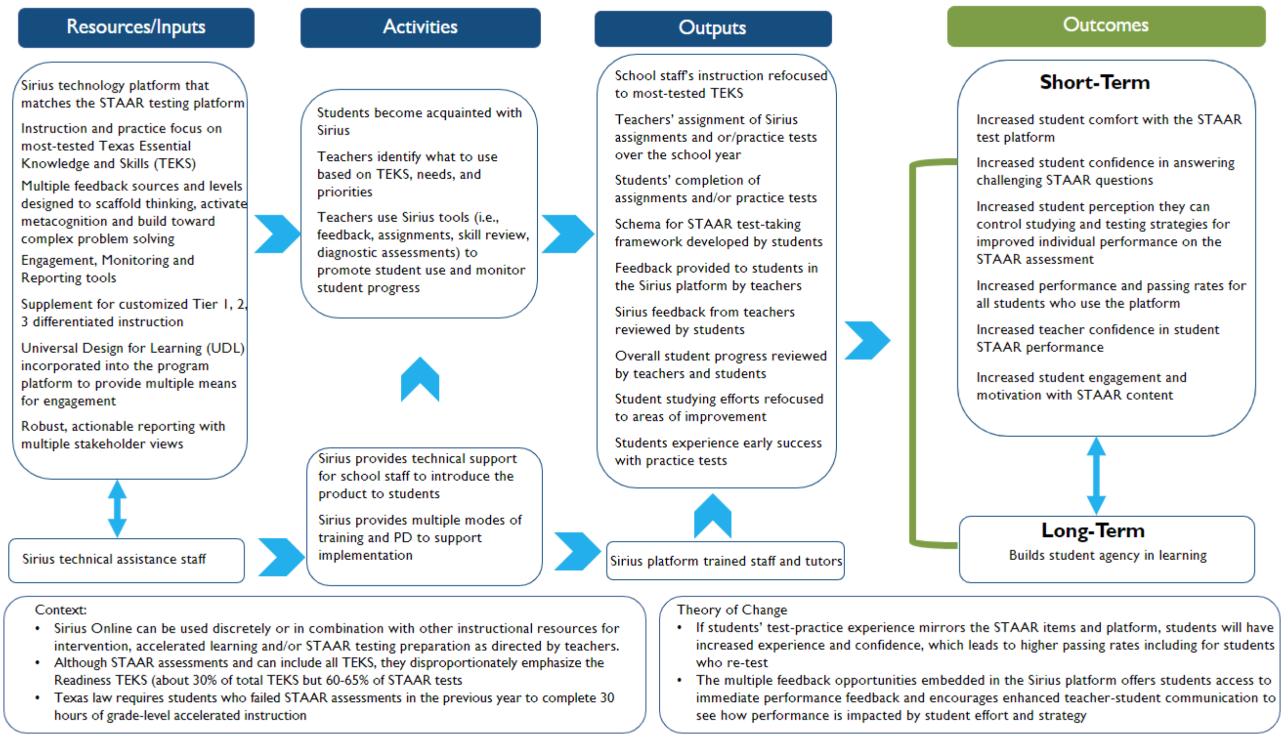
Sirius Education Solutions decision makers are interested in further understanding whether and how the company's test preparation platform can support students' "State of Texas Assessments of Academic Readiness" (STAAR) testing outcomes. The proposed project will leverage current and future relationships with *Sirius Education Solutions* and schools/districts utilizing the Sirius platform to conduct an implementation and efficacy study of the product. Recently, *Sirius Education Solutions* partnered with McREL to co-create a theory of action logic model, identify the research base for the product, and co-design an evaluation plan (phase 1). The theory of action logic model describes the specific change mechanisms that lead to intended outcomes and serves as a guide for *Sirius Education Solutions* continuous improvement processes. Moreover, it guides the implementation and efficacy study of the product. Phase 1 activities are designed so *Sirius* is eligible to meet Tier 4 ESSA evidence requirements.

McREL proposes conducting an implementation and evaluation study designed to make Sirius Education Solutions *eligible to meet* the Tier 3 ESSA evidence level. The purpose of the implementation study is to capture and describe how teachers are implementing the materials in their classroom practice including how they are used as a supplement to Tier 1 classroom instruction. Moreover, the implementation study will enable McREL to capture the contextual factors (external environment, district policies, and district initiatives) that can facilitate or impede effective implementation.

Sirius Education Solutions Product Elements

The Sirius platform was designed with several components and processes in mind for supporting student success on STAAR in subject areas of math, language arts, science, and social studies. Sirius is designed to align with STAAR testing in content, structure, and format, and provides multiple sources of feedback for students. It offers various monitoring and reporting tools for teachers to use and is designed as a supplement to be customizable for Tiers 1, 2, or 3 use. The platform also offers accessible Universal Design for Learning (UDL) engagement features such as "text-to-speech." The theory-of-change behind these Sirius Online components supporting STAAR performance and other outcomes is displayed in the logic model below [Figure 1], as well as in the research foundation section supporting how these components have been associated with learning and other outcomes in prior research.

