Part B: Supporting Statement for Paperwork Reduction Act Submission

REL Pacific Efficacy and Implementation Evaluation of the Secondary Writing Toolkit

March 2023

Prepared for:
Elizabeth Nolan
U.S. Department of Education
550 12th Street, SW
Washington, DC 20024

Submitted by:
REL Pacific
745 Fort Street, Suite 1915
Honolulu, HI 96813

Table of Contents

B. Collection of Information Employing Statistical Methods

Introduction

The U.S. Department of Education (ED), through its Institute of Education Sciences (IES), requests clearance for the recruitment materials and data collection protocols under the OMB clearance agreement for activities related to the Regional Educational Laboratory Pacific Program (REL PA). Literacy, including writing, is closely tied to student success throughout K–12 education, which impacts high school graduation (National Institute for Literacy, 2008; NCES, 2020) and, ultimately, income beyond graduation (U.S. Bureau of Labor Statistics, 2019). Despite the importance of writing to life and learning, teachers report that the training they receive on teaching writing, both prior to entering the field and while teaching, is minimal or insufficient (Graham, 2019).

To address this problem, the REL PA Toolkit development team is developing a Secondary Writing Toolkit to support teachers in implementing evidence-based instructional strategies. The theory of change is that improving teachers' writing instruction will lead to improved writing among students in grades 6–8. The instructional strategies and resources in the Toolkit that will be selected and developed are based on the three recommendations in the *[Teaching Secondary Students to Write Effectively What Works Clearinghouse (WWC) Practice Guide](#)*: 1) explicitly teach appropriate writing strategies using a Model-Practice-Reflect instructional cycle; 2) integrate reading and writing to emphasize key writing features; and 3) use assessments of student writing to inform instruction and feedback. The Toolkit is being developed in collaboration with district and school partners in Hawaiʻi.

The REL PA Toolkit study team is requesting clearance to conduct an independent evaluation that will assess the efficacy of the Toolkit. The evaluation will also assess how teachers and facilitators implement the Toolkit to provide context for the efficacy findings and guidance to improve the Toolkit and its future use. The evaluation will take place in 40 schools in Hawaiʻi and focus on all students in grades 6–8.

B.1   Respondent Universe and Sample Design

The evaluation will employ a school-level cluster-randomized controlled design and take place in the state of Hawaiʻi, which is considered a single district. The pool of eligible schools includes schools that employ at least two English Language Arts (ELA) teachers in at least one of the target grades (grades 6–8). This criterion has been set by the Toolkit developer as the minimum number of teachers needed for effective implementation of the Toolkit activities. The study team will use Hawaiʻi Department of Education (HIDOE) school-level enrollment data from the 2022/23 school year to identify schools that meet this criterion and expects that there will be approximately 110 schools in this group. The study team will approach all eligible schools to confirm eligibility and ascertain their interest in participating in the study. Schools will need to meet three additional criteria to be included in the sample: 1) the school principal will need to agree to participate; 2) at least 80 percent of the ELA teachers in grades 6–8 will need to agree to participate; and 3) the school must have a staff member who can serve as a peer facilitator for the Toolkit Professional Learning Community (PLC)—the study team anticipates this will be an ELA teacher in most cases, but it could also be a coach or teacher leader, depending on the school. The first criterion is to ensure buy-in at the level of school leadership; the second is necessary to maximize the impact estimate of access to the Toolkit components and professional development activities, since all students in a school, regardless of whether their teacher accesses the Toolkit components and professional development activities or not, will be included in the school-level means used in the intent-to-treat impact analysis; and the third is necessary for effective implementation of the

Toolkit professional development activities. Within schools, the sample will include all grades 6–8 ELA teachers (20 of these teachers will likely be peer facilitators), their students, and school administrators (principals or assistant principals).

Exhibit B2 shows the sample sizes and expected response rates for each level of data collection.

