

Evaluation of the Innovative Assessment Demonstration Authority Pilot Program-Preliminary Activities

Supporting Statement for Paperwork Reduction Act
Submission

PART B: Collection of Information Employing Statistical
Methods

July 2020

Contract # 91990019C0059

Submitted to:
Institute of Education Sciences
U.S. Department of Education

Submitted by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Contents

		<u>Page</u>
B.1.	Respondent Universe and Sample Design.....	1
	B.1.1. District Sample.....	1
	B.1.2. School Sample.....	2
	B.1.3. Teacher Sample.....	3
B.2.	Information Collection Procedures.....	3
	B.2.1. Notification of the Sample and Recruitment.....	3
	B.2.2. Statistical Methodology for Stratification and Sample Selection.....	3
	B.2.2.1. District Sample.....	4
	B.2.2.2. School Sample.....	4
	B.2.2.3. Teacher Sample.....	5
	B.2.3. Estimation Procedures.....	5
	B.2.4. Degree of Accuracy Needed.....	6
	B.2.5. Unusual Problems Requiring Specialized Sampling Procedures.....	7
	B.2.6. Use of Periodic (less than annual) Data Collection to Reduce Burden.....	7
B.3.	Methods for Maximizing the Response Rate.....	7
B.4.	Test of Procedures.....	8
B.5.	Individuals Consulted on Statistical Aspects of Design.....	8
	Appendix A. Instructions for Preparing and Submitting the Teacher List.....	A-1
	Appendix B. Notification Letters and Follow-up Emails.....	B-1

Table

B-1	Subjects, grades, participating districts, and eligible schools involved with the IADA assessment in 2019–20, by IADA program.....	1
B-2	Proposed sample size targets.....	5
B-3	Precision of survey-based estimates.....	7

Part B. Collection of Information Employing Statistical Methods

Since 1994, federal law has required states to regularly administer assessments to students in selected grades and subjects.¹ The purpose of these assessments is to inform teaching and learning, and to hold schools accountable for student performance. To improve the quality and usefulness of these assessments, the law was most recently updated in 2015 to create an Innovative Assessment Demonstration Authority (IADA) Pilot Program. The program (Title I, Section 1204 of the *Every Student Succeeds Act*, or ESSA) allows the U.S. Department of Education (ED) to exempt a handful of states from certain testing requirements if they agree to pilot new types of assessments. ED, through its Institute of Education Sciences (IES), is requesting clearance to recruit school districts and collect teacher lists for the Congressionally mandated evaluation of the IADA program. A second package will request clearance for district, principal, and teacher survey instruments and the collection of these data.

B.1. Respondent Universe and Sample Design

As of July 2020, ED has approved five states for the program: Louisiana and New Hampshire in 2018, North Carolina and Georgia in 2019, and Massachusetts in 2020. The evaluation includes surveys of all pilot districts in the first four IADA states and representative samples of principals and teachers in participating schools within the pilot districts. Surveys will be administered in spring 2021 and spring 2022. Under the IADA pilot program, states are to scale up the innovative assessment to statewide use over a period of five years. As a result, the evaluation team will increase the number of districts, principals, and eligible teachers surveyed between the first and second round of surveys so that the findings from the second year will better represent the larger mix of participants expected at that time.

B.1.1. District Sample

All participating districts in the first four pilot states (Louisiana, New Hampshire, North Carolina, and Georgia) will be surveyed in each of the two survey years. There is one IADA program per state except for Georgia, which is piloting two innovative assessments in different sets of districts. Program implementation may vary across IADA assessments and across districts that are piloting the same innovative assessment, particularly as new districts join the pilot. Including all participating districts in the district survey will allow the evaluation to fully capture variation in implementation across districts. For context, Table B-1 illustrates the grades and subjects, as well as estimated number of districts and schools involved in each innovative assessment for the 2019–20 school year, which is the most recent school year with data available.

Table B-1. Subjects, grades, participating districts, and eligible schools involved with the IADA assessment in 2019–20, by IADA program

IADA program	Subjects and grades involved with the IADA assessment	Estimated number of participating districts and eligible schools
New Hampshire (PACE assessment)	Subjects: ELA, math, science Grades: Varies by subject	15 districts 48 schools (has any of grades 3–8 or

¹ Improving America’s Schools Act of 1994, P.L. 103-382, 20 U.S.C. § 6301 *et seq.*

		high school)
Louisiana (LEAP assessment)	Subjects: Humanities, which integrates ELA and social studies content Grades: 7 in 2019-20 expanding to 6-8 in 2020-21	15 districts 119 schools (has any of grades 6-8)
North Carolina (NCPAT assessment)	Subjects: ELA, math Grades: 4 and 7	3 districts 44 schools (has grade 4 or 7)
Georgia (GMAP assessment)	Subjects: ELA, math, science Grades: 3-8	8 districts 62 schools (has any of grades 3-8)
Georgia (NAVVY assessment)	Subjects: ELA, math, science Grades: 3-8 for ELA and math, 5 and 8 for science, selected high school courses	12 districts 98 schools (has any of grades 3-12)

Notes: The evaluation team does not have lists of all participating schools for each IADA program. The number of eligible schools in the table reflects the number of schools from the Common Core of Data that contain any of the grades that will use the IADA assessment in 2019-20. This number provides an upper-bound estimate of the number of participating schools.