Figure 1. Logic Model



Research Foundation

The Sirius Education Solutions test preparation platform lacks an independent evidence base that shows its efficacy due to the newness of the product. Currently, the evidence is limited to teacher descriptions of use and changes in students' test performance. Student and staff users' testimonials show they value the supplemental instructional tool. Furthermore, an internal study showed that after adopting Sirius, 65% of students in United Independent School District (ISD) demonstrated mastery on the Algebra I End of Course (EOC) STAAR assessment in 2019 compared to the 39% state-wide average. This descriptive analysis provided a quick look at Sirius' potential to increase student achievement in mathematics.

The scholarly literature is rich with examples of how specific cognitive processes, as well as curriculum instruction and assessment structure, contribute to student learning and academic performance. Sirius developers designed the product to target these cognitive processes so students can improve performance on STAAR. This brief focuses on a few of those processes: metacognition, self-regulated learning, student motivation, and curriculum alignment while also demonstrating how elements of the platform (e.g., feedback and practice) support the interaction of cognitive processes and curriculum instruction and assessment structure that can lead to increased student achievement. Sirius leaders believe that the platform elements that support these processes will increase the proportion of students who pass STAAR assessments the first time and increase the passing rates for students who are re-tested.

Cognitive Processes

Metacognition, Self-Regulated Learning, and Student Motivation

The cognitive processes that underlie student learning and mastery include metacognition, self-regulated learning, and student motivation. The literature shows these three processes interact with one another in the student's environment to facilitate mastery. For example, Tian and colleagues (2018) argued that the interaction between metacognition, motivation, and self-efficacy play a role in the development of self-regulated learning. A growth mindset and self-efficacy are also associated with students' self-regulated learning strategies (Bai & Wang, 2020; NAS, 2018). The National Academy of Sciences, Engineering, and Medicine (NAS; 2018) report on how people learn described the multiple cognitive and contextual factors that influence learning including the interplay of underlying processes (executive functioning, self-regulation, memory, and goal setting). These underlying processes comprise metacognition, self-regulated learning, and motivation, which ultimately lead to learning mastery. Moreover, the NAS (2018) review showed that the interaction of these processes is necessary for mastering complex content including developing problem-solving and critical thinking skills, which are imperative for successful performance on the STAAR. This literature summary briefly describes the overarching cognitive processes and their relationship with underlying mechanisms that contribute to student learning.

Metacognition

Metacognition is a conscious process, monitoring and regulating students' academic and emotional behaviors; it is associated with improving student performance (NAS, 2018). Tachie (2019) showed that strategies such as monitoring, checking work, reflecting on tasks and work, planning, and analyzing tasks needed for learning help students solve mathematical problems.

Metacognitive skills, however, are not necessarily innate—they can be taught. Educators, for example, help students build metacognitive skills through specific prompts and feedback designed to guide students' awareness of the cognitive processes needed to solve problems (Dori et al., 2018; Hsu & Lin, 2017; and Tian et al., 2018). The Sirius platform is designed to help teachers promote metacognition and help students strengthen metacognitive skills by individualizing feedback to students and guide their awareness of cognitive processes involved in problem solving.

Another metacognitive process students experience involves their perception of the test-taking process. Familiarity with the format and structure of assessment may be beneficial for students by reducing cognitive load involved in engaging with the assessment (Choi et al., 2014; Van Gog et al., 2011). The Sirius platform is designed to mirror the STAAR test platform so that students are learning in the same context in which they will ultimately be evaluated. Through this alignment, students may develop a schema which allows them to navigate the assessment, and focus their cognitive resources on content rather than structure (Choi & Kim, 2021; Paas et al., 2003).

Metacognition, however, does not singularly contribute to improved student performance. Tian and colleagues (2018) showed that the association between metacognitive knowledge and student math performance is "mediated by self-efficacy and intrinsic motivation," (p. 1). Moreover, Vettori and associates (2018) demonstrated that students' internal mindset about learning success further mediates the relationship between academic achievement and metacognitive activities and skills. Student motivation will be discussed in greater detail below.

Self-regulated learning

Self-regulated learning (SRL) is associated with students' learning strategies and behaviors by allowing students to control their cognitive processes (NAS, 2018). Lai and Hwang (2021) argue three processes comprise SRL: forethought, performance, and self-reflection. Executive functioning plays a supporting role in students' use of SRL strategies such as monitoring, planning, effort, and goal setting (Bai & Wang, 2020; Rutherford et al., 2018). Executive functioning controls the cognitive and neural processes that regulate thinking, sequencing, initiating, and sustaining and adjusting learning behavior when exposed to feedback (NAS, 2018).