Exhibit B2. Sample Sizes and Expected Response Rates

Level of Sample	Expected Sample Size	Response Rate
District	1	100%
Schools	40	100%
School Administrators	40	85%
Teachers[a]		
Peer Facilitators	20	90%
Other Participating Teachers	180	85%
Student School-level Averages (SBA Writing)	40	100%

[a] Based on current data from the Hawaiʻi Department of Education, the study team estimates an average of five ELA teachers per school.

B.2   Information Collection Procedures

B.2.1   Notification of the Sample and Recruitment

To support the study team's recruitment efforts, HIDOE (the "school district") has agreed to endorse the study and will provide a letter of support to document approval from HIDOE's Office of Curriculum and Instruction, in addition to support from the HIDOE State Superintendent. This will likely be accomplished prior to August 2023. Following HIDOE's research policy, the study team will make initial contact with the 15 superintendents of the 110 schools via email, with a phone call follow-up shortly thereafter (within two days of the email), to apprise them of the opportunity to participate in the study, gauge interest, and obtain buy-in and approval to reach out to school administrators. With superintendent support, we will target all eligible schools with two or more teachers of grades 6–8 on any island. The study team will leverage connections of on-island REL Pacific staff and study liaisons to reach out via email and phone calls to school administrators of the 110 eligible schools. Schools will need to meet two criteria to be included in the sample: 1) the school needs to agree to participate; and 2) at least 80 percent of the ELA teachers in grades 6–8 will need to agree to participate. The study team, in close coordination with the study liaisons, will work with school administrators to obtain consent from eligible teachers and to execute a Memorandum of Understanding between REL Pacific and each school to clarify with schools what participation entails and provide an outline of activities and timing. After the school administrator expresses interest and agrees to participate, the study team will ask school administrators about the best approach to continuing the conversation with eligible teachers (as a condition of participation, school administrators will have already apprised teachers of the opportunity) and will be flexible in accommodating each school's suggestions on how to present the study to teachers, which we believe may include REL Pacific staff emailing teachers directly, attending a school staff meeting (virtually or in person), and/or hosting an informational webinar. Recruitment materials that will be used for this study are included in Appendix A.

Primary data collection will occur in schools that consent to be part of the study. Each round of data collection will include an initial email outreach, multiple follow-up emails, and phone calls for non-respondents. Phone calls will be done by study liaisons, who will have relationships with the participating peer facilitators, teachers, and administrators in the schools for which they are responsible. Data

collection communication email texts and follow-up phone call talking points are included in Appendix B.

B.2.2 Statistical Methodology for Stratification and Sample Selection

The Toolkit's professional learning activities are intended to be a school-wide approach to improving teachers' writing instruction and students' writing achievement and are delivered through a professional learning community among all grade 6–8 ELA teachers, rather than individually. Therefore, the study team proposes using the school as the unit of assignment. Methodologically, school-level random assignment reduces the risk of within school spillover as all teachers within a school are implementing the treatment or not. Random assignment will take place in May of 2024—immediately after recruitment closes and prior to the Toolkit implementation year—to allow schools to select a peer facilitator and prepare for participation in the treatment. Plans for random assignment will be communicated with school officials early in the recruitment process to ensure buy-in and randomization assignments will be carefully documented.

Prior to random assignment, the study team will generate randomization blocks to increase the explained variance at the school level and increase power. Randomization blocks will include school characteristics based on publicly available data, including the log of student membership counts, the natural log of the ratio of minority to white students,[1] student–teacher ratios, log proficiency rates on the Spring 2024 Smarter Balanced Assessment (SBA) Writing score, and median income of the school catchment area (the geographical boundaries of typical school assignment for families). The study team will use k-means clustering to maximize the multivariate between-block differences, minimize the multivariate within-block differences, and create the optimal number of randomization blocks. The goal will be to create randomization blocks with more than two schools since withdrawal of a school from a block with only two schools would necessitate the removal of the other pair member from the analysis and affect attrition. Within each randomization block, the study team will employ a sampling function (such as the sample function in R) without replacement to randomly assign half of the schools to treatment and half to comparison group status.

