REL Peer Review: Pilot Data Collection Methods for Examining the Use of Research Evidence

Part B: Collection of Information Employing Statistical Methods

November 6, 2023

REL Peer Review: Pilot Data Collection Methods for Examining the Use of Research Evidence

Part B: Collection of Information Employing Statistical Methods

November 6, 2023

CONTENTS

APPENDIX A: REL USE OF RESEARCH EVIDENCE (URE) SURVEY

APPENDIX B: FOLLOW-UP INTERVIEW PROTOCOL

APPENDIX C: OUTREACH MATERIALS

Exhibits

B. Collection of information employing statistical methods

Introduction

The Institute of Education Sciences (IES) within the U.S. Department of Education (ED) requests clearance for data collection activities to support a pilot study of the reliability and validity of survey items used to assess the use of research evidence (URE) among education agencies and other partners served by the Regional Educational Laboratories (RELs). The REL program is an essential IES investment focused on partnering with state and local education agencies use evidence to improve education outcomes by creating tangible research products and providing engaging learning experiences and consultation. IES seeks to better understand how REL partners use research evidence to improve education outcomes and the role of RELs in promoting URE among partners.

This study will test the reliability and validity of new and extant URE items in the REL context. Specifically, the study will (1) assess how existing items from the URE literature perform in a REL context and (2) assess the reliability and validity of a small set of items from the Stakeholder Feedback Surveys (SFS) that are currently administered to REL partners and used by IES to improve the work of REL contractors, inform the REL program as a whole, and address internal requests such as the Congressional Budget Justification. The reliability and validity of the new and existing survey items will be assessed through two data collection activities: an online survey administered to a set of partnerships across RELs and follow-up interviews with a subset of REL partners.

IES contracted with Mathematica to conduct this study. At the end of the study, the study team will provide a finalized list of valid and reliable survey items that RELs can use in future iterations of the SFS. The current SFS is already approved by OMB.

B1. Respondent universe and sampling methods

For each data collection activity proposed, Exhibit B.1 summarizes the respondent universe and expected response rate. These are also described in more detail following the exhibit.

Exhibit B.1. Respondent universe, sampling method, and response rate by data source

Data source	Respondent	Respondent sample	Expected response rate
Web survey	REL partners	85	85%
Follow-up interview	REL partners	30	60%

Study sample. The respondent universe for the web survey and follow-up interviews consists of REL partners across 10 REL regions. Each REL contains several partnerships that are each designed, developed, and executed to improve long-term student success on a focused, high-leverage topic within a specific state or district. Each of the 10 RELs has at least one partnership serving each of the states and territories in their region; as a whole the REL program currently supports 62 partnerships nationwide. Each partnership consists of a set of REL partners from schools, districts, state agencies, and other organizations, including a total of 450 core partners and more than 2,500 non-core partners across the full REL program.

1. **Core partners** are people involved in planning the overall direction of the partnership with REL staff to address high-leverage needs. These people may or may not attend REL activities but regularly attend planning meetings for projects and activities under the partnership.

2. **Non-core partners** are people not highly involved in the overall planning, direction, or strategy of the work of the partnership. However, they receive REL services, such and training and coaching, and contribute their expertise to applied research and dissemination projects. They may also attend planning meetings for projects and activities under the partnership.

The study team will select a set of partnerships across REL regions to participate in the study and invite all members of those partnerships to participate in the web survey; a smaller subset of partners will be invited to participate in the follow-up interviews. The team will work with IES and RELs to ensure the sample represents variation across partner categories, roles, and locations (for example, core/non-core partners, different education roles, multiple RELs and regions of the country.)

Web survey. The study team will select a sample of partnerships for the survey that reflects a range of REL partners, including both core and non-core partners in different education-related roles, such as teachers, administrators, and staff at the district, post-secondary, and state level.

The study team expects the best candidates for participating in the survey will be REL partnerships with activities ending in spring 2024. The study team will determine which partnerships among those with an activity ending in spring 2024 best represent a cross-section of the various types of REL partners, including core and non-core partners and respondents at the school, district, and state levels. The team will then work with their Contracting Officer's Representative – the Branch Chief for the REL Program to gather names and email addresses of potential study participants from each of those partnerships.