Self-regulated learning, closely related to metacognition, can also be fostered when teachers create the environmental context for learning, "tune-in" to students' interest in learning, apply focused and explicit instructional strategies, and provide meaningful evaluation of student learning (Alvi & Gillies, 2020). The environmental context is influenced by students' perceptions of their relationships with teachers, teachers' beliefs about student development and use of SRL strategies, and teachers' self-efficacy in teaching SRL strategies, which are all strongly associated with SRL implementation (Alvi & Gillies, 2021; DeSmul et al., 2019; and Zee & deBree, 2017). Educators can also provide models of meaningful evaluation. An experiment conducted in a biology classroom showed that students who provided high-quality explanations of their learning (after exposure to a teacher video in which it was modeled) scored higher on problem-solving tasks than those who did not (Baars et al., 2018). The Sirius platform is designed for teachers to provide individualized feedback that could offer students the opportunity to evaluate their learning, which could positively affect their self-regulated learning. Further, the Sirius platform is designed to mirror the STAAR test platform so that students are learning in the same context in which they will ultimately be evaluated.

Student Motivation

Student motivation is associated with students' academic outcomes, including how they perform on high stakes standardized assessments (Andersen & Nielsen, 2016; Bostwick et al., 2019; Rattan et al., 2015; Sriram, 2014). Growth mindset and mastery learning are two constructs related to students' motivation in the classroom. Growth mindset is the underlying idea that intelligence, skill, and ability is not fixed, but instead something that can be fostered (Dweck, 2006). Closely related, mastery learning is focused on students mastering, or understanding and being able to apply, the material and content not just learning for the sake of a grade or test score (Bloom, 1973; Slavin, 1990). Sirius' practice tests and the platform that aligns with the STAAR assessment could potentially positively affect student achievement motivation. When students are repeatedly and consistently exposed to opportunities to learn and test their content knowledge, they could potentially increase their mastery of academic content and recognize that they can learn strategies for effective test performance.

Research findings show that students are more likely to adopt a growth mindset when feedback and grading is mastery orientated (Fernandez, 2021; Harsy et al., 2021; Lenarz & Pelatt, 2020). Although test-taking and test-prep might not seem like a natural setting to foster growth mindset or mastery orientation, there is a growing body of literature related to mastery-based testing. Mastery-based testing creates an assessment environment where students repeatedly take the test until they show mastery. This practice fosters deep understanding of content and students' perseverance in the face of challenging content. Moreover, mastery-based testing removes the high stakes pressure some students may feel when one test accounts for a large percentage of their grade (Lenarz & Pelatt, 2020; Linhardt, 2020; Maier, 2020). The Sirius platform is designed as a mastery-based testing environment because there are no negative consequences of using the platform. Sirius encourages teachers to assign grades on participation and engagement rather than practice test performance. Both the platform and teachers' direction encourage students to answer questions until they feel they have mastered the content.

Curriculum Alignment

Curriculum alignment refers to how curriculum content, its instruction, and its assessment are in-line with each other (Leitzel & Vogler, 1994). This alignment supports students through ensuring that students are being assessed on the intended learning outcomes as well as how they were taught, and also ensures instructors are structuring their content delivery in ways that are focused on the learning objectives. There have been unclear links between curriculum and school achievement tests (Traynor et al., 2020), suggesting poor or poorly done (English & Steffy, 2001) direct efforts to align material and structure of content in courses with their summative assessment. Sirius explicitly designs their platform to reflect the content, structure, and format of STAAR. Students may benefit from this direct alignment in metacognition and self-regulated learning, as well as motivationally in recognizing the link between instruction and later summative assessment.

Through Sirius' explicit alignment with STAAR and guiding of student awareness, students receive feedback on linked activities that may allow them to reflect on, plan, and adjust problem-solving strategies during STAAR assessment (Alvi & Gillies, 2020; Dori et al., 2018; Hsu & Lin, 2017; Tachie, 2019; and Tian et al., 2018), which are important components of both metacognition and self-regulated learning. Similarly, as mentioned earlier, the alignment of Sirius with STAAR may allow for a reduction in cognitive load experiences through schema development during the test-taking process by creating a familiar format and structure (Choi et al., 2014; Van Gog et al., 2011) and so facilitating the use of freed cognitive resources for the test content and acquired problem-solving strategies.

Alignment of Sirius with STAAR may also support student motivation. Curriculum alignment allows students to see efforts in making progress towards learning objectives, demonstrating to students that learning outcomes are flexible with effort and strategy, helping to develop growth mindset (Dweck, 2006). With this alignment also comes support for mastery learning through student awareness of the link between Sirius engagement as a means of preparing for STAAR rather than just as an assignment for an immediate grade or score in the course (Bloom, 1973; Slavin, 1990).