B.2.3 Estimation Procedures

The study team will use school averages of students' SBA Writing scores, instead of individual student-level data, which produce the same impact estimate as if all student data were available, while greatly reducing burden on schools, parents, and students, as parents and students do not have to individually, actively consent to the use of these school averages. Using school averages also improves the generalization of the impact estimate, as it precludes bias due to compliance non-response, while maintaining a low risk of privacy violations. The estimation strategy for the student impact analysis is a simple multivariate linear model (RQ E1). The analysis of teacher outcomes based on teacher logs and the administrator survey (questions on teachers' receipt of professional development and coaching at the school level) will utilize mixed (or hierarchal) generalized linear models (RQ E2, RQ I3). Because the assignment to treatment or comparison is at the school level, the use of mixed models is required to estimate the appropriate standard errors of the impact estimate of the teacher-level dependent variables. The study team will use these mixed generalized models to appropriately model the outcome distributions for both categorical variables and positive integers. All impact estimates will represent the differences

---

[1] Ratios are not normally distributed and sensitive to extreme values. By taking the natural log of such ratios, the k-means clustering procedure assumptions are better met.

between sample schools randomly assigned to treatment and those assigned to comparison, also known as the intent-to-treat (ITT) impact estimate.

Student Outcome. To address RQ E1 about student outcomes, the study team will use an Ordinary Least Squares regression model to estimate the average ITT impact of the treatment on the school-level averages of students' SBA Writing scores. The model for outcome y of school j in regression notation is:

$$\overline{Z}_{.j} = \beta_0 + \beta_1 T_j + \sum_c^{\square} \gamma_c \overline{x}_{c.j} + \sum_q^{\square} \delta_{qj} + u_j,$$

where $\overline{Z}_{.j}$ is the weighed school average of the teacher and grade-specific SBA Writing scores standardized on grade-specific pooled standard deviations; β_1 is estimated impact of the treatment assignment, T_j (coded 1 for treatment assignment and 0 for comparison), on the school average SBA Writing score; γ_c are associations of school means of demographic covariates (median income for the school catchment area and proportion of students who are members of minority groups) and school-specific averages of the SBA Writing scores from the previous year, noted collectively as $\overline{x}_{c.j}$; $\sum_q^{\square} \delta_{qj}$ are the randomization block fixed effects;[2] and u_j are school residuals on the outcome. The model will be fit with standard regression procedures and tested against the null hypothesis with classical statistical methods.

The study team will create the school-level student outcome variable as a weighted mean of the teacher and grade-specific averages of the SBA Writing score, standardized on the grade-specific sample means and standard deviations for each teacher. Each teacher for each grade of each school will be summarized by three values: the total number of students tested by teacher, the average SBA Writing scale score for each teacher, and the standard deviation of the SBA Writing scale score for each teacher. For the school means analysis, the school mean will be a weighted average of each teacher's average within the school, standardized by the grade-specific pooled standard deviation. This z-score, $\overline{Z}_{.j}$, will be a standardized measure of how each school's mean deviates from the sample's mean.

Teacher Outcomes. To address RQ E2 about impacts on teacher practice,[3] the study team will convert the data from the teacher log into aggregate teacher outcome variables that summarize the totals, counts, numbers, and Yes/No dichotomous responses, as shown in Exhibit B3 below. These outcome variables correspond to the recommendations in the Practice Guide (PG).

Exhibit B3. Writing Instruction Outcomes and Metrics from Teacher Log Data

Writing instruction outcome	Metric
Number of minutes of time spent on writing instruction and writing activities.	Total minutes
Number of different types of writing instruction/activities implemented.	Count
Any instruction or feedback that was informed by assessments of student writing.	Yes/No
Any instruction that integrated writing and reading.	Yes/No

---

[2]   Should the number of blocks be large, such as each block is only a pair of schools, the study team will employ alternative methods to explain this variance, such as de-meaning or adding the blocks as a higher random effect level to increase sensitivity.

[3]   Data from the teacher log serves two purposes: 1) it illustrates the effect of the Toolkit on teachers' self-reported practice (RQ E2); and 2) it provides information on the treatment-control contrast in classroom instruction, which is a requirement of the implementation evaluation (RQ I3, sub-question 2).

