Evaluation of Preschool Special Education Practices Efficacy Study

Part B: Supporting Statement for Paperwork Reduction Act Submission

OMB Information Collection Request 1850-0916

March 25, 2019

This page has been left blank for double-sided copying.

CONTENTS

TABLES

This page has been left blank for double-sided copying.

PART B. SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This package requests clearance for data collection activities to support a rigorous efficacy study of an instructional framework designed to address the needs of all preschool children in inclusive classrooms.[1] The efficacy study is part of the Evaluation of Preschool Special Education Practices (EPSEP), which is exploring the feasibility of a large-scale effectiveness study of an intervention for preschool children in inclusive classrooms. The Institute of Education Sciences (IES) in the U.S. Department of Education has contracted with Mathematica Policy Research and its partners the University of Florida, University of North Carolina at Chapel Hill, and Vanderbilt University to conduct EPSEP (ED-IES-14-C-0001).

The main objective of the efficacy study is to test whether the Instructionally Enhanced Pyramid Model (IEP_M) can be implemented with fidelity. IEP_M is comprised of three established individual interventions for children with disabilities integrated together into a single comprehensive intervention for use with all children in inclusive preschool classrooms (IEP_M is described in detail in Part A section A1.c). The secondary objective is to provide initial evidence about IEP_M's impacts on classroom and child outcomes. This study provides an important test of whether strategies for delivering content in a manner that meets the needs of each child with a disability can be integrated with an existing framework of teaching practices for inclusive preschool classes, thus helping all children participate and make progress in the general preschool curriculum. These strategies, which are called targeted instructional supports, have been tested separately but have not been tested as part of this framework.

Findings from an earlier EPSEP survey data collection (OMB 1850-0916, approved March 26, 2015) and systematic review provided little evidence that curricula and interventions that integrate targeted instructional supports are available for school districts to use in inclusive preschool classrooms. These earlier findings justify the need for an efficacy study to obtain more information before IES decides whether to conduct a large-scale evaluation. In addition, the results can inform preschool instructional practices and policy objectives in the Individuals with Disabilities Education Act (IDEA) that support inclusion.

The efficacy study will include data collection to conduct both implementation and impact analyses. The implementation analysis will use observation data to describe the fidelity of training and implementation. It also will draw on coaching logs and coach interviews to describe program implementation.[2] In addition, responses to a teacher survey and teacher focus groups will provide information on teachers' backgrounds, professional experiences, and perspectives on IEP_M implementation. The impact analysis will use data from observations of classroom inclusion quality and engagement, a child observation, a direct child assessment, and teacher reports on child outcomes. The implementation and impact analyses also will use district administrative records to offer additional contextual and background information on the preschool program, its teachers, and enrolled children.

---

[1] We define inclusive classrooms as classrooms in which children with disabilities are educated alongside other children and receive most or all of their special education services.

[2] IES is not requesting approval for the collection of data that the study team will collect and that will not impose any burden on teachers or district staff. Examples include coaching logs, coach interviews, and observations.

This supporting statement describes the study sample, our plans for maximizing response rates, methodological tests we will conduct, and the data collection procedures we will use.

Collection of information employing statistical methods

B1. Respondent universe and sampling methods

The efficacy study will rely on a purposive sample of 26 schools from three school districts in the United States. Approximately 40 inclusive preschool classrooms (that is, preschool classrooms that include children with and without disabilities) and associated teaching staff (two per classroom) will be included from the 26 schools. We plan to recruit districts, schools, and teachers during winter 2018 and spring 2019. In late spring 2019, we will randomly assign half of the schools in each district to the IEP_M intervention group and half to the business-as-usual control group. The IEP_M implementation period will include the 2019–2020 and 2020–2021 school years. The child sample will include all preschool children attending study classrooms; we anticipate up to 1,440 children across the two years combined. The study will not statistically sample districts, schools, classrooms, or children; therefore, we will not make statements that generalize beyond the study sample.