Evaluation Design

Purpose

The intent of the retrospective correlational study focuses on usability for continuous product improvement, marketing to current and future users, and to become eligible to meet ESSA Tier 3 evidence requirements. The evaluation approach used to support this purpose focuses on Context (C), Input (I), Process (P), and Purpose (P), described at the CIPP model. Stufflebeam (2000) developed the CIPP model because he recognized the inherent limitation of measuring outcomes based solely on program objectives. Rather, the CIPP model includes "… a critical evaluation of the program's objectives, what is needed to make the program work, the extent the program is implemented as planned, and what the outcomes are," (Mertens & Wilson, 2018, p. 88).

The CIPP model is designed to evaluate needs and opportunities, participant characteristics, implementation, and intended outcomes. Moreover, the process sheds light on the unintended consequences of a specific product. The findings generated from this study are useful for making decisions about the intended outcomes, resource allocations, changes in product delivery and refinement, and continuous assessment of the product's intended targets and goals (Mertens & Wilson, 2018).

The CIPP approach allows evaluators to focus on the factors which facilitate or impede realizing the intended outcomes. The data collection and analytic strategies build understanding about how users interact with the platform, the components that contribute to use, and how these two elements interact to increase STAAR passing rates for all students who use it, including those who retest. Sirius leaders can use this information to modify specific elements to strengthen use for both teachers and students. Moreover, the theory of action logic model development process with Sirius leaders highlighted specific assumptions that could help or hinder use. This evaluation approach captures the relationship between assumptions, use, and intended outcomes. "Knowing why a project meets its goals is more important than just knowing that it does … an implementation evaluation allows you to put this outcome data in the context of what is actually done," (W. K. Kellogg Foundation, 2004, p. 27). This design asks three fundamental questions related to how the Sirius platform is used, what happens when it is used, and "can the results be attributed to what happens" (Patton, 2012, p. 192.)

Patton (2012, p. 116) created a set of questions to help leaders identify how the evaluation will influence future decisions:

1. What decisions, if any, are the evaluation findings expected to influence?
2. When must evaluation findings be presented to be timely and influential?
3. How much influence do you expect the evaluation to have—realistically?
4. What data and findings are needed to support decision making?

McREL evaluators recommend Sirius leaders consider these questions before the evaluation plan is implemented. This will enable both Sirius and the evaluators to identify specific areas where the implementation data can influence product changes, marketing decisions, and service modifications. The theory of action logic model provides a useful resource to consider these questions.

Evaluation Questions

The evaluation is guided by five over-arching evaluation questions:

1. What, if any, effects do students' performance on Sirius activities have on STAAR Math and Reading/Language Arts performance?
2. To what extent is engagement with the SIRIUS platform associated with improved STAAR scores, student perceptions of controllability and self-efficacy in solving complex problems, and teacher confidence in student performance on the STAAR assessment?
3. How do students use the Sirius Education Solutions platform?

4. How do teachers use the Sirius Education Solutions platform?
5. What characteristics of the Sirius Education Solutions platform have facilitated or impeded the intended outcomes?

To answer these questions, McREL recommends conducting a retrospective, correlational mixed methods study. The study is retrospective in that student assessment data will be collected from the prior year (2022-23). Teachers and school administration will be asked to reflect on the prior year's implementation of the Sirius platform. McREL and Sirius will use convenience sampling to draw from schools, teachers, and students currently using the Sirius products. For Sirius to be eligible to meet the ESSA Tier 3 evidence level, the correlational study must employ statistical controls to account for selection bias (REL Midwest, 2019). All analytic models will control for previous years' assessment scores, students' gender, and economic status. The correlational study can identify differences in outcomes based on students' exposure to and use of the Sirius platform. In addition, it will provide information on the relationship between the variables of interest (Lau, 2017).

Sirius will invite all middle/high school clients to participate in the evaluation study with the goal of including 400 students across 20 classrooms in the study, the minimum amount needed for medium effect sizes for hierarchical linear modeling given the nature of students nested within classrooms with potentially different models of exposure to Sirius. This sample will ideally include students who did not use Sirius in 2021-22 to avoid contamination. Otherwise, degree of engagement with Sirius during 2021-22 as determined by platform data can be used as a covariate, or a sampling frame of non-users can be identified and undergo propensity score matching to construct a proper comparison group for 2022-23 users. Although this is a convenience sample, inviting all clients within these grade levels affords students an equal chance of participation in the study. McREL will create a flyer that describes the purpose of the study, what is required for participation, and appropriate contact information. School leaders will be the audience of the flyer. Upon agreeing to participate, school leaders will provide a list of classrooms and teacher contact information needed to conduct the study. The relationship between Sirius and the schools and districts currently using the products reduces recruitment costs because the evaluation team will not engage in district, school, and classroom recruitment.