Length of longest writing piece that students were assigned/worked on.	Number
Any explicit instruction of *planning* strategies.[a]	Yes/No
Number of different *planning* strategies explicitly taught.[a]	Number
Any modeling of *planning* strategies.[a]	Yes/No
Any opportunities for students to practice *planning* strategies in their own writing.[a]	Yes/No
Any opportunities for students to evaluate or reflect on their use of *planning* strategies in their own or their peers' writing.[a]	Yes/No
Any instruction that follows the full writing process (teach, model, practice, and reflect) within each of the six components of writing.	Yes/No
[a] These five outcomes are repeated for the other five components of the writing process: *goal setting*, *drafting*, *evaluating, revising,* and *editing*.	

These outcomes will be analyzed separately using generalized mixed models, where the dependent variable is "transformed" into a linear metric during estimation. For example, totals, numbers, and counts will be analyzed using a Poisson distribution with a log-link function (as negative and non-integer values are not possible), and dichotomous indicators will be analyzed using a binomial distribution with a logistic function. The model will be a mixed generalized linear model, as noted by the $g(.)$ link function and the lack of a teacher-level residual. The model for outcome y of the ith teacher in school j in generalized mixed notation is:

$$g\left(y_{ij}\right)=\beta_0+\beta_1 T_j+\sum_c \lambda_c x_{cij}+\sum_c \gamma_c \overline{x}_{c.j}+\sum_q \delta_{qj}+r_j,$$

where β_1 is the estimated impact of the treatment assignment, T_j (coded 1 for treatment assignment and 0 for comparison). Depending on the link function, the natural exponent (exp) of this regression slope will have different interpretations. For Poisson models of totals, numbers, and counts, $\exp\left(\beta_1\right)$ will be the ratio of the average value of treatment teachers to comparison teachers. For dichotomous outcomes, $\exp\left(\beta_1\right)$ will represent the odds ratio of the outcome for treatment teachers relative to comparison teachers. Other terms in this model include λ_c, which are associations of teacher covariates (such as the baseline measure, gender, race/ethnicity, years of teaching experience, and educational credentials), noted as x_{cij}; γ_c are associations of school means of demographic covariates (proportion of students eligible for free or reduced price lunch and proportion of students who are members of minority groups) and school-specific averages of the SBA Writing scores from the previous year, noted collectively as $\overline{x}_{c.j}$; $\sum_q \delta_{qj}$ are the randomization block fixed effects; and r_j is the random effect at the school level. Note that generalized outcomes do not produce residual variation, so this model omits a teacher-level residual.

Sensitivity Analyses. We assume most peer facilitators will be classroom teachers who will implement the evidence-based strategies in their classrooms and therefore plan to include them in our impact analyses. However, because peer facilitators have a unique role in implementing the intervention, have received additional training, and have no similar counterparts in the counterfactual condition (since there may not be PLCs in comparison schools), we will also conduct sensitivity analyses, in which we exclude the peer facilitators (n = 20) from the treatment group to determine if their outcomes affect the results for the full sample. To do this, we will re-estimate the impact estimate, as described above, using school averages but omit the data from the peer facilitators to compute school averages for the sensitivity analysis. To maintain comparability with the primary student outcome analysis, we will use the same pooled standard

deviations to standardize the school means. Sensitivity analyses for teacher outcomes will also exclude peer facilitators but use the same procedures described above for estimating teacher outcome differences. If results differ when peer facilitators are omitted, we will present results for both the full sample and the sample that excludes peer facilitators.

***Strategies for Correcting for Multiple Hypothesis Testing*.** According to the [What Works Clearinghouse Study Review Protocol, version 1.0](#) (U.S. Department of Education, 2020), the proposed analysis does not require correction for multiple hypothesis testing, as it includes only one confirmatory analysis comparison within the Writing Quality Domain (treatment vs. comparison outcome for all students).