Each state's Annual Performance Report (APR) identifies the participating districts for the school year that just ended (i.e., the reporting year for the APR) and those for the upcoming school year. The APR that covers the 2019-20 school year (to be submitted to ED in fall 2020) will identify districts participating for the 2020-21 school year, and the APR that covers the 2020-21 school year (to be submitted to ED in fall 2021) will identify the districts participating for the 2021-22 school year.

Based on information in the New Hampshire and Louisiana APRs for the 2018-19 school year and IADA application information for Georgia and North Carolina, there were 53 districts participating in the IADA program in 2019-20. Using this information, the evaluation team estimates 64 participating districts for the spring 2021 data collection and 77 districts for the spring 2022 data collection (20 percent growth in participating districts between 2020-21 and 2021-22). These numbers will be verified or updated if needed based on the number of participating districts identified on future APRs.²

B.1.2. School Sample

The evaluation team will use lists of participating schools from state APRs to build the school sampling frame for each survey year. As illustrated in Table B-1, participating schools may vary by grade span (elementary, middle, or high school) across IADA program.

The evaluation team will draw an equal probability sample of schools within each district for each data collection year. The evaluation team will select two schools per district for the spring 2021 survey and

² The estimated number of districts for Georgia and North Carolina in 2019-20 are based on their IADA applications. These states have not yet submitted the APR for their first year of implementation (i.e., the 2019-20 school year).

three schools per district for the spring 2022 survey.³ This approach is expected to yield 128 sampled schools for spring 2021 and 231 sampled schools for spring 2022. These numbers will be verified or updated based on information in future APRs. If the number of districts is larger than expected, the total number of schools sampled will be accordingly revised upward. See B.2 for details on the school sampling approach.

B.1.3. Teacher Sample

Participating teachers are those whose grade and subject (or course for high school teachers) is identified as part of the IADA program in the data collection year.

The evaluation team will select all participating teachers in schools with no more than six participating teachers and randomly select exactly six participating teachers in all other schools in the sample. This approach is expected to yield 640 sampled teachers for spring 2021 and 1,155 sampled teachers for spring 2022. These numbers will be verified or updated based on information in the future APRs on the number of districts participating for each survey year. If the number of districts is larger than expected, the number of teachers sampled will be accordingly revised upward. See B.2 for details on the teacher sampling approach.

B.2. Information Collection Procedures

B.2.1. Notification of the Sample and Recruitment

for each school sampled for the evaluation. The list will include the teacher's grade and subject, which the evaluation team will use in the sampling process (see Section B.2.2.3 for additional details). District Coordinators will prepare a list of participating teachers according to written instructions (see Appendix A for instructions).

B.2.2. Statistical Methodology for Stratification and Sample Selection

The evaluation team's sample design is guided by the key analysis objectives to produce reasonably precise pooled estimates of (1) district and educator perspectives in 2020–21 and 2021–22, and (2) year-to-year change in these perspectives (based on comparisons of the cross-sectional data in 2020–21 and 2021–22). The following sections provide details on the sampling approaches.

B.2.2.1. District Sample

There are no sampling methods for the district survey. **All participating districts in the four pilot states will be surveyed** in each of the two survey years (spring 2021 and spring 2022). By including all districts in the evaluation, the team prioritized the precision of the district estimates, which is expected to have a high degree of between-district variation in survey responses given the differences in pilot programs and degree of scale up.

B.2.2.2. School Sample

³ If a district has fewer than two participating schools for spring 2021 or three participating schools for spring 2022, all participating schools in that district will be selected for the evaluation.

The evaluation team will **select two schools per district for the spring 2021 data collection, and three schools per district for the spring 2022 data collection**. This number of schools per district will yield a large enough sample size to make reasonably precise estimates for the population of participating schools in the four IADA states, while prioritizing the precision for the district estimates (see Section B.2.4 for additional details on the precision of estimates). At least two schools per district are needed to be able to estimate the variance of survey estimates.⁴ The additional school in each district for spring 2022 reflects the anticipated scaling up of the intervention in the district. Using these minimum sample sizes for schools in each district limits the burden and study costs, while allowing for a focus on obtaining the most precise estimates for districts. Variation in survey responses are expected to be larger across districts, relative to the variation across schools within the same district.