Data Collection and Measures

To answer the evaluation questions, data will be collected through the Sirius platform, teacher and student surveys, and student-level administrative data (e.g., demographic information, GPA, student STAAR Math and Reading/Language Arts performance [scaled score and passing status], etc.). Program data will be collected from Sirius throughout the school year to inform the development of the teacher and student surveys based on emerging use patterns. Teachers and students will be surveyed in April of 2023. McREL and Sirius will co-develop the survey items, however, in the below section, McREL provides detail on potential measures for each construct of interest. Administrative data will be collected from school districts after STAAR testing is completed and ready to be shared, likely late Summer. Table 1 provides an overview of where the data will come from and how the data will be collected. Table 2 provides an overview of the specific measures we will look to collect.

Table 1. Evaluation Questions, Design, Measures, and Evidence

Evaluation Question	Source/Measure	Evidence
1. What, if any, effects do students' performance on Sirius activities have on STAAR performance?	-Extant District Data -STAAR outcomes -Sirius assignment and practice test outcomes	-STAAR performance in Math and Reading/Language Arts for all students (including those who retest) who complete assignments and practice tests
2. To what extent is engagement with the SIRIUS platform associated with a. Improved STAAR scores, b. Student perceptions of controllability and self-efficacy in solving complex problems, and c. teacher confidence in student performance on the STAAR assessment? -To what extent is the relationship between SIRIUS platform engagement and STAAR performance mediated by student perceptions of controllability and self-efficacy?	-Platform Use Data -Retrospective Pre/Post Teacher and Student surveys -Extant Data (STAAR)	-Student completion rates for Sirius assignments and practice tests -STAAR Math and Reading/Language Arts performance -Frequency of teacher feedback to students and their access through the Sirius platform -Students reported confidence in answering challenging questions in both Sirius and STAAR -Student users' descriptions of their comfort in taking STAAR assessment and their perceptions of academic self-efficacy and controllability due to Sirius -Descriptions of teachers' confidence in students' ability to perform on STAAR
3. How do students use the Sirius Education Solutions platform?	-Retrospective Pre/Post Student Survey	-Student completion rates for Sirius assignments and practice tests -Student descriptions of their use -Information about the time of year students use the platform -Student expectations about what could happen if they use the platform -Student perceptions of activities and levels of interest in and engagement with Sirius assignments and practice tests -Frequency of students' responses to teacher feedback -Student reasons for using the platform
4. How do teachers use the Sirius Education Solutions platform?	-Retrospective Pre/Post Teacher Survey -Usage data	-Teacher Sirius assignments to students -Teacher descriptions of how they use Sirius -Information about the time of year teachers make Sirius assignments to students -How often teachers provide feedback to students through Sirius -Teachers' goals and purpose in requiring students complete Sirius assignments and/or practice tests -Teacher perceptions of student engagement with Sirius platform
5. What characteristics of the Sirius Education Solutions have facilitated or impeded the intended outcomes?	-Retrospective Pre/Post Student and Teacher Surveys -Usage Data -Extant student data	-Incomplete assignments and/or practice tests -Student and teachers' perceptions about which aspects are most effective for improved STAAR preparation -Student and teacher descriptions of Sirius use barriers and facilitators

Table 2. Overview of Measures to Be Collected for Predictive Questions

Variable Type	Measure-specific Information

Independent Variables	• Sirius performance (Program Data)
	    o Sirius Practice test scores, calculated as:
	        ▪ Last score, or
	        ▪ Growth score (Best score – worst score), or
	        ▪ Average score
	• Composite of Classroom Level Use (Teacher Survey, used as moderating predictor if large enough subsample of use cases)
	    o Cramming vs Distributed dummy-coded
	    o Use of platform features
	• Composite Student Engagement with Platform (Program Data)
	    o Student completion rates for Sirius assignments and practice tests
	    o Frequency of accessing teacher feedback to students through the Sirius platform

Outcomes	• STAAR Assessment Scores in Math and Reading/Language Arts (Admin Data)
	• Student Self-Efficacy, Controllability, and Comfort with Complex Problems
	    o Self-Report Scales in Student Survey
	• Teacher Confidence in Student Performance on STAAR
	    o Self-Report Scales in Teacher Survey

Control Variables	• Previous years' assessment scores (Admin Data) • Students' gender and socio-economic status (Admin Data)