Implementation Fidelity. To assess fidelity at the school level (RQ I2), the study team will calculate quantitative scores for fidelity of implementation of the Toolkit by comparing information on the implementation as-conducted (actual implementation in Hawaiʻi) compared to the implementation as-planned, which is represented by the logic model the study team created based on information from the Toolkit developer. Using data collected from peer facilitators, teachers, and the Toolkit developer, the study team will calculate fidelity of implementation at the school level in two ways: 1) separately for the six components of the Toolkit activities (peer facilitator training, administrator training, peer facilitator preparation work before eight PLC sessions, peer facilitator leadership of activities across eight PLC sessions, teacher independent work before eight PLC sessions, and teacher participation activities in eight PLC sessions); and 2) across the six components. For each component and overall, the study team will score school-level fidelity as high, moderate, or low (Exhibit B4). The study team will also determine the level of fidelity of implementation of the Toolkit overall (at the program level) by reporting the overall percentage of schools with high, moderate, or low fidelity. This approach will allow the study team to determine which of the Toolkit activities were met with fidelity or may require additional supports to ensure fidelity in a future implementation. These data will also provide valuable contextual information for the impact evaluation and may help explain variation in impacts.

Exhibit B4. Fidelity of Implementation Score Matrix to be Calculated for Each Treatment School (across four modules)

Level of fidelity for school	Component 1: Peer facilitator training	Component 2: Administrator training	Component 3: Peer facilitator preparation work before 8 PLC sessions	Component 4: Peer facilitator leadership of activities across 8 PLC sessions	Component 5: Teacher independent work before 8 PLC sessions	Component 6: Teacher participation in activities in 8 PLC sessions	Across All Components
High	85–100% of total possible points	85–100% of total possible points	85–100% of total possible points	85–100% of total possible points	85–100% of total possible points	85–100% of total possible points	"High fidelity" for all components
Moderate	50–84% of total possible points	50–84% of total possible points	50–84% of total possible points	50–84% of total possible points	50–84% of total possible points	50–84% of total possible points	"High fidelity" for components 3–6, but not 1 and 2
Low	< 50% of total possible points	< 50% of total possible points	< 50% of total possible points	< 50% of total possible points	< 50% of total possible points	< 50% of total possible points	"High fidelity" for only some of components 3–6

Implementation Experiences, Challenges, and Recommendations for Improvements to the Toolkit. To describe the experiences of peer facilitators and teachers using the Toolkit, identify challenges for completing Toolkit activities, and summarize recommendations for improvements in future implementations (RQ I1, sub-questions 1, 2, & 3), the study team will collect and analyze quantitative data from the Professional Learning Tracker (PLT) and qualitative data from teacher and peer facilitator focus groups. The PLT asks teachers Yes/No questions about what Toolkit resources they used and activities they participated in during the PLCs and in their independent work in their classrooms, such as "During today's PLC session, did you use a protocol for reflecting on and discussing the video?" and "Before today's PLC session, did you read relevant sections from the What Works Clearinghouse Practice Guide?" It also asks teachers to rate how helpful the resources and activities were for them on a Likert scale. The study will conduct descriptive analyses of these responses. The focus groups ask teachers more open-ended questions about their experiences in using the Toolkit resources and activities, the challenges they encountered, and recommendations for improving the Toolkit, such as "How challenging was it for you to take on the PLC activities in addition to your regular classroom and school/district responsibilities?" and "What additional supports might have been helpful in making participation in the PLC more manageable?" The study team will apply thematic analysis (Guest, MacQueen, & Namey, 2012) to analyze the focus group responses in NVivo 12.0. Through this process, the study team will generate a set of themes that emerge in the data and develop summaries for the most commonly occurring challenges and recommendations for improvement. The study team will use the results of the qualitative analysis to supplement the quantitative PLT data.

