MATHEMATICA
Policy Research

Impact Evaluation of Departmentalized Instruction in Elementary Schools

Part B: Collection of Information Employing Statistical Methods

March 8, 2018

Submitted to:

U.S. Department of Education
National Center for Education Evaluation
Institute of Education Sciences
550 12th Street, S.W.
Washington, DC 20202
Project Officer: Tom Wei
Contract Number: ED-IES-17-C-0064

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Alison Wellington
Reference Number: 50533

CONTENTS

APPENDICES

TABLES

PART B. SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This package requests clearance for data collection activities to support an evaluation of departmentalized instruction in elementary schools. The Institute of Education Sciences (IES), National Center for Education Evaluation and Regional Assistance, U.S. Department of Education (ED) has contracted with Mathematica Policy Research, Inc. (Mathematica) and its partners (Public Impact; Clowder Consulting, LLC; Social Policy Research Associates; and IRIS Connect) to conduct this evaluation.

By the upper elementary grades, low-income students' achievement lags several years behind that of higher-income students (Duncan and Magnuson 2011). Departmentalized instruction, where each teacher specializes in teaching one subject to multiple classes of students instead of teaching all subjects to a single class of students (self-contained instruction), has recently become more popular as an improvement strategy in elementary schools. This strategy, which secondary schools already use almost universally, holds promise for several reasons. Many teachers are, to some degree, more effective at teaching particular subjects (Condie et al. 2014; Fox 2016; Goldhaber et al. 2013). Assigning teachers to those subjects could raise student achievement. It also allows teachers to concentrate planning on fewer subjects, which may lead to more thoughtful lessons and deeper instructional or content knowledge in those subjects (Chan and Jarman 2004). However, some experts worry that departmentalization could harm struggling students, particularly low-income students, by compromising student-teacher relationships (McPartland and Braddock 1993). In particular, teaching more students may make teachers less aware of each student's needs; having more teachers may make students feel less connected to each teacher. These factors could decrease student achievement, offsetting any gains from being taught by teachers who are more effective in the subjects they teach.

Despite concerns about departmentalization in elementary grades, elementary schools are increasingly adopting it. The percentage of elementary teachers in departmentalized settings more than doubled over a recent 12-year period, from 6 percent in 1999–2000 to 15 percent in 2011–2012 (U.S. Department of Education [ED] 2009; Goldring et al. 2013).

Currently, virtually no evidence exists on the effectiveness of departmentalized instruction relative to the more traditional self-contained approach to instruction. Given the increased use of departmentalization and numerous ways it might affect students, there is an urgent need for more evidence on its effects. This evaluation will help to fill the gap by examining whether departmentalizing fourth and fifth grade teachers improves teacher and student outcomes. The evaluation will focus on math and reading, with an emphasis on low-performing schools that serve a high percentage of disadvantaged students.

To help schools that are selected to transition to departmentalized instruction, the study team will provide implementation support. This support will include two design meetings before the start of the 2018–2019 school year to help schools determine the most effective structure for departmentalization and provide principals with advice on how to assign teachers to subjects. It will also include support calls to schools implementing departmentalized instruction throughout

the 2018–2019 and 2019–2020 school years to help them navigate any challenges related to departmentalization, as needed.

We will estimate the impact of departmentalized instruction in two different types of districts—those with teacher effectiveness measures based on student achievement growth and those without these measures. Impacts of departmentalized instruction could vary across these two sets of districts. In districts with these scores, principals can use the scores to determine teachers' relative effectiveness in reading and math and to assign teachers to the subjects they teach best. However, 63 percent of districts nationwide do not have teacher effectiveness measures based on student achievement growth (Troppe et al. 2017). These districts need evidence on whether principals, despite not having standardized information on teacher effectiveness in each subject, can accurately assess teachers' strengths in a way that enables departmentalization to succeed. The evaluation will provide evidence to guide decisions about departmentalization in both types of districts.

The evaluation will include implementation and impact analyses. The implementation analysis will describe schools' approaches to departmentalization and benefits and challenges encountered. The analysis will be based on information from school agreement forms, meetings to design each school's approach to departmentalization; monitoring and support calls; principal interviews; and teacher surveys. The impact analysis will draw on data from teacher surveys, videos of classroom instruction, principal interviews, and district administrative records to estimate the impact of departmentalized instruction on a range of outcomes. These outcomes include quality of instruction and student-teacher relationships, teacher satisfaction and retention, and student achievement and behavior.

