U.S. Department of Education - EDCAPS
G5-Technical Review Form (New)

Status: Submitted
Last Updated: 09/04/2018 11:33 AM

Technical Review Coversheet

Applicant: Sonoma State University (U411C180146)
Reader #2: **********

		Points Possible	Points Scored
Questions			
Selection Criteria			
Quality of the Project Evaluation			
1. Project Evaluation		20	10
	Total	20	10

Technical Review Form

Panel #8 - Early Phase Tier 2 - 8: 84.411C

Reader #2: **********

Applicant: Sonoma State University (U411C180146)

Questions

Selection Criteria - Quality of the Project Evaluation

1. **In determining the quality of the project evaluation to be conducted, the Secretary considers the following factors:**

   (1) The extent to which the methods of evaluation will, if well implemented, produce evidence about the project's effectiveness that would meet the What Works Clearinghouse standards with or without reservations as described in the What Works Clearinghouse Handbook (as defined in the NIA).

   (2) The extent to which the evaluation will provide guidance about effective strategies suitable for replication or testing in other settings.

   (3) The extent to which the methods of evaluation will provide valid and reliable performance data on relevant outcomes.

   (4) The extent to which the evaluation plan clearly articulates the key project components, mediators, and outcomes, as well as a measurable threshold for acceptable implementation.

Strengths:

Sonoma State University proposes a multi-site randomized control design with treatment and control groups which is the strongest evaluation design meeting the WWC standards without reservations. A power analysis using Power Up software was conducted and appears to be generalizable. The sample size is large reaching 800 9th grade students over the five years of the grant.
A fidelity of implementation study is proposed where the applicant will conduct a series of usability tests on the treatment implementation providing data on the program as implemented at different sites. Additionally, student and teacher guides along with teacher slide sets will be produced to further support replication and testing in other settings. These activities are strategies that provide information and support for possible future replication and testing.
The applicant proposes to use the High School Student Attitudes toward STEM Survey, Teacher Instructional Practice and Competency Survey, Northeast Evaluation Association Measures of Academic Progress (MAP) for both math and science and the use of items from the Certica item bank to measure the effects of the implementation on both teachers and students. All of these instruments with the exception of the Certica item bank test questions have been tested for validity and reliability. Certica items may be valid and reliable; however, the applicant did not provide information on them. Students will be tested for academic gains in years 2 – 5 of the study.
Goals, objectives and outcomes are provided (Section B, Table 3). Acceptable measureable thresholds for outcomes are provided in this table. A 2 level HLM will be used to analyze the treatment effect.
Qualitative data collection methods providing important contextual information about the implementation include the use of classroom observations with protocols, teacher interviews, student focus groups and reviews of lesson plans.

Weaknesses:

During years 1 -3 teachers are provided with professional development and experiments with training guides and teacher slides are developed. The RCT begins in years four and five using a multi-site block trial design. Student academic achievement is expected to improve by 80% in years 2-5 as measured by the Certica science and math assessment items. Use of the MAP for math and science assessment is not included to determine student academic achievement. Explaining why Certica assessment items were used during years 2-5 and during the RCT to determine student achievement instead of or in addition to use of the MAP math and science assessments would have strengthened the proposal. How students and teachers would be assigned during the RCT was not provided. For example, it is unclear

whether the RCT is taking place at the student, classroom, or school level. A rationale for waiting until years 4 and 5 to implement the RCT would have been beneficial. Providing this information would have strengthened the application. Short, medium and long term outcomes were provided on the logic model, there was no discussion around how mediators and moderators would be identified. While the sample size is large, attrition is not discussed. For example, it is unclear how the study would handle data for students who move from school to school or drop out of the program.
Because of the lack of rationale for waiting until years 4 and 5 to implement the RCT, use of Certica assessment items to document student achievement instead of the MAP math and science assessments years 2 – 5 and lack of detail on how mediators and moderators will be determined, this evaluation most likely would not provide enough information on valid and reliable performance data on relevant outcomes.

Reader's Score: **10**

Status: Submitted
Last Updated: 09/04/2018 11:33 AM

Status: Submitted
Last Updated: 09/03/2018 12:10 AM

Technical Review Coversheet

Applicant: Sonoma State University (U411C180146)
Reader #1: **********

		Points Possible	Points Scored
Questions			
Selection Criteria			
Quality of the Project Evaluation			
1. Project Evaluation		20	12
	Total	20	12

Technical Review Form

Panel #8 - Early Phase Tier 2 - 8: 84.411C

Reader #1: **********

Applicant: Sonoma State University (U411C180146)

Questions

Selection Criteria - Quality of the Project Evaluation

1. **In determining the quality of the project evaluation to be conducted, the Secretary considers the following factors:**

   (1) The extent to which the methods of evaluation will, if well implemented, produce evidence about the project's effectiveness that would meet the What Works Clearinghouse standards with or without reservations as described in the What Works Clearinghouse Handbook (as defined in the NIA).

   (2) The extent to which the evaluation will provide guidance about effective strategies suitable for replication or testing in other settings.

   (3) The extent to which the methods of evaluation will provide valid and reliable performance data on relevant outcomes.

   (4) The extent to which the evaluation plan clearly articulates the key project components, mediators, and outcomes, as well as a measurable threshold for acceptable implementation.

Strengths:

1) The application proposes a multisite trial design where students will be randomly assigned to treatment or control groups. This approach qualifies the evaluation plan to meet the WWC standards without reservations. Furthermore, the inclusion of a statistical power analysis that factors in the large sample demonstrates the potential to produce strong evidence of program effectiveness.
2) The discussion on the extent to which the evaluation will provide valid and reliable data was thorough and explained the considerations that were factored into the validity and reliability analysis. Not only does the evaluation plan rely on already established instruments with proven validity and reliability in the field, but it also tests for validity and reliability using factor analysis. These steps ensure that performance data is thoroughly vetted and tested, which increases confidence the analysis.
3) In the project design, the proposal has built in the relationship between the key components, mediators, outcomes and thresholds for acceptable implementation. The threshold targets are clearly articulated and represent measurable targets that will aid in evaluating implementation fidelity as well as help with replication moving forward.

Weaknesses:

1) Even though the evaluation plan includes random assignment of students into treatment or control groups, there is no discussion of how baseline equivalence will be tested and/or adjusted. Without testing for baseline equivalence across treatment and control groups, it is unclear the extent to which these groups are statistically matched according to WWC standards, where the difference between them is no greater than .25 SD. Without a thorough understanding of the baseline equivalence between groups, any conclusions made about program impact may be affected by the inherent differences in the groups from before the treatment began.
2) One aspect that was not clear in the evaluation plan was the extent to which the other aspects of the proposal would be assessed for effectiveness. The random assignment of students in the experiment is designed to test the curriculum and instruction components on student achievement. However, it is unclear how the professional development and partnership components of the proposal will be evaluated for successful outcomes.
3) The discussion of replicability was not included, neither were mediators and moderators that might impact program effectiveness. It is not clear which confounding variable or other aspects of the program, such as site, attrition, or teacher characteristics might influence outcomes. It is also not clear how this study can be replicated in other settings as

no data was provided to guide other schools and districts.

Reader's Score: **12**

Status: Submitted
Last Updated: 09/03/2018 12:10 AM