Follow-up interviews. The study team plans to conduct follow-up interviews with up to 30 survey respondents. Follow-up interviews will occur approximately three or four months after a participant completes the survey. The study team will invite a mix of both core and non-core partners representing a variety of roles. The follow-up interviews will test whether survey responses map on to what respondents actually do in practice to help confirm whether the survey items are a valid proxy for actual behaviors. For this reason, the sample for the follow-up interviews will be determined based on respondents' answers to the survey. For example, the study team could identify a subset of respondents who indicate in the survey that they often use their REL's study findings to inform improvement to policy or practice; in the follow-up interviews, the study team could ask these respondents whether they had used REL resources to inform improvements in policies or practices over the past few months. The study team will also ask partners whether and how they embed research evidence into tools and resources, as this information is too nuanced to measure accurately in the survey. The study team might also select interview participants who can help explain unexpected survey results. For example, if certain questions yield responses different from what the study team anticipated, the team might follow-up with respondents who scored very high or very low on those items to learn more about the unexpected results. The study team could also ask respondents optional questions aimed at unpacking potential reasons or context for answers to survey questions. For example, survey respondents who reported that REL resources and supports were not helpful might be more likely than other respondents to report during the interview that they felt less valued by their REL colleagues.

B2. Statistical methods for sample selection and degree of accuracy needed

a.   Sample selection

The survey sample will need to be large enough to test the reliability and validity of survey items after allowing for some nonresponse. A sample that is too small can undermine the internal and external validity of the study and compromise the findings due to the insufficient statistical power to answer the key questions about URE that are of interest to IES. On the other hand, a sample that is larger than necessary can add unnecessary respondent and data collection burden, without providing significant benefits in terms of reliability and validity (Andrade, 2020; Faber & Fonseca, 2014). Therefore, it is important to identify the optimal sample size for the study to both render accurate findings while minimizing the extra effort and expense involved in recruiting additional participants.

The REL Peer Review team conducted power calculations (Arifin, 2017; Bonett, 2002) and Monte Carlo simulations (Lee, 2015) to determine the sample sizes it would need to replicate previous types of analyses conducted on existing URE items. Results from the power calculations suggested that at least 30 participants are needed to calculate Cronbach's alpha for each construct and at least 50 participants are needed to calculate confirmatory factor analyses (CFA).

To provide a large enough sample size for these analyses and ensure the sample represents key variation across categories (for example, core/non-core partner, different education roles, multiple RELs), the study team aims to obtain a final sample of 70 participants. To achieve this sample, the study team anticipates needing to invite at least 85 partners to participate in the survey and achieve an 85 percent response rate.

b.   Estimation procedures

The study team will use the following analyses to test the reliability and validity of the survey items:

- **Cronbach's alpha.** Other URE studies typically use Cronbach's alpha to measure the reliability, specifically the internal consistency, of items within a construct (May et al., 2022; Penuel et al., 2017). The study team plans to calculate Cronbach's alpha separately for each construct across all study participants. The study team will also provide descriptive statistics for each construct separately by key subgroups (e.g., school-level, district-level, and state-level staff; core and non-core REL partners). Cronbach's alpha is computed as:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \omega_{c_i}^2 \sigma_{c_i}^2}{\sigma_S^2}\right)$$

  where K is the number of items, σ_S^2 is the variance of the overall score, $\omega_{c_i}^2$ is the weight on indicator i in the calculation of the overall score, and $\sigma_{c_i}^2$ is the variance of indicator i.

- **Confirmatory Factor Analysis.** The study team will use CFA to test the reliability and validity of at least some of the constructs and survey items. Findings from CFA are more psychometrically robust than the findings based on Cronbach's alpha alone, as CFA can be used to calculate a McDonald's omega for each construct, providing another measure of reliability (Hayes & Coutts, 2020). McDonald's omega corrects for the amount of measurement error associated with each survey item and thus offers a more accurate assessment of reliability than Cronbach's alpha. In addition,

Cronbach's alpha is positively affected by the number of items in a scale, whereas McDonald's omega does not depend on the number of items.