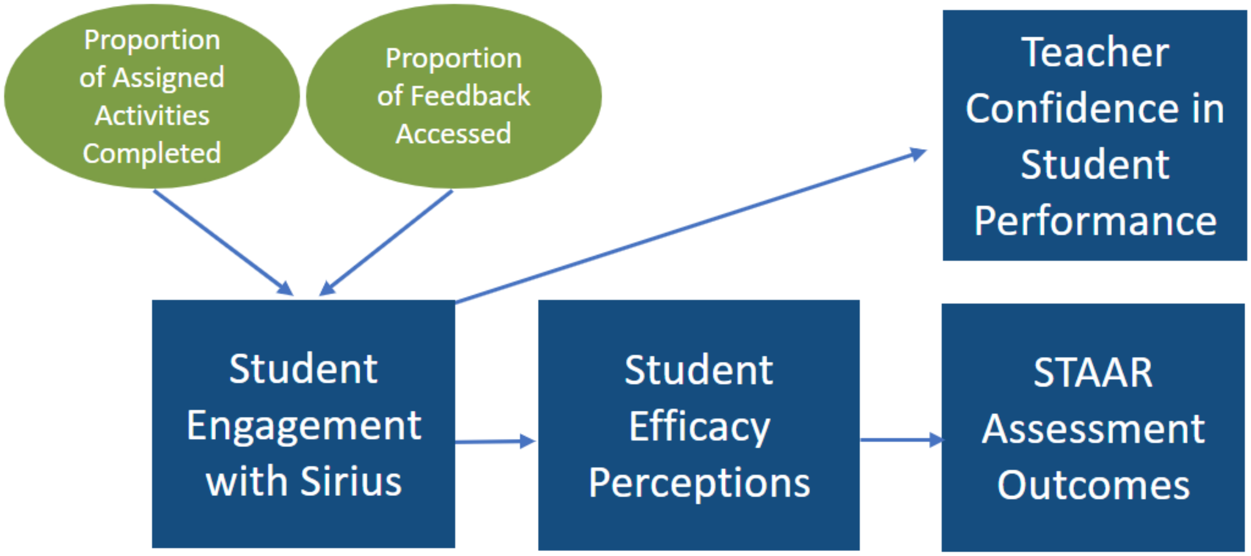
Analysis Plan

McREL will use a variety of analytic techniques to answer the research questions. First, McREL will use hierarchical linear modeling to answer evaluation questions one and two, given the nature of students nested within classes with varying implementation. Multiple linear regression or ANCOVA may also be used if a suitable sample size for the HLM is not collected. Figures 2 and 3 depict the planned analysis models for evaluation question one and two, respectively, and Figure 4 represents a planned model for question two if there is a large enough subsample of different levels of implementation among teachers to use as a moderator in the model. Evaluation questions three through five will be answered with descriptive statistics and thematic analysis (for open-ended questions).

Figure 2. Model to Answer Evaluation Question One

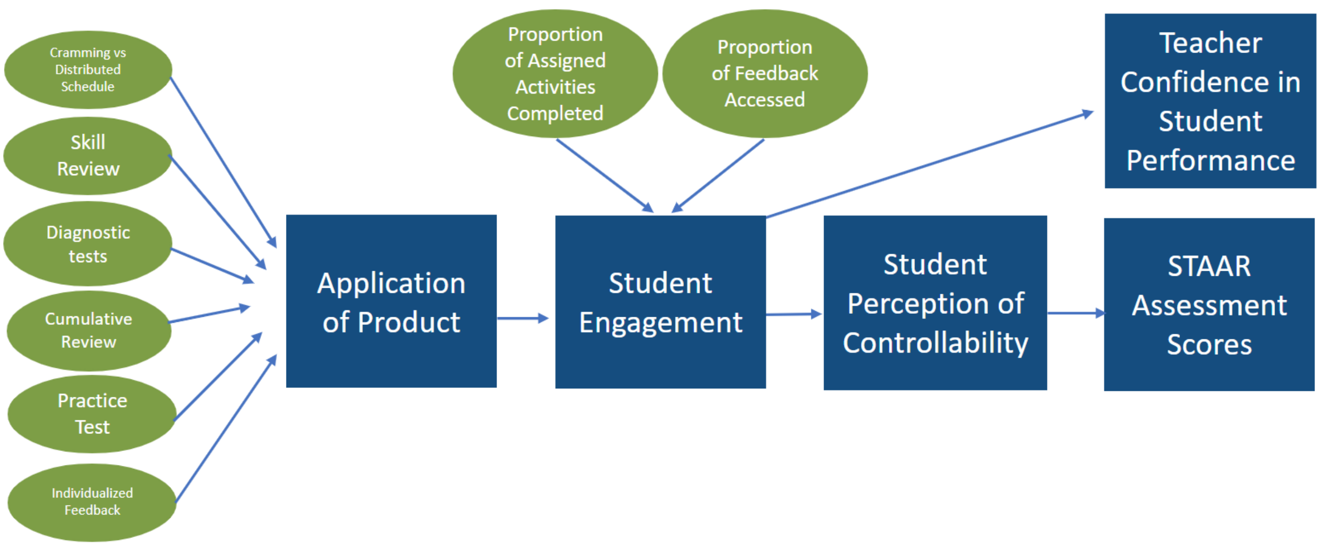
*The model will control for previous years' assessment scores, students' gender and socio-economic status.