B2. Procedures for the collection of information

a. Statistical methods for sample selection

The sample will be chosen purposively in support of the study's objective to learn whether the components of IEP_M can be implemented together with fidelity and have positive effects on preschoolers' social-emotional/behavioral skills and language outcomes. For the efficacy study, the goal for sample selection is not to represent a broad population of schools and children, but to preliminarily test implementation with teaching staff willing to pilot IEP_M in their preschool classroom. If IEP_M can succeed in a favorable context, a larger effectiveness trial can be conducted to evaluate whether IEP_M can succeed with a more representative population of schools and children. Findings from this efficacy study will inform the decision whether to conduct a larger effectiveness trial.

Selection of districts, schools, and classrooms. The best context for evaluating IEP_M is one in which children with disabilities are taught in inclusive preschool classrooms and in which classrooms are not currently using any components of IEP_M or its related fidelity instruments. Members of the IEP_M provider team have relationships with the public preschool programs near their locations that satisfy these selection criteria.

We will work with district staff to identify schools to participate in the study that have inclusive preschool classrooms. To minimize the potential for families to make school enrollment decisions based on the availability of IEP_M, schools must be neighborhood schools (rather than schools of choice). The participating schools in each district must be using the same general preschool curriculum, to which IEP_M will be added. The study will include inclusive preschool classrooms and associated teaching staff (two per classroom) from participating schools. Based on findings from the EPSEP school district survey (OMB 1850-0916, approved March 26, 2015), we anticipate that schools will have one or two inclusive preschool classrooms, on average.

Selection of children. We will seek to include all children in the inclusive preschool classrooms in participating schools. Based on the EPSEP school district survey data and on input from the provider team, we anticipate that study classrooms will have an average of 18 children, 5 of whom will have a disability. Therefore, we anticipate the total number of children will be 720 per year across the 40 classrooms, or up to 1,440 across both school years. We assume that we will obtain data at the end of each school year for a consented sample that includes, on average, 4 children with disabilities and 9 children without disabilities from each study classroom (a total of 520 children per year, or up to 1,040 across both school years).

b.   Estimation procedures

The efficacy study will include four broad sets of analyses: (1) implementation analyses; (2) impact analysis of the average treatment effect on classroom outcomes and child outcomes for the full child sample; (3) impact analysis of the average treatment effect for key subgroups of children; and (4) impact analysis of quantile treatment effects for the full child sample (that is, an analysis of how the intervention affects the *entire outcome distribution*, not just the mean).

Implementation analyses. A core objective of this efficacy trial is to assess the implementation of IEP_M. If IEP_M cannot be implemented with fidelity in this efficacy trial, it may not make sense to proceed with a larger effectiveness evaluation. In addition, understanding the implementation experiences and challenges of districts, schools, and teachers participating in the intervention, as well as implementation costs, will provide important information for a later effectiveness trial.

We will conduct two types of implementation analyses. First, we will describe implementation supports and experiences using measures available only for the intervention group. These measures include training fidelity observations, coaching logs, focus groups with teachers, interviews with coaches, and data on IEP_M costs. Second, we will examine the difference between the intervention and control groups on measures appropriate for use in both groups. These measures include a teacher background and experiences survey, as well as the intervention fidelity measures associated with IEP_M's three component interventions. (Part A, Section A1, provides for more information on IEP_M.) In essence, these differences between the intervention and control groups can be regarded as estimates of impacts on intermediate outcomes. For example, differences on the intervention fidelity measures would signify that IEP_M is having effects on teachers' practices in the classroom.

Average treatment effect for the full sample. We will estimate the impact of IEP_M on the average outcome of all children and classrooms after both year 1 and year 2 of the evaluation. We will use a regression model to adjust for baseline differences in the characteristics of the schools, teachers, and children in the intervention and control groups. Such differences could arise by chance despite the fact that schools were randomly assigned to intervention and control groups.