The evaluation team will **select overlapping samples of participating schools for the spring 2021 and spring 2022 data collections**. For districts that entered the sample for spring 2021, the evaluation team will maximize the overlap of schools selected for spring 2021 and spring 2022 using a “keyfitzing” approach.⁵ Note that increasing the overlap through keyfitzing does not guarantee that all spring 2021 sampled schools will be in the spring 2022 sample, as keyfitzing is designed to assure that the probabilities of school selection in the second year are the same for all schools.⁶ Increasing the overlap through this approach will improve the precision of estimates for change over time while preserving the precision of the spring 2022 estimates. For districts that enter the program for spring 2022, the three schools selected for the sample will be new to the study.

The evaluation team will **implicitly stratify participating schools by grade span**. In cases where a participating district includes more than one grade span (e.g., primary and secondary), the evaluation team will implicitly stratify participating schools within districts by grade span prior to drawing the school sample. This approach will increase representativeness of the school sample regarding teacher experiences that differ between primary and secondary schools.

B.2.2.3. Teacher Sample

The evaluation team will **select all participating teachers in schools with no more than six participating teachers and randomly select exactly six participating teachers in all other schools** in the sample. Participating teachers are those whose grade and subject (or course for high school teachers) is identified as part of the IADA program in the data collection year. This number of teachers per school

⁴ Unless there is only one participating school in the district, in which case there is no variance.

⁵ The original reference for this is Keyfitz, N. (1951). Sampling with Probabilities Proportionate to Size: Adjustment for Changes in Probabilities. *Journal of the American Statistical Association*, 46, 105-109. A comprehensive reference for the later extensive work on this (up to 1999) is Ernst, L. R. (1999). The maximization and minimization of sample overlap problems: A half century of results. *Bulletin of the International Statistical Institute, Proceedings*, 58.

⁶ For example, if a school has a probability of selection of 90% for spring 2021 and only 10% for spring 2022, then to assure the school really has a 10% chance of selection for spring 2022, it is necessary to give the school some chance of deselection (not being selected for spring 2022). A fully longitudinal sample would ignore this consideration and retain the school through spring 2022 with its too-large chance of selection, which, when weighted would erode the precision of the cross-sectional spring 2022 estimate. Under keyfitzing, the school's weight for spring 2022 will reflect the correct chance of selection.

will yield a large enough sample size to make reasonably precise estimates of the population of teachers participating in the four IADA states, while prioritizing the precision for the district estimates (see Section B.2.4 for additional details on the precision of estimates). In addition, because larger within-school correlations for teacher survey responses are expected (because implementation experiences within schools are more likely to be similar), having larger teacher sample sizes in some schools would not confer much benefit.

The evaluation team will **select overlapping samples of participating teachers in Year 1 and Year 2**. As with the school sample, the evaluation team will increase the overlap of the teacher samples for the 2021 and 2022 data collections through keyfitzing. This approach will improve the precision of estimates for change over time without eroding the precision of the Year 2 estimates.

The evaluation team will **implicitly stratify teachers by grade and subject** where relevant to improve the representation of teacher experiences in the sample across grades and subjects. For example, in cases where a school goes from administering innovative assessments only in grade 7 in 2020–21 to grades 6, 7, and 8 in 2021–22, implicit stratification by grade level will improve the mix between teachers who are new to these assessments (i.e., grade 6 and 8 teachers) and teachers who have already been using them (i.e., grade 7 teachers).

Table B-2 presents the sample size targets based on the sampling approached described above.

Table B-2. Proposed sample size targets

Respondent	Data collection Year 1* (2020–21)	Data collection Year 2* (2021–22)
Districts	64	77
Schools	128	231
Teachers	640	1,155

*Targets are based on the current information available from the 2019-20 school year. In 2019-20, there were 53 participating districts across the five IADA programs. The table assumes 20 percent growth in the number of participating districts between the 2019–20 and 2020–21 school years and between the 2020–21 and 2021–22 school years. The team will revisit sample size targets once lists of participating districts and schools are available from the IADA state APRs.

B.2.3. Estimation Procedures

The primary analysis of the subsequent survey data will be pooled analyses (drawing on data across participants). Responses to survey questions will be tabulated into descriptive statistics (such as percentages) and simple statistical tests (such as tests for differences between percentages across years). These tabulations will provide a snapshot of district, school, and teacher experiences at each time point, as well as aggregate changes over time. Because of the use of a statistical sample, survey data presented for schools and teachers will be weighted to generalize findings to participating schools and participating teachers in the IADA program in 2020–21 and in 2021–22.

B.2.4. Degree of Accuracy Needed

Program implementation may vary across IADA assessments and across districts that are piloting the same innovative assessment, particularly as new districts join the pilot. Given the possible variation at the district level, the evaluation sampling approach is designed to prioritize the precision of the district estimates by selecting all participating districts for the surveys. We expect schools-within-districts to be more highly correlated than across districts, and teachers-within-schools to be more highly correlated than across schools. This expectation influenced the number of sampled schools per district and sampled teachers per school, as described in B.2.2.2 and B.2.2.3.