Figure 3. Model to Answer Evaluation Question Two



*The model will control for previous years' assessment scores, students' gender and socio-economic status.

Figure 4. Evaluation Question Two Including Application of Product Moderation



*The model will control for previous years' assessment scores, students' gender and socio-economic status.

Generic Timeline

Table 3. Generic Timeline

	Phase 1 Preparation	Phase 2 Data Collection	Phase 3 Analysis & Reporting
Project Management			
Recruit Schools	X		
Process MOUs/IRB	X		
Instrument Development Using Preliminary Use Data			
Collect Early Sirius Use Data	X		
Finalize Teacher Survey	X		
Finalize Student Survey	X		
Data Collection and Analysis			
Administer Teacher Survey		X	
Administer Student Survey		X	
Collect Extant Student Data		X	X
Collect Final Sirius Data		X	
Data Cleaning and Analysis		X	X
Deliverables			
Final Report			X
District Data Summaries			X

References

Alvi, E., & Gillies, R. M. (2020). A case study of a grade 7 teacher's perspectives and practices related to self-regulated learning (SRL). *Asia-Pacific Journal of Teacher Education*, *48*(2), 147–167. https://doi.org/10.1080/1359866X.2018.1542663

Alvi, E., & Gillies, R. M. (2021). Investigating the individual nature of teacher beliefs and practices about self-regulated learning and how this shapes their practices for supporting student learning. *Educational Studies*, 1–23. https://doi.org/10.1080/03055698.2021.1966757

Andersen, S. C., & Nielsen, H. S. (2016). Reading intervention with a growth mindset approach improves children's skills. *Proceedings of the National Academy of Sciences*, *113*(43), 12111-12113. https://doi.org/10.1073/pnas.1607946113

Bai, B., & Wang, J. (2020). The role of growth mindset, self-efficacy and intrinsic value in self-regulated learning and English language learning achievements. *Language Teaching Research*, 1362168820933190. https://doi.org/10.1177/1362168820933190

Baars, M., Leopold, C., & Paas, F. (2018). Self-explaining steps in problem-solving tasks to improve self-regulation in secondary education. *Journal of Educational Psychology*, *110*(4), 578–595. https://doi.org/10.1037/edu0000223

Bloom, B. S. (1973). Recent developments in mastery learning. *Educational Psychologist*, *10*(2), 53-57. https://psycnet.apa.org/doi/10.1080/00461527309529091

Bostwick, K. C. P., Martin, A. J., Collie, R. J., & Durksen, T. L. (2019). Growth orientation predicts gains in middle and high school students' mathematics outcomes over time. *Contemporary Educational Psychology*, *58*, 213–227. https://doi.org/10.1016/j.cedpsych.2019.03.010

Choi, Y., & Kim, J. (2021). Learning Analytics for Diagnosing Cognitive Load in E-Learning Using Bayesian Network Analysis. *Sustainability, 13*(18), 10149. https://doi.org/10.3390/su131810149

Choi, H. H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review, 26*(2), 225-244. https://doi.org/10.1007/s10648-014-9262-6

De Smul, M., Heirweg, S., Devos, G., & Van Keer, H. (2019). School and teacher determinants underlying teachers' implementation of self-regulated learning in primary education. *Research Papers in Education*, *34*(6), 701–724. https://doi.org/10.1080/02671522.2018.1536888

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Dori, Y. J., Avargil, S., Kohen, Z., & Saar, L. (2018). Context-based learning and metacognitive prompts for enhancing scientific text comprehension. *International Journal of Science Education*, *40*(10), 1198–1220. https://doi.org/10.1080/09500693.2018.1470351

English, F. W., & Steffy, B. E. (2001). Deep curriculum alignment: *Creating a level playing field for all children on high-stakes tests of educational accountability*. Scarecrow Press.