Average treatment effect for key subgroups. IEP_M is a multitiered intervention intended to provide appropriate supports to preschool children with disabilities. Therefore, we will estimate impacts separately for children with identified disabilities (IEPs). However, there may also be children who are not yet identified but who are at risk. To address this category of students, we will define an analytic subgroup that includes children who have the greatest difficulties with

social interaction and behavior challenges as measured by fall teacher reports on the Social Skills Improvement System (SSIS). This subgroup would be defined without considering whether children have an identified disability. By defining this subgroup, we will be able to estimate impacts of IEP_M separately for children at risk for social-emotional or behavioral challenges.

Quantile treatment effects. An important analytic issue in an evaluation of a multitiered intervention like IEP_M is the need to *examine impact heterogeneity*. IEP_M is specifically designed to provide differentiated learning experiences to preschool children based on their individual abilities and needs, creating the strong potential for heterogeneous effects. Calculating an average treatment effect for the full sample could mask important variation in impacts across the full outcome distribution.

To address this issue, we will complement traditional subgroup analysis with quantile treatment effects (Doksum 1974; Firpo 2007; Koenker and Bassett 1978; Lehmann 1974; Schochet et al. 2014). A quantile treatment effect is the impact of the intervention on a specific quantile of the outcome distribution (for example, the 25th percentile). Just as an unadjusted average treatment effect can be calculated as the difference in means between the intervention and control groups, an unadjusted quantile treatment effect can be calculated as the difference in quantiles between the intervention and control groups (for example, the difference in the 25th percentile between the two groups).

Compared to calculating average treatment effects on the full sample, this approach will provide a more complete representation of the impact of IEP_M and increase study power. The potential advantage of this approach over traditional subgroup analysis is that it may not be possible to precisely identify the subgroups of interest. For example, we would ideally want to calculate average treatment effects for the subgroup of children at risk of eventually needing IDEA services; however, precisely identifying those children at baseline may prove challenging, especially in preschool. Calculation of quantile treatment effects does not depend on identifying subgroups of children at baseline.

c.   Approach to interpreting impact estimates

Our approach to interpreting impact estimates is designed to support the primary decision that will be informed by findings from this efficacy study: whether to conduct a larger-scale effectiveness evaluation of IEP_M. Most studies that IES's National Center for Education Evaluation and Regional Assistance (NCEE) conducts are effectiveness trials whose findings are intended to inform the decisions of policymakers at the federal, state, and local levels. In those studies, the approach to interpreting impact estimates is driven primarily by a desire to avoid the mistake of concluding an intervention is effective when it is not. That mistake is called a Type I error under the Neyman-Pearson (1933) framework for statistical testing of competing hypotheses. Consequently, those studies typically conduct two-tailed hypothesis tests using alpha = 0.05.

In this efficacy study, the real-world consequences of a Type I error are much smaller than in most other studies that NCEE conducts. The primary decision that will be informed by this study is whether to conduct another study on a larger scale. This means that a Type I error will not adversely affect schools, teachers, and children. Meanwhile, the cost of concluding that IEP_M is ineffective when in fact it is effective (a Type II error) could be considerable. By failing to

conduct a subsequent effectiveness study, federal, state, and local policymakers would be deprived of important evidence establishing the effectiveness of IEP_M.

In light of these considerations, we will use an approach to interpreting impact estimates that strikes a different balance between Type I and Type II errors than most NCEE evaluations. Our approach to balancing these errors is informed by Westlund and Stuart (2017) and by Lee et al (2014). Westlund and Stuart (2017) show that two-tailed testing with alpha=0.05 can lead to a very high Type II error rate, meaning that many effective interventions would never be studied at scale using that approach. Lee et al (2014) offered guidance for a better approach to inference in efficacy trials; they recommend 85 or 75 percent confidence intervals and Bayesian methods.

Instead of conducting two-tailed hypothesis tests with alpha = 0.05, we will conduct a one-tailed test with alpha = 0.10 to reduce the probability of a Type II error. This approach is equivalent to using an 80 percent confidence interval, which is consistent with the guidance from Lee et al (2014). Furthermore, we will supplement the traditional hypothesis testing approach with Bayesian posterior probabilities that can be used to directly assess the probability that an intervention had positive effects—something that *p*-values and statistical significance cannot do (Greenland et al. 2016; Wasserstein and Lazar 2016).