McDonald's omega is typically calculated within the context of a broader CFA. The observed score variance includes both the variance of the expected (true) scores and the variance of error scores. Reliability is the ratio of true variance of the scores (signal) and the total variance of the scores (signal + noise):

$$\omega = \frac{\sigma^2_{true\,variance}}{\sigma^2_{true\,variance} + \sigma^2_{error\,variance}}$$

Calculations of the McDonald's omega coefficient account for the differential strength of the relationship between each item and a factor (Deng & Chan, 2017; Trizano-Hermosilla & Alvarado, 2016).

The study team will also use CFA to confirm the structure of the constructs. This CFA could help indicate whether two sets of related items load on two separate factors, or if they load together on one factor. If the study team finds that items generally load on one factor, it may recommend combining items across the two groups into one construct. The factor loadings could also provide information about which items might be the best candidates to drop, if there are opportunities to pare down the items while still fully capturing the construct(s).

- **Pearson's product moment correlation coefficient.** To test convergent validity, which is the degree to which constructs are related to other similar variables and measures, the study team will calculate Pearson's product moment correlation coefficients (Brennan et al., 2017) between measures. For example, the study team may examine the correlation between six items it developed based on existing URE literature to measure respondents' confidence in ability to use research evidence and two existing SFS items that measure respondents' capacity to use research evidence and data to inform decisions.

- **Construct validity.** To test construct validity, the study team will conduct follow-up interviews with a small subset of the study sample to test whether survey responses map on to what respondents *actually* do in practice. This will confirm whether the survey items are a proxy for actual behaviors. For example, the study team could identify a subset of respondents who indicate in the survey that they are planning to use their REL's study findings to inform improvements to policy or practice. The study team could then follow up with these respondents a few months later to see whether they followed through with that plan. The study team will determine which constructs to follow up on, based on the results of the survey.

- **Descriptive statistics.** In addition to testing the reliability and validity of constructs, the study team will also calculate descriptive statistics of the pilot study data including means, standard deviations, and ranges for each construct. The study team will present descriptive statistics across the full sample and for key organizational subgroups that IES is interested in (school-level, district-level, and state level staff; core and non-core partners). As requested by IES, the study team will also produce descriptive statistics for each participating REL region and partnership.

c. *Degree of accuracy needed*

The study team will use a multi-step approach to assess composite reliability of the constructs. First, to inform the psychometric properties of the constructs, to the extent possible, the study team adapted

existing scales with known reliability and validity (for example, the team used existing constructs and survey items with Cronbach's alphas ≥ .70).

Second, to ensure these scales remain sufficiently reliable when used with REL partners, the study team computed the minimum sample size required to reach the desired reliability threshold for Cronbach's alpha within +/-0.10 of the desired reliability (Bonett, 2002, 2003). Desired (or expected) reliability was informed by the reliability of the original scales. If the reliability estimate was not available, the study team computed the minimum sample size required to reach a Cronbach's alpha of 0.70, 0.75 and 0.80, with the precision of +/- 0.10 at the 0.05 significance level. For example, reported reliability of the existing Conceptual Use of Research scale (Penuel et al., 2017), on which the study's survey items are based, is 0.88. Power calculations allowed the study team to determine the sample size that is sufficient to reach the reliability of 0.88 +/- 0.10 (or 0.78 to 0.98). In other words, 0.78 and 0.98 represent the lower and upper endpoints of an exact confidence interval for the construct reliability.

Third, to ensure the proposed sample size is sufficient to conduct a CFA, the study team generated 1,000 samples with varying sample sizes (with N ranging from 30 to 75, depending on the scale and the expected alpha). The study team then evaluated simulated samples to determine the sufficient number of observations to achieve the following criteria (Muthén & Asparouhov, 2002):

- **Parameter estimate bias.** Average of parameter estimates of Monte Carlo samples relative to the known parameter values is within 10 percent.
- **Standard error bias.** Average of standard errors of Monte Carlo samples relative to the known standard errors is within 10 percent.
- **Coverage.** Proportion of the replications where a 95 percent confidence interval covers the known parameter values is between 0.91 and 0.98.
- **Statistical power.** Proportion of significant tests across replications (power to reject the hypothesis that estimated parameter mean is zero when it is not zero) is ≥ 0.80.

The study team found that a sample size of 50 participants is generally sufficient to ensure adequate coverage and statistical power, while minimizing the standard error bias. To provide a large enough sample size for these analyses and ensure the sample represents key variation across categories (for example, core/non-core partner, different education roles, multiple RELs), the study team aims to obtain a final sample of 70 participants. Exhibit B.2 summarizes the accuracy of the statistics informing construct reliability.