Collection of information employing statistical methods

B1. Respondent universe and sampling methods

The evaluation will rely on a purposive sample of approximately 200 elementary schools from approximately 12 school districts from across the United States. The study will not statistically sample districts or schools, and thus we will not make statements that generalize beyond the districts and schools in the study. The 200 elementary schools will currently have self-contained classrooms in grades 4 and 5. We will group these 200 schools into pairs based on the similarity of their characteristics (such as average baseline school performance and socioeconomic status of students, as measured by free or reduced-price lunch receipt). We will then randomly assign schools within each pair to either implement departmentalized instruction in grades 4 and 5 (the treatment group) or to continue to use self-contained classrooms in these grades (the control group). Schools will remain in their assigned group for two school years (2018-2019 and 2019-2020).

B2. Procedures for the collection of information

a. Statistical methods for sample selection

The study will include a purposive sample of approximately 12 districts that together include about 200 schools (approximately 17 per district) that are eligible for and willing to participate in the study. Schools are eligible for the study if they contain fourth and fifth grades that are not currently departmentalized. We will also target schools that serve disadvantaged students (at

least 30 percent of students receiving free or reduced-price lunch) and have below-average performance (average proficiency rate in the bottom half of schools in the state). This will result in a purposive sample of districts that are willing to participate and schools from within those districts that are willing and eligible to participate. Although we will not be able to generalize to all schools, we will obtain valid estimates of the impact of departmentalizing instruction for a policy-relevant sample of schools that meet our eligibility requirements and are willing to participate. Below we explain in more detail how we will select districts, schools, and students for the study.

Selection of school districts. The 12 recruited school districts must together contain 200 schools that are eligible and willing to participate in the study. To identify districts likely to yield a sufficient number of eligible schools, we will target districts with at least 18 disadvantaged elementary schools with below-average performance. Data from the Common Core of Data and EDFacts suggest there are 168 such school districts. To help ensure that about half of the districts in the study have student growth scores for teachers, we will use information from the National Center on Teacher Quality supplemented by online research and our experience working with states and districts to classify districts into two groups, based on the availability of these scores. We will draw on information from other Mathematica studies to eliminate districts that we know are already mostly or fully departmentalized in fourth and fifth grades. We then focus on the largest remaining districts with and without growth scores (as these are the districts most likely to have a sufficient number of eligible schools), and screen out districts that cannot participate because they already departmentalize most or all fourth and fifth grade classrooms. We will recruit suitable districts until we reach our sample size target of 12 districts, about half of which will have student growth scores for teachers.

Selection of schools. Within the participating districts we will invite eligible schools to participate in the study. We will include 200 elementary schools with fourth and fifth grades for the 2018–2019 school year. Schools will be randomly assigned to the treatment or control group as described in section B1 above.

Selection of students. We will include all fourth and fifth grade students enrolled in the schools participating in the study. The study team will have access to administrative data on student characteristics and test scores through a Memorandum of Understanding (MOU) established with each participating district. Additionally, the study team will request parent consent for students to be included in video recordings of the study classrooms (Appendix E).

b. Data collection

This study includes multiple data collection efforts, including the activities summarized here.

Principal interview. The study team will conduct interviews in spring 2019 with principals of study schools to collect standardized information on factors principals considered when deciding teachers' subject (if appropriate) and grade assignments, teachers' communication with parents, and the schools' and teachers' handling of disciplinary issues (Appendix A). We will also ask principals of treatment schools about their perceptions of the challenges and benefits of departmentalization.

Teacher survey. A thirty-minute, web-based teacher survey will collect information about treatment and control teachers' time devoted to instruction, planning, and professional development, as well as their opportunities to coordinate with other teachers, and their perceptions of the successes and challenges related to planning and providing instruction and building relationships with students and parents (Appendix B). The survey will also measure teacher satisfaction and confidence in their teaching and level of awareness of student learning styles.

District administrative records. The study team will collect district administrative data on teacher effectiveness, student records, and teacher assignments (Appendix C). We will use information on teachers' effectiveness in math and reading from the 2016–2017 school year to examine the impact of departmentalization in districts that do and do not have teacher effectiveness measures. To estimate the impact of departmentalized instruction on student achievement and behavior, we will collect district administrative data on students' test scores in reading and math, as well as data on student attendance and disciplinary incidents. We will use district administrative data on teachers' school assignments to estimate the impact of departmentalized instruction on teacher retention.

Videos of classroom instruction. To measure the quality of instruction and teacher-student interactions, the study team will video-record, on average, two 30-minute lessons of fourth grade classes selected to be video-recorded. Study team videographers will record and upload the videos, and study team members will rate the videos using the Classroom Assessment Scoring System (CLASS) observation instrument. The CLASS measures the quality of student-teacher interactions, is valid and reliable (Kane and Staiger 2012; Pianta et al. 2012), and has strong procedures for training raters (Pianta et al. 2012). It is also suitable for teachers in multiple subjects.

c. Estimation procedures

The evaluation will include three broad sets of analyses: (1) impact analyses, estimating the effect of departmentalized instruction on student and teacher outcomes; (2) subgroup analyses, estimating the effects of departmentalized instruction on various subgroups of interest; and (3) implementation analyses, to learn about study schools' experiences and challenges implementing departmentalized instruction.