Table B.1 provides an example of how impact estimates can be presented and interpreted using Bayesian posterior probabilities alongside traditional measures. In this table, we consider a hypothetical scenario in which the estimated impacts are 0.25, 0.05, and 0.11 standard deviations for children with disabilities (or separately, for children at risk for disabilities), children without disabilities, and all children, respectively. Given these example impact estimates, we calculate *p*-values using standard errors based on our anticipated sample size and assumptions regarding the intraclass correlation and regression R^2. To calculate the probability that the impact is truly positive, we also use prior evidence from the What Works Clearinghouse.[3] In this example, the impact for children with disabilities at baseline is 0.25 standard deviations, the *p*-value is 0.04, and the probability that the impact is truly positive is 94 percent.

Table B.1. Example of our approach to interpreting impact estimates

Groups of children	Estimated impact (effect size)	*p*-value	Probability that impact is truly positive
Disability at baseline	0.25	0.04	0.94
At risk of social-emotional or behavioral challenges at baseline	0.25	0.04	0.94
No disability at baseline	0.05	0.32	0.67
All children	0.11	0.13	0.85

Notes: These calculations assume that the data for the impact analysis come from 26 schools and 1.5 classrooms per school. We assume an intraclass correlation of 0.10, a school-level R^2 of 0.70, a child-level R^2 of 0.40, and one-tailed hypothesis testing when calculating standard errors and *p*-values. The "estimated impacts" are hypothetical (not based on real data). The prior distribution used to calculate the probability that an

[3] We conducted a preliminary analysis of all impact estimates that meet WWC evidence standards. In this analysis, we made statistical adjustments for the varying precision of the prior evidence and for a strong positive correlation between impact estimates and the standard errors of those estimates (such a correlation is consistent with the phenomenon known as "*p*-hacking" or "file drawer bias" After those adjustments, the impact estimates in the WWC database are approximately normally distributed, with mean 0 and a standard deviation of 0.20.

impact is truly positive is normal, with mean 0 and standard deviation 0.20. This prior distribution is based on evidence from the What Works Clearinghouse.

d.   Degree of accuracy needed

When calculating the average treatment effect, we estimate that the efficacy study will achieve a minimum detectable effect size of 0.21 standard deviations on child outcomes for the full sample (Table B.2). For the subgroups of children with disabilities or, those at risk for social-emotional or behavioral challenges, the minimum detectible effect size is 0.30 standard deviations on child outcomes. The minimum detectable effect size for teacher and classroom outcomes are 0.39 and 0.53. These targets represent meaningful but realistic impacts in an efficacy study. Prior studies of the components of IEP_M have found effect sizes larger than these minimum detectable effect sizes. For example, the effect sizes that Hemmeter et al. (2016) estimated for the Pyramid Model for Promoting Social and Emotional Competence in Young Children were 0.43 on a measure of social skills and -0.29 standard deviations on a measure of problem behaviors. Strain and Bovey (2011) estimated impacts of the Learning Experiences Alternate Program for Preschools and Their Parents (LEAP) on social-emotional/behavior and language outcomes ranging from 0.64 to 1.41 standard deviations. Our proposed sample sizes will be sufficient to detect impacts of these magnitudes.

Table B.2. Minimum detectable effect sizes

Groups of children in each classroom	Minimum detectible effect size		
	Child outcomes	Teacher survey outcomes	Classroom observation outcomes
4 children with disabilities	0.30		
4 children at risk for social-emotional or behavioral challenges	0.30	0.39	0.53
9 children without disabilities	0.23		
Full child sample (13)	0.21		

Notes: These calculations assume that the data for the impact analysis come from 26 schools and 1.5 classrooms per school. We assume an intraclass correlation of 0.10, a school-level R^2 of 0.70, and an R^2 of 0.40 at both the classroom and child levels. The minimum detectable effect size is calculated assuming a one-tailed test, alpha = 0.10, and 80 percent power. We assume outcomes from the teacher survey will be available for 80 teachers (two staff per classroom).