Exhibit B.2. Accuracy of the statistics informing construct reliability

	Minimum detectable effect size
Cronbach's Alpha (at the scale level)	+/- 0.10 of the expected alpha
Parameter estimate bias (at the item level)	≤10%
Standard error bias (at the item level)	≤10%
Coverage (at the item level)	between 0.91 and 0.98
Statistical power (at the item level)	≥ 0.80

d.   Unusual problems requiring specialized sampling procedures

The study team does not anticipate any unusual problems that require specialized sampling procedures.

- **Strategies to reduce burden.** The proposed data collection approach reflects several strategies to limit the burden on study participants while ensuring the study team obtains high-quality data. First, the survey was developed with input from the study team's content experts in URE. For each of the eight constructs, the study team carefully reviewed all the survey items that could fit within each construct, and then pared down the number of survey items to only include those that are relevant to REL partners and will allow the study team to adequately measure that construct. This ensures the data collection is limited to a brief 20-minute survey focused on only the key URE topics. Additionally, the follow-up interviews will be with only a subset of the sample members who completed the survey and will focus on only the key questions needed to inform validity of the survey constructs.

B3. Methods to maximize response rates and deal with nonresponse

a.   Maximizing response rates

As shown in Exhibit B.1, the study team expects an 85 percent response rate on the web survey and a 60 percent response rate on the follow-up interviews. REL partners will complete the survey once in spring 2024. The follow-up interviews will take place with a subset of REL partners in summer 2024. The study team will work closely with REL partners to schedule the interviews at a time that is most convenient for them. The study team will follow up with respondents by email to ensure it reaches response rate targets and will provide $30 incentives for participation in the survey and interview.

Across all aspects of data collection, the study team will use strategies that have proved successful on other IES studies. Specific methods for maximizing response rates and minimizing nonresponse in the collection of data in each study component are as follows:

- **Developing and testing web-based surveys to maximize ease of completion and reduce respondent burden.** The study team will develop a web-based version of the survey (Appendix A) to administer to REL partners. It will follow processes that have proved successful for other web-based data collections. The web-based version of the survey will include (1) automated skip patterns so respondents see only the questions that apply to them (including those based on answers provided previously in the survey), (2) logical rules for response options so respondents' answers are restricted to those intended by the question, and (3) easy-to-use navigation tools so respondents can navigate between survey sections and access previous responses. Additionally, the survey will automatically save entered responses, and respondents will be able to revisit the web link as many times as needed to complete the survey.

- **Providing clear and informative emails to REL partners (Appendix C).** The study team will send an advance email to each identified REL partner with an overview of the data collection effort. After sending that email, the study team will follow up with an invitation email that contains specific instructions on how to access and complete the survey. The study team will also send reminder emails during the data collection period to encourage responses. After the survey ends, the study team will send a follow-up email to a subset of REL partners who completed the survey, inviting them to participate in the follow-up interview.

- **Assuring confidentiality to respondents.** The study team will assure respondents about the protection of their responses in language included in the study instruments. This language will state that responses to data collection materials will be used only for statistical purposes, and that the reports prepared for the study will summarize findings across the sample and not associate responses

with a specific district, school, institution, or individual. In addition, all study staff who have access to confidential data will obtain security clearance from ED by completing personnel security forms, providing fingerprints, and undergoing a background check.

- **Conducting the follow-up interviews with trained, experienced research staff.** Trained research staff will conduct all follow-up interviews, following a structured protocol (Appendix B). To minimize interview length and burden, the study team will draw on responses to the survey to collect information that is specific to the respondent's experiences.

b. *Dealing with nonresponse*

To reduce item nonresponse, the web-based questionnaire will include programmed checks alerting respondents to missing or out-of-range responses they enter. These checks will allow respondents to change their response based on guidance provided on the pop-up screen or leave their answer and continue to the next question. The study team will thoroughly test the instrument for clarity, accuracy, length, flow, and wording.

They study team will have few or no options to assess nonresponse bias, given the small sample and the lack of data on sample member characteristics, outside of their position and REL affiliation.