Table B-3 provides information on the precision of survey-based estimates given the sample size targets. No margin of error estimates are provided for the district survey because the survey will be conducted in all participating districts, not a sample. The estimates assume a response rate of 85 percent for principals and teachers each survey year. The estimates assume a 0-1 outcome measure and that 20 percent of the variation across schools is explained by the district in which they are located. The estimates assume that variation in the teacher selection probabilities increases the variance of the teacher survey estimates by 10 percent, and that there is a 35 percent clustering effect. The estimates assume that the correlation between principal survey responses between spring 2021 and spring 2022 for the same schools is 67 percent, and the correlation between responses to the teacher survey between spring 2021 and spring 2022 is 50 percent. Minimum detectable differences (MDD) between spring 2021 and spring 2022 assume 95 percent confidence in a two-sided test of the null hypothesis of no difference and 80 percent power.

Under these assumptions, the evaluation team expects margin of error estimates based on a 95 percent confidence interval of plus or minus 10 percentage points and plus or minus 8 percentage points at the principal and teacher levels in spring 2021, respectively. The margins of errors improve to plus or minus 8 percentage points and plus or minus 7 percentage points in spring 2022 with the increase in sample sizes. The MDD of year-to-year change is 11 percentage points at both the principal and teacher levels. These MDDs reflect a reasonable compromise between the need to detect policy relevant changes in an evolving program while limiting burden on schools and districts. This threshold is similar to the one used to identify policy-relevant differences in recent IES implementation studies.⁷

Table B-3. Precision of survey-based estimates

	Margin of error for the percent of schools or teachers		Minimum detectable difference of year- to -year change
	Spring 2021	Spring 2022	
Principal survey	10.3 pp	8.3 pp	11.1 pp
Teacher survey	8.4 pp	7.4 pp	11.3 pp

Notes: The estimates are measured in percentage points (pp). For the teacher estimates, the evaluation team assumed a mean value of five sampled, participating teachers per sampled school. Schools with six or more eligible teachers will have a sample size of six, but we anticipate there will be many schools with fewer than six eligible teachers.

⁷ For example: Epps, S.R., Jackson, R.H., Olsen, R.O., Shivji, A., & Roy, R. (2016). *Upward Bound at 50: Reporting on Implementation Practices Today* (NCEE 2017-4005). Washington, DC: National Center for Education Evaluation, Institute of Education Sciences, U.S. Department of Education.

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

District Coordinators will be asked to provide teacher lists a year apart to reflect possible changes in a school's teaching staff from year to year, as well as to account for program scale up within a district to more schools. The IADA program requires states to scale up the new assessments on a relatively aggressive timeline (approximately five years). A longer period between data collections would make it difficult for the survey findings to be timely enough to inform the expected rapid developments in the IADA program, along with future development and use of innovative assessment systems. In fact, a longer period of time between data collections would prevent the evaluation from meeting the Congressionally mandated timeline for the evaluation report.

B.3. Methods for Maximizing the Response Rate

The evaluation will mail the District Coordinator the teacher list collection request (see Appendix A) as part of the Coordinator's notification letter (see Appendix B). Obtaining teacher rosters is critical to ensuring the quality of the teacher sample. The evaluation team will engage in a number of activities to maximize the teacher list response rates. The team will offer, where allowed, District Coordinators a small incentive for providing the teacher lists. The team will accept the teacher list in whatever format is most convenient for the district (e.g., an excel file, hard copy, scanned images). The team will begin reaching out to District Coordinators about two weeks after the initial mail out to see if they have any questions about the evaluation or the teacher list request. The team will begin telephone follow-up for nonresponse about three weeks after the initial request has been mailed. Experienced telephone interviewers will be trained in prompting the nonrespondents and will be monitored by Westat supervisory personnel during all interviewing hours. The evaluation team expects to achieve at least an 85 percent response rate for the teacher lists for sampled schools. Teacher list nonresponse will be considered during the development of weights for the teacher survey data (to be described in a subsequent OMB package requesting clearance for administering the actual teacher survey).

B.4. Test of Procedures

The participating teacher list request will be pretested with nine or fewer district respondents. The pretests will be conducted via telephone calls and focus on ensuring that terminology used in the request is consistent with the way districts identify participating teachers. The pretests also will be a check to see if the teacher list request can be completed within the estimated 120-minute timeframe.

B.5. Individuals Consulted on Statistical Aspects of Design

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

Patty Troppe, Westat, Project Director
Art Thacker, HumRRO, Principal Investigator
Lou Rizzo, Westat, Senior Statistician
Rob Olsen, Westat, Quality Support Advisor