When calculating quantile treatment effects, the study has the potential to detect an impact of IEP_M even when the average effect is smaller than the minimum detectable effect sizes reported in Table B.2. This potential can be realized if the impacts of IEP_M vary across the outcome distribution, as in Table B.3.[4] For example, Table B.3 shows that calculating quantile treatment effects gives the study 79 percent power to detect an impact of IEP_M when the true

---

[4] The pattern of impact heterogeneity examined in Table B.3 also yields a small increase in power for the average treatment effect (relative to homogenous impacts) because the intervention reduces the variance of the outcome in the intervention group. This advantage can be seen when comparing the second row of this table to the last row of Table B.2. In this table, the average treatment effect analysis has 82 percent power to detect an impact of 0.20 standard deviations. This is more power than in Table B.2, where there is 80 percent power to detect an impact of 0.21 standard deviations.

average treatment effect is just 0.15 standard deviations *if* impacts are much larger for children at the bottom of the outcome distribution than at the top.[5]

Table B.3. Calculating quantile treatment effects can increase study power

True average treatment effect	True quantile treatment effects on the:			Power	
	10th percentile	50th percentile	90th percentile	Average treatment effect	Quantile treatment effect
0.25	0.50	0.23	0.04	0.93	0.99
0.20	0.39	0.18	0.04	0.82	0.92
0.15	0.31	0.13	0.02	0.64	0.79

Notes: Effects are reported in standard deviations of the outcome; power is reported as a percentage. Power for quantile treatment effects is calculated using simulations and takes into account multiple hypothesis testing across quantiles. We assume 26 schools, 1.5 classrooms per school, a total of 520 children in the analysis sample, ICC of 0.10, school-level R^2 of 0.70, and child-level R^2 of 0.4. Power is calculated assuming a one-tailed test with alpha = 0.10.

e.   Unusual problems requiring specialized sampling procedures

We do not anticipate any unusual problems that require specialized sampling procedures.

f.   Use of periodic (less frequent than annual) data collection cycles to reduce burden

These data will be collected during the 2019–2020 and 2020–2021 school years.

g.   Who will collect the information and how it will be done

Field staff from the study team will collect data from districts, schools, teachers, and parents/children, as we describe next.

Training field staff. Field staff will conduct data collection activities in schools during the fall and spring of each year. This includes classroom observations and teacher-child reports at both points in time while child observations, child assessments, and the teacher background and experiences survey take place each spring. Field staff will also work closely with school staff to coordinate gathering consents each fall. We will hold three trainings for field staff per study year. The first training, to be held the summer before each study year, will prepare a team of observers to collect intervention fidelity data. The second training will focus on conducting the observations of classroom inclusion quality and engagement, collecting the teacher-child reports, and gathering consents. At this training, field staff will be trained and certified on the classroom observation protocols. The third training, held in early spring, will focus on child assessments, individual child observations and distributing and collecting teacher background survey. It will also include a short refresher on classroom observations.

Obtaining parental consent. We will collect parental consent in the fall of both study years. The consent process will begin two to four weeks after the start of school, after class rosters have stabilized. We anticipate that it will take three to four weeks to obtain consents (by the end of September). Field staff will visit each school and coordinate with school staff to

---

[5] In this example, 79 percent power means that if the impacts truly look like what is in the third row of Table B.3, there is a 79 percent chance that at least one of those three individual quantile treatment effects will be statistically significant after adjusting for multiple comparisons.

gather classroom rosters, explain the study to teachers, answer any questions teachers have about the consent process, and confirm fall data collection dates. They will then ask teachers to distribute consent packets for each child in the study classrooms. Teachers will give out consent packets for children to take home and will collect and return signed consent forms to the field staff team.

Fall data collection activities. Fall data collection will be conducted between the start of the school year and October. On the scheduled data collection dates, a trained field staff will conduct intervention fidelity observations over 1.5 school days in both intervention and control schools. Another trained field staff member will visit each study school to conduct an observation of classroom inclusion quality and engagement in each study classroom (one day per classroom) and distribute teacher-child report forms. Field staff will work with the lead teacher to determine which teacher-child reports the lead teacher should complete and which ones the other teacher in the classroom should complete.