Treatment Comparison Contrast. To assess contrasts between treatment and comparison group teachers (RQ I3), the study team will conduct analyses of two key aspects of implementation: 1) teachers' receipt of professional development and coaching on writing; and 2) the use of best practices in writing instruction in the classroom. For the first aspect, the study team will examine the contrast in the professional development and coaching on writing received by treatment and comparison teachers (above and beyond the Toolkit for treatment schools) by administering a survey to administrators in both treatment and comparison schools about the types of professional development and coaching on writing that their teachers have received during the 2024/25 school year. These data will represent professional activities at the school level (for treatment schools, the study team will also include the Toolkit professional development and coaching). Because teachers may have also individually pursued professional development and/or coaching outside of their school, the second teacher log (in February 2025) asks treatment and comparison teachers about any professional support on writing they have received (above and beyond the Toolkit for treatment teachers) during the 2024/25 school year. The study team will combine these two data sources (school and individual teacher-level reports of receipt of professional development and coaching on writing) for each teacher to create a new variable to use in the treatment comparison contrast analysis. For the second aspect, the study team will examine the contrast in the writing instruction implemented by treatment and comparison teachers using data from the teacher log, which asks teachers about their implementation of best practices in writing instruction. The analysis approach for both these aspects of implementation is explained above under Teacher Outcomes. The data collected from the administrator survey and the teacher log will also provide insights into what "business as usual" looks like in the counterfactual.

B.2.4   Degree of Accuracy Needed

The study team used school-level enrollment data from HIDOE for the 2021/22 school year to identify the schools that are eligible for the recruitment pool (schools that employ at least two ELA teachers in at least one of the target grades of 6–8). The student impact will be estimated as the difference in the school

means of the SBA Writing score between the treatment and comparison schools. As such, the power analysis is based on a simple independent t-test comparing 20 treatment schools and 20 comparison schools using formulas found in Hedges & Rhoads (2010). The minimum detectable difference, based on including a covariate measuring each school's average Writing score from the previous year and randomization blocks, is expected to explain 80 percent of the residual variance (see Hedges & Hedberg [2013] Table 8, which indicates .80 is a typical value for within-district school-level reading score pretest R-squares; to our knowledge, design parameters for writing are not available in the literature), which leads to an effective minimum detectable effect size (MDES) of 0.41 ($0.91 \times \sqrt{1-.8} = 0.41$) in school-level standard deviation units. To translate this value into a student-level effect size metric, a school-level intraclass correlation coefficient (ICC) of 0.15 for the SBA Writing score outcome was estimated by using available data from HIDOE on SBA ELA proficiency for the 110 eligible schools.[4] Using this school-level ICC of 0.15, the school-level effect size of 0.41 translates into a student-level effect size of 0.17 ($0.41 \times \sqrt{.15} = 0.17$).

The teacher impact will be estimated as the differences in teacher log variables between the treatment and comparison teachers. The expectation is that a pool of 40 recruited schools will translate into 200 teachers in the sample, which will result in a MDES of 0.36 for the teacher outcomes. This MDES will allow the study team to detect the likely size of effects on teacher outcome since it is likely that the differences between treatment and comparison teachers on writing instruction will be large. The study team bases this expectation on the assumption that comparison group means are likely to be very low relative to the treatment group and have low overall standard deviations, given that teachers are not often provided effective professional development in writing instruction (Graham, 2019) and that HIDOE has indicated that teachers in Hawai'i receive very little, if any, professional development in writing. Moreover, this MDES is consistent with previous studies on teacher writing practice (Gallagher et al., 2017). Based on the parameters for the ICC and R-square values in Exhibit B5, which are consistent with Westine et al. (2020) and Kelcey & Phelps (2014) for other professional development (PD) teacher outcomes related to science and math instruction, the study team estimates a maximum MDES of 0.36 standard deviation (SD) for power of 0.80 for a two-tailed test at α = 0.05. The teacher log on writing instructional practices will be administered in both fall 2024 and spring 2025, so there will be a natural pretest for this outcome.