Fernandez, O. E. (2021). Second chance grading: An equitable, meaningful, and easy-to-implement grading system that synergizes the research on testing for learning, mastery grading, and growth mindsets. *PRIMUS*, *31*(8), 855–868. https://doi.org/10.1080/10511970.2020.1772915

Harsy, A., Carlson, C., & Klamerus, L. (2021). An analysis of the impact of mastery-based testing in mathematics courses. *PRIMUS*, *31*(10), 1071–1088. https://doi.org/10.1080/10511970.2020.1809041

Hsu, Y.-S., & Lin, S.-S. (2017). Prompting students to make socioscientific decisions: Embedding metacognitive guidance in an e-learning environment. *International Journal of Science Education*, *39*(7), 964–979. https://doi.org/10.1080/09500693.2017.1312036

Kellogg, W. K. (2004). Logic model development guide. *Michigan: WK Kellogg Foundation*. https://wkkf.issuelab.org/resource/guiding-program-direction-with-logic-models.html

Lai, C.-L., & Hwang, G.-J. (2021). Strategies for enhancing self-regulation in e-learning: A review of selected journal publications from 2010 to 2020. *Interactive Learning Environments*, 1–23. https://doi.org/10.1080/10494820.2021.1943455

Lau, F. (2017). Methods for correlational studies. In F. Lau & C. Kuziemsky (Eds.), *Handbook of eHealth Evaluation: An Evidence-based Approach* [Internet]. University of Victoria. PMID: 29431951.

Leitzel, T. C., & Vogler, D. E. (1994). Curriculum Alignment: Theory to Practice. https://files.eric.ed.gov/fulltext/ED371812.pdf

Lenarz, J., & Pelatt, K. E. (2020). A transition to mastery-based testing with the hope of increasing student persistence. *PRIMUS*, 1–10. https://doi.org/10.1080/10511970.2020.1843209

Linhart, J. M. (2020). Mastery-based testing to promote learning: Experiences with discrete mathematics. *PRIMUS*, *30*(8–10), 1087–1109. https://doi.org/10.1080/10511970.2019.1695236

Maier, U. (2021). Self-referenced vs. reward-based feedback messages in online courses with formative mastery assessments: A randomized controlled trial in secondary classrooms. *Computers & Education*, 174. https://doi.org/10.1016/j.compedu.2021.104306

Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice.* Guilford Publications.

National Academies of Sciences, Engineering, and Medicine 2018. *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24783.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist, 38*(1), 1-4. https://doi.org/10.1207/S15326985EP3801_1

Patton, M.Q. (2012). *Essentials of Utilization-Focused Evaluation*. Sage Publications.

Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging mindsets to promote academic achievement. *Perspectives on Psychological Science, 10*(6), 721-726. https://doi.org/10.1177/1745691615599383

REL Midwest (2019). ESSA Tiers of Evidence: What You Need To Know. https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf

Rutherford, T., Buschkuehl, M., Jaeggi, S. M., & Farkas, G. (2018). Links between achievement, executive functions, and self-regulated learning. *Applied Cognitive Psychology*, *32*(6), 763–774. https://doi.org/10.1002/acp.3462

Slavin, R. E. (1990). Mastery learning re-reconsidered. *Review of Educational Research*, *60*(2), 300-302. https://doi.org/10.3102/00346543060002300

Sriram, R. (2014). Rethinking intelligence: The role of mindset in promoting success for academically high-risk students. *Journal of College Student Retention, 15*(4), 515-536. https://psycnet.apa.org/doi/10.2190/CS.15.4.c

Stufflebeam, D. L. (2000). The CIPP model for evaluation. In *Evaluation models* (pp. 279-317). Springer, Dordrecht.

Tachie, S. A. (2019). Meta-cognitive skills and strategies application: How this helps learners in mathematics problem-solving. *Eurasia Journal of Mathematics, Science & Technology Education*, *15*(5), 1–12. https://doi.org/10.29333/ejmste/105364

Tian, Y., Fang, Y., & Li, J. (2018). The effect of metacognitive knowledge on mathematics performance in self-regulated learning framework—Multiple mediation of self-efficacy and motivation. *Frontiers in Psychology*, 9, 1-11. https://doi.org/10.3389/fpsyg.2018.02518

Traynor, A., Li, T., & Zhou, S. (2020). Gauging uncertainty in test-to-curriculum alignment indices. *Applied Measurement in Education, 33*(2), 141–158. https://doi.org/10.1080/08957347.2020.1732387

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*(3), 212-218. https://psycnet.apa.org/doi/10.1016/j.cedpsych.2010.10.004

Vettori, G., Vezzani, C., Bigozzi, L., & Pinto, G. (2018). The mediating role of conceptions of learning in the relationship between metacognitive skills/strategies and academic outcomes among middle-school students. *Frontiers in Psychology*, 9, 1-13. https://doi.org/10.3389/fpsyg.2018.01985

Zee, M., & de Bree, E. (2017). Students' self-regulation and achievement in basic reading and math skills: The role of student–teacher relationships in middle childhood. European Journal of *Developmental Psychology*, *14*(3), 265–280. https://doi.org/10.1080/17405629.2016.1196587