Spring data collection activities. Spring data collection activities will occur during two 4- to 6-week periods starting in March and in April, respectively. In March, trained field staff will conduct intervention fidelity observations over 1.5 school days in both intervention and control schools. In addition, one or two researchers from the Mathematica study team will visit each district for two days during this period to conduct in-depth interviews with coaches and focus groups with teachers receiving the IEP$_M$ program.

The April data collection will include the teacher-child reports, observations of classroom inclusion quality and engagement, child observations, and child assessments. It will also involve distributing and collecting the teacher background and experiences survey. This data collection effort will require more field staff and longer visits at each school than the fidelity observations. Table B.4 provides an example of a typical week during the April data collection. On the scheduled data collection dates, two field staff members will visit each school for five days. In addition to conducting classroom observations and distributing and collecting the teacher-child reports, they will conduct four child observations and child assessments with all study children in the classroom. Child assessments will be conducted one-on-one with children taken out of the classroom for the assessment and returned after the assessment. To the extent possible, we will ask teachers to complete teacher-child reports for the same children they reported on in the fall. Make-up visits for any missed child assessment and child observations will be conducted as needed.

Table B.4. Example of typical data collection week: April

	Monday	Tuesday	Wednesday	Thursday	Friday
School 1					
Classroom A	Classroom observation[1] *(2.5 hours)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	5 Child assessments[3] *(30 minutes/child)*
Classroom B	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	Classroom observation[1] *(2.5 hours)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	5 Child assessments[3] *(30 minutes/child)*
School 2					
Classroom C	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	Classroom observation[1] *(2.5 hours)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	5 Child assessments[3] *(30 minutes/child)*
Classroom D	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	4 Child Observations[2] *(20 minutes/child)* 3 Child assessments[3] *(30 minutes/child)*	Classroom observation[1] *(2.5 hours)*	5 Child assessments[3] *(30 minutes/child)*

[1] Classroom observations will be conducted using the Inclusive Classroom Profile and Engagement Check.

[2] Child observations will be done with the Target Child Observation System on up to four children selected by the teacher over three days.

[3] Child assessments will be conducted with the Test of Early Reading Ability, Third Edition on an average of 14 children per classroom over five days.

In both fall and spring, each team member conducting classroom observations will receive a quality assurance visit from a gold standard[6] observer to ensure the quality of the data collection. In spring, gold standard observers and assessors will also conduct at least one quality assurance visit to observe each team member conducting child assessments and observations. All hard-copy instruments will be returned to Mathematica via express mail and reviewed for quality. All instruments will be data entered and 100 percent verified.

Obtaining administrative records from districts. Mathematica research staff will work with a district liaison to collect administrative records for all study children from the district in summer 2020 and 2021. These electronic records, uploaded to a secure project website, will include demographic, attendance, and curriculum-linked assessment information. We will also ask districts to provide cost information, including staff, substitute teacher, and facilities rates.

---

[6] A gold standard observer/assessor is a staff member who has received special training on the classroom and/or child observations and/or child assessments and usually is involved in the on-site quality assurance visits.

B3. Methods to maximize response rates and deal with nonresponse

The EPSEP efficacy study will use several approaches to maximize response rates, while minimizing burden on respondents. To maximize response rates for this information collection, we will take the following steps:

- **Use trained and experienced data collection staff.** Field staff will work with a coordinator at each school who will serve as a liaison between the study team and school staff. All research and field staff assigned to the study will participate in extensive project-specific training to ensure that they are ready to respond effectively to respondents' questions and develop their skills for securing respondent cooperation. They also will be trained in techniques to conduct study activities efficiently and with minimal disruptions to school staff and children.