Exhibit B5. Power Parameters for Teacher Outcome Analyses

Minimum Detectable Effect Size	Samples at Each Level		School Intraclass Correlation	Covariate R-Squares	
	School	Teacher		School	Teacher
0.36	40	5	.1	0.3	0.5
0.35		6			

---

[4] While it would be better to use the ICC for the SBA Writing score, this value was not available. Therefore, the study team computed this value on the next most correlated outcome, ELA scores, using the following procedure. After downloading the 2020–21 state report card data, the study team filtered the data to the 110 schools and the rows for "all students." The percent proficient in ELA metric represents the probability of being proficient for each school, multiplied by 100. The study team then converted these percentages into log-odds metrics, with $\Lambda = \ln(\%/(100-\%))$, and found the variance of Λ across the 110 schools, $V[\Lambda]$. This approximates the school-level variance component (variances in means) of a logistic regression predicting the chance of a student being proficient. The intraclass correlation for such an outcome employs the variance of the logistic function, $\sigma^2 = \dfrac{\pi^2}{3}$, which is then $ICC = \dfrac{V[\Lambda]}{V[¿\Lambda]+\sigma^2}$.

0.33		7			

B.2.5 **Unusual Problems Requiring Specialized Sampling Procedures**

There are no unusual problems requiring specialized sampling procedures.

B.2.6 **Use of Periodic (less than annual) Data Collection to Reduce Burden**

This evaluation will collect data one time for recruitment. Similarly, student outcome and background data will only be requested from HIDOE once per year (July 2024 and September 2025). Administrators will only be surveyed once, and peer facilitator and teacher focus groups will only happen once (March/April 2025). However, teacher data will need to be collected more frequently than annually because the evaluation is occurring within one school year, and some measures will need to be assessed pre- and post-treatment in September 2024 and February 2025 (teacher log) or after each PLC (about twice per month for four–five months using the PLT). A longer period between data collection for the teacher log would make it difficult for the study team to meet the requirements for the efficacy study (by preventing baseline and follow-up data collection in the timeframe necessary for the evaluation). A longer period between data collections for the PLT would mean that the evaluation would not be able to gather timely data on implementation and meet the requirements to conduct an implementation evaluation.

B.3   Methods for Maximizing the Response Rate

The study team is committed to obtaining complete and high-quality data for this evaluation. The primary efficacy evaluation question on student writing relies entirely on administrative data from one district—HIDOE—on student SBA Writing scores averaged at the school level, and student and school characteristics. Given the use of school-level averages from administrative data, the study team does not expect to have any missing student writing data at the school level. While not every student enrolled during the full academic year is likely to complete the assessment at each school, schools are legislatively mandated to ensure that 95 percent of students are tested, which poses a low risk of bias in the derived administrative averages of schools. Differential bias between treatment and comparison schools is also not likely to be an issue. The study team has already been in contact with HIDOE and will carefully check all data received for completeness. Follow-up emails and calls will be made to ensure that all data needed for the evaluation are received in a timely manner.

For all other data collection activities, the study team will use multiple strategies to ensure high response rates and high-quality data and to protect respondents' privacy, regardless of data type. As noted in Exhibit B2 above, the study team expects a 90 percent response rate for peer facilitators (who are also teachers) and 85 percent response rate for other participating teachers and administrators. The study team will: 1) use well-designed and respondent-tailored instruments; 2) develop strong partnerships at both the district and school levels and obtain buy-in at the district and school levels to encourage participation of teachers and students and coordinate scheduling of study data collection with school and testing schedules; 3) use multiple approaches to engage participants in the data collection, including reminder emails and telephone calls to ensure high response rates; 4) offer incentives for completion ($10/day to teachers and peer facilitators for the log for 10 days, $50 for the PLT, and $30 for the focus group; $15 to

administrators for the survey; and $2,500 to comparison schools);[5] and 5) employ a comprehensive tracking system to ensure follow-up with non-response during data collection in a timely manner using intentionally varied strategies.

The study team will hire temporary data collection staff in Hawaiʻi who will serve as study liaisons to the schools to help with data collection. With 40 schools, the study team assumes that approximately 8–10 field staff are needed, each of whom will cover multiple schools. The study liaisons' tasks will be to cover all aspects of data collection, including alerting teachers about missing logs daily during the data collection period, following up with peer facilitators and teachers about PLT completion, recruiting participants for focus groups, and being available for general data collection questions. The study team has used this approach in the past, and it has resulted in efficient and secure data collection and high response rates.