B4. Tests of procedures and methods to be undertaken

During the 60-day public comment period, the study team pretested the survey with fewer than 10 individuals currently in REL partnerships who collectively represented each REL partner type (i.e., core, non-core, school-level, non-school level). After pretesting, the team reviewed feedback and made appropriate revisions to clarify the wording of survey items and response options and calculated the average response time needed to complete the survey.

The study team will program the survey instrument for administration via computer-assisted web surveying methods. Before deployment, the team will test the survey instrument to ensure it functions as designed. This will include extensive manual testing for skip patterns, fills, and other logic. This testing will increase the accuracy of data collected while minimizing respondent burden. Please refer to section B2 for additional information on the methodologies to be used in this study.

B5. Individuals consulted on statistical aspects of the design

The following offered feedback on the statistical aspects of the study:

Name	Title	Telephone number
Mathematica staff		
Ruth Neild	Senior fellow	(609) 275-2341
Martha Bleeker	Principal researcher	(609) 275-2269
Dmitriy Poznyak	Senior researcher, statistics and methods	(609) 716-4382
Felicia Hurwitz	Researcher	(609) 945-3379
Sally Atkins-Burnett	Senior fellow	(202) 484-5279
IES staff		
Liz Eisner	Associate commissioner, Knowledge Utilization Division	(202) 245-6614
Christopher Boccanfuso	Branch chief, Regional Educational Lab Program	(202) 245-6832

Name	Title	Telephone number
Study content experts		
Dr. Elizabeth Farley-Ripple	Professor and director of the Partnership for Public Education, School of Education at the University of Delaware	--
Dr. Drew Gitomer	Professor, Graduate School of Education at Rutgers University	--
Dr. Caitlin Farrell	Director, National Center for Research in Policy and Practice, and associate research professor, School of Education at University of Colorado Boulder	--
Dr. Mark Rickinson	Associate professor, School of Education at Monash University, Melbourne, Australia	--

References

Andrade, C. (2020). Sample size and its importance in research. *Indian Journal of Psychological Medicine, 42*(1), 102–103. https://doi.org/10.4103/ijpsym.ijpsym_504_19

Arifin, W. N. (2017). *Sample size calculator* (Version 2.0) [Spreadsheet file]. http://wnarifin.github.io

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*(4), 335–340.

Bonett, D. G. (2003). Sample size requirements for comparing two alpha coefficients. *Applied Psychological Measurement, 27*(1). https://doi.org/10.1177/0146621602239477

Brennan, S. E., McKenzie, J. E., Turner, T., Redman, S., Makkar, S., Williamson, A., Haynes, A., & Green, S. E. (2017). Development and validation of SEER (Seeking, Engaging with and Evaluating Research): A measure of policymakers' capacity to engage with and use research. *Health Research and Policy Systems, 15*(1). https://doi.org/10.1186/s12961-016-0162-8

Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement, 77*(2), 185–203. https://doi.org/10.1177/0013164416658325

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics, 19*(4), 27–29. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo

Hayes, A., & Coutts, J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures, 14*(1). https://doi.org/10.1080/19312458.2020.1718629

Lee, S. (2015). Implementing a simulation study using multiple software packages for structural equation modeling. *SAGE Open, 5*(3). https://doi.org/10.1177/2158244015591823

May, H., Blackman, H., Van Horne, S., Tilley, K., Farley-Ripple, E. N., Shewchuk, S., Agboh, D., & Micklos, D. A. (2022). Survey of Evidence in Education for Schools (SEE-S) Technical Report. The Center for Research Use in Education (CRUE) & the Center for Research in Education and Social Policy (CRESP), University of Delaware. Survey of Evidence in Education for Schools (SEE-S) Technical Report - Center for Research Use In Education (udel.edu)

Muthén, B., & Asparouhov, T. (2002). *Using Mplus Monte Carlo simulations in practice a note on non-normal missing data in latent variable models* (Mplus Web Notes No. 2). https://www.statmodel.com/download/webnotes/mc2.pdf

Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C., & Allen, A. (2017). How school and district leaders access, perceive, and use research. *AERA Open, 3*(2). https://files.eric.ed.gov/fulltext/EJ1194150.pdf

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7,* 1–8. https://doi.org/10.3389/fpsyg.2016.00769

Mathematica Inc.

Our employee-owners work nationwide and around the world.

Find us at **mathematica.org** and **edi-global.com**.