- **Use data collection procedures that follow district requirements, protect confidentiality, and minimize burden.** We will adhere to any data collection requirements that districts may have, such as preparing research applications and seeking institutional review board approvals. We will also include a statement on confidentiality and data protection (Education Sciences Reform Act of 2002, Title I, Part E, Section 183) in all letters and data collection instruments. We plan to rely on administrative records where possible to minimize burden on schools and parents. We do anticipate full district participation addressing requests for administrative records, consistent with federal rules permitting the U.S. Department of Education and its designated agents to collect student demographic and existing achievement data from schools and districts without prior parental or student consent (Family Educational and Rights and Privacy Act (FERPA) (20 U.S.C. 1232g; 34 CFR Part 99)). To minimize burden on district staff, we will provide a secure website to upload administrative records. To minimize burden of other data collection activities on school staff, our team's field staff will be available to go to schools to pick up hard copies of completed data collection instruments.

- **Secure school engagement.** We will follow the districts' lead on how to best engage their schools in this study. To help ensure a smooth data collection effort at the school-level, we will work with the school to appoint one school staff member as the on-site coordinator. The research and field staff will work closely with the on-site coordinator to schedule the site visits, secure space for conducting focus groups, disseminate and collect study forms and surveys, and follow-up with non-respondents. We will identify how best to handle these complex logistics in a way that is least burdensome for the school and its staff.

- **Secure teacher engagement.** We will provide teachers with professionally designed flyers about the study and the IEP_M program. In addition, a one-hour orientation session will give teachers the opportunity to ask questions in person and fully understand what is involved in participating in the study. Teachers also will receive advance notification about upcoming visits and a toll-free number to ask any questions.

- **Use tiered incentives to secure parental consent.** As described in Part A, Section A9, we will provide teachers (one per classroom) with incentives to ensure we obtain a high rate of returned parental consent forms. Teachers will receive $25 to help us distribute and collect consent forms. We will provide an additional $25 to teachers in classrooms with at least an 85 percent return rate on the consent form (whether or not the parents agree to participate in

the study). We have successfully used tiered incentives to boost return rates on consent forms on multiple studies, including the Impact Study of Feedback for Teachers Based on Classroom Videos. We plan to offer respondents gifts of appreciation for other data collection activities as well.

- **Use an efficient and flexible approach to completing teacher-child reports.** Our plan to divide the teacher-child reports between the lead teacher and the assistant teacher will help reduce the burden on any one teacher and result in a higher response rate. We also will create separate teacher-child report packets for each child and be flexible in how teachers would like to receive them (for example, in batches or all at once). Our team has obtained high response rates by staggering teacher-child report packets in batches. We plan to ask teachers for their preference and accommodate their request. The field staff will have extra copies of the teacher-child reports available on-site should teachers need them and will also work with the on-site coordinator at each school to address any questions as they arise. The coordinator will be available to pick up completed packets and local field staff will be available to collect the completed materials and return them to Mathematica.

- **Use a flexible and sensitive approach to conducting child assessments.** We will work closely with the on-site coordinator and classroom teachers to schedule and conduct the child assessments in a manner that is the least disruptive as possible to the class. The assessments will be conducted one-on-one with a trained data collector and there will be multiple short breaks during the assessment to help the child stay engaged and on track. Whatever adaptations or adjustments that are needed to accommodate children with disabilities will also be practiced.

- **Use an in-person approach to encourage completion of the teacher background and experiences survey.** Because teachers will receive full information on study commitments and we will use several methods to secure teacher engagement, we anticipate high levels of cooperation. To ensure the completion of surveys, the field staff will distribute the hard copy survey to the teachers at the start of the visit while introducing themselves and explaining what the visit will entail. The field staff will be available to answer any questions, follow up with the teachers, and collect the completed surveys at the end of the site visit. Field staff will work with the on-site coordinator to collect any missing surveys after the visit week.

B4. Tests of procedures or methods to be undertaken

We selected data collection instruments and measures that have been used with populations similar to the EPSEP sample. The proposed observations, teacher-child reports, and child assessments have good psychometric properties and have been used with similar populations, including preschool teachers and preschool-aged children with and without disabilities. Because these data collection instruments are standardized, the content cannot be altered. However, we will pre-test the protocol for the classroom observation where we will be combining two observation instruments (the Inclusive Classroom Profile and the Engagement Check) to confirm that there are no unforeseen difficulties in conducting these two observations simultaneously. Likewise, we pre-tested the child assessment and teacher-child reports to confirm our burden estimates and that our procedures and instructions were clear.