In addition to the activities focused on maximizing response rates, the study team will use What Works Clearinghouse-accepted techniques to deal with missing data due to non-response (for example, creating non-response and analysis weights, imputation). The study team will also consider the most recent statistical literature to examine other additional methods. If such methods are necessary, results using data not adjusted for missingness will also be included in an appendix for the report.

B.4   Test of Procedures

The student outcome measure used to analyze the impact of the Toolkit is a standardized state assessment that has demonstrated validity and reliability and will not require pretesting. The teacher measures for impact and implementation, including the teacher log, the PLT, and the focus group protocols, are new instruments. For the teacher log, which will be used to measure the impact of the Toolkit on teacher practice, the study team will conduct a pilot study in Spring 2023 with six teachers (two per grade level). The objective of the pilot study will be to validate teachers' report of instructional practices on the log against observer report. The study team will train two observers on the log and have them observe in the teachers' classrooms during ELA instruction for one week. The study team will also conduct a training for the pilot teachers on how to complete the log and then compare teacher to observer reports to check for consistency. The study team will also conduct cognitive interviews with the teachers and observers to obtain feedback on how they understood and interpreted the questions in the log, which will identify any issues related to question wording, ordering, and format that need to be addressed in preparing the final version of the log. For the remaining teacher measures (PLT and focus groups) and the administrator survey, all of which are used to measure implementation, the study team will conduct at least two, but not more than nine, cognitive interviews with teachers in the schools that are participating in the Toolkit usability testing during the 2022/23 school year. The instruments are included in Appendix C.

B.5   Methods to Minimizing Burden on Small Entities

The individuals consulted on the statistical aspects of the design include:

Paul J. Burkander, senior education researcher, SRI Education, (650) 859-2000, paul.burkander@sri.com

Christina Tydeman, executive program director, Regional Educational Laboratory of the Pacific at McREL International, (303) 337-0990, ctydeman@mcrel.org

---

[5]   This incentive amount is necessary to encourage participation in the study and has been approved by IES.

References

Gallagher, A. H., Arshan, N., & Woodworth, K. (2017). Impact of the National Writing Project's College-Ready Writers Program in high-need rural districts. *Journal of Research on Educational Effectiveness, 10*(3), 570–595. https://doi.org/10.1080/19345747.2017.1300361

Graham, S. (2019). Changing how writing is taught. *Review of Research in Education, 43*(1), 277–303. https://doi.org/10.3102%2F0091732X18821125

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). Applied thematic analysis. SAGE Publications, Inc. https://www.doi.org/10.4135/9781483384436

Hedges, L. V., & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning two- and three-level cluster-randomized experiments in education. *Evaluation Review, 37*(6), 445–489.

Hedges, L. V., & Rhoads, C. (2010). *Statistical power analysis in education research* (NCSER 2010-3006). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf.

Institute of Education Sciences. (2021). *Study review protocol, version 1.0*. Washington, D.C.: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Document/1297

Kelcey, B., & Phelps, G. (2014). Strategies for improving power in school-randomized studies of professional development. *Evaluation Review, 37*(6), 520–554. https://doi.org/10.1177%2F0193841X14528906

National Center for Education Statistics (NCES). (2020, September 21). National Assessment of Educational Progress (NAEP): Writing. nces.ed.gov/nationsreportcard/writing/

National Institute for Literacy. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. National Center for Family Literacy. lincs.ed.gov/publications/pdf/NELPReport09.pdf.

U.S. Bureau of Labor Statistics. (2019). *Median weekly earnings $606 for high school dropouts, $1,559 for advanced degree holders*. U.S. Department of Labor. https://www.bls.gov/opub/ted/2019/median-weekly-earnings-606-for-high-school-dropouts-1559-for-advanced-degree-holders.htm

Westine, C. D., Unlu, F., Taylor, J., Spybrook, J., Zhang, Q., & Anderson, B. (2020). Design parameter values for impact evaluations of science and mathematics interventions involving teacher outcomes. *Journal of Research on Educational Effectiveness, 13*(4), 816–839. https://doi.org/10.1080/19345747.2020.1821849