We also pre-tested the teacher background and experiences survey which has been developed for this study and designed to place as little burden as possible on respondents. Based

on the pre-test, we confirmed that the length was as expected, and made some minor revisions to question wording to ensure that questions are understandable, use language familiar to respondents, and are consistent with the concepts they aim to measure.

B5. Individuals consulted on statistical aspects of the design and on collecting and/or analyzing data

The people listed here worked closely in developing the survey instruments and will have primary responsibility for the data collection and analysis. Contact information for these people (including content experts serving as consultants to Mathematica) is provided below.

Cheri Vogel, Ph.D.
Project director
cvogel@mathematica-mpr.com
(609) 716-4546

John Deke, Ph.D.
Co-principal investigator
jdeke@mathematica-mpr.com
(609) 275-2230

Margaret Burchinal, Ph.D.
Co-principal investigator
burchinal@unc.edu
(919) 966-5059

Patricia Snyder, Ph.D.
Co-principal investigator
patriciasnyder@coe.ufl.edu
(352) 273-4291

Stephen Lipscomb, Ph.D.
Deputy project director
Slipscomb@mathematica-mpr.com
(617) 674-8371

Laura Kalb, B.A.
Survey director
Lkalb@mathematica-mpr.com
(617) 301-8989

Barbara Carlson, M.A.
Statistician
bcarlson@mathematica-mrp.com
(617) 674-8372

Harshini Shah, Ph.D.
Deputy survey director
hshah@mathematica-mpr.com
(617) 674-8360

REFERENCES

Doksum, K. "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case." *Annals of Statistics*, vol. 2, 1974, pp. 267–277.

Firpo, Sergio. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica*, vol. 75, no. 1, 2007, pp. 259–276.

Greenland, Sander, et al. "Statistical Tests, *P* Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology*, vol. 31, no. 4, 2016, pp. 337–350.

Hemmeter, M.L., P.A. Snyder, L. Fox, and J. Algina. "Evaluating the Implementation of the Pyramid Model for Promoting Social-Emotional Competence in Early Childhood Classrooms." *Topics in Early Childhood Special Education*, vol. 36, no. 3, 2016, pp. 133–146.

Koenker, Roger and Gilbert Bassett. "Regression Quantiles." *Econometrica*, vol. 46, no. 1, 1978, pp. 33–50.

Lee, E.C., A.L. Whitehead, R.M. Jacques, and S.A. Julious. The Statistical Interpretation of Pilot Trials: Should Significance Thresholds Be Reconsidered? *BMC Medical Research Methodology*, vol. 14, no. 41, 2014. doi: 10.1186/1471-2288-14-41.

Lehmann, E. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden-Day, 1974.

Neyman, J., and E.S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society*, Series A, Containing Papers of a Mathematical or Physical Character, vol. 231, 1933, pp. 289-337.

Schochet, Peter Z., Mike Puma, and John Deke. "Understanding Variation in Treatment Effects in Education Impact Evaluations: An Overview of Quantitative Methods." Report submitted to the Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Princeton, NJ: Mathematica Policy Research, April 2014.

Strain, P.S., and E. Bovey. "Randomized Controlled Trial of the LEAP Model of Early Intervention for Young Children with Autism Spectrum Disorders." *Topics in Early Childhood Special Education*, vol. 31, no. 3, 2011, pp.133–154.

Wasserstein, Ronald L., and Nicole A. Lazar. "The ASA's Statement on *p*-Values: Context, Process, and Purpose." *American Statistician*, vol. 70, no. 2, 2016, pp. 129–133.

Westlund, Erik, and Elizabeth A. Stuart. "The Nonuse, Misuse, and Proper Use of Pilot Studies in Experimental Evaluation Research." *American Journal of Evaluation*, vol. 38, no. 2, 2017, pp. 246-261.

MATHEMATICA
Policy Research