Impact analyses. We will estimate the impact of departmentalized instruction after the first and second years of implementation, using regression models to compare the outcomes of students and teachers in schools randomly assigned to departmentalize instruction and those assigned to the control group.

Key outcomes of interest for the impact analysis include:

- Students' reading and math achievement
- Student behaviors, including attendance and disciplinary incidents
- Teachers' instructional planning and professional development
- Teachers' instructional practices

- Teachers' job satisfaction and confidence in their teaching abilities
- Teacher retention (overall and for higher- and lower-performing teachers)
- The quality of student-teacher relationships

To estimate impacts on student achievement and teachers' outcomes, we will use the following regression model:

(1) $y_{isbd} = \beta T_{isbd} + \gamma X_{isbd} + \delta Z_b + \epsilon_{isbd}$,

where y_{isbd} is the outcome for individual i (either teacher or student) in school s, block b, and district d; X_{isbd} is a set of student-, teacher-, and school-level covariates; Z_b is a set of indicators for the study's random assignment blocks (matched pairs of schools); T_{isbd} indicates whether the school was assigned to departmentalize instruction; ϵ_{isbd} is an individual-level error term; and δ and γ are parameter vectors. The coefficient β represents the average impact of departmentalized instruction. The baseline characteristics in X_{isbd} will include:

- (for student-level outcomes) student characteristics, such as test scores from the year before the intervention, gender, race/ethnicity, free or reduced-price lunch eligibility, special education status, and English learner status
- (for teacher-level outcomes) teacher characteristics, such as demographic characteristics, age, experience, and educational background
- (for both student- and teacher-level outcomes) school-level characteristics, such as school-level student achievement and demographics.

When estimating student achievement models, the outcome of interest will be a student's state standardized test score in reading or math. For comparability across states, we will convert state test scores to z-scores, subtracting off the mean and dividing by the standard deviation of scores for all students in that state and grade level. To estimate impacts on teacher-level outcomes, such as teachers' practices, and survey responses, we will estimate a similar model at the teacher level. In each analysis, we will weight schools equally and cluster standard errors at the school level.

Subgroup analyses. To help districts and schools decide whether to switch from self-contained classrooms to departmentalized instruction, it can be valuable to know if the impact of departmentalization differs when principals have access to teachers' math and reading effectiveness scores and when they do not. Districts and schools may also want to know whether the impact of departmentalization varies for different types of students. For example, if departmentalization has a positive impact on high-achieving students, but a negative impact on low-achieving students, schools with many low-achieving students may decide not to implement departmentalization.

We will estimate impacts for various subgroups, including:

- Districts that do and do not have teacher effectiveness measures
- Students with high and low pre-intervention achievement
- Students who are and are not eligible for free and reduced-price lunch
- Special education students

Implementation analyses. Understanding the implementation experiences and challenges of schools selected to departmentalize instruction for the study will provide important information for other districts and schools considering departmentalizing instruction in upper elementary grades. The implementation analyses will support replication of study schools' approaches to departmentalized instruction in other districts and provide important context for interpreting the impact results.

Our implementation analysis will describe schools' approaches to departmentalization and benefits and challenges encountered, from the perspective of both teachers and principals. We will document the structure of departmentalization in treatment schools, how schools assigned teachers to subjects, and any implementation challenges. In both treatment and control schools, we will document time for instruction, planning, and teacher professional development.

d. Degree of accuracy needed

We estimate that the targeted sample sizes for the study will achieve a minimum detectable effect size of 0.08 standard deviations on student achievement, 0.33 standard deviations on teacher classroom observation scores, 11 percentage points on teacher satisfaction outcomes based on responses from the teacher survey, and 8 percentage points on teacher retention. Using a 50 percent subsample of schools – such as for the subgroup analyses based on whether principals have access to effectiveness scores when they make teacher assignment decisions – the study will achieve minimum detectable effects of 0.11 standard deviations on student achievement, 0.47 standard deviations on teacher classroom observation scores, 15 percentage points on teacher satisfaction outcomes, and 11 percentage points on teacher retention.

These target minimum detectable effects represent meaningful and realistic impacts that balance policy relevance against the costs of data collection. Prior studies of teacher-focused interventions have found effect sizes larger than these. The minimum detectable effects for student achievement (0.08 to 0.11) are smaller than the impacts of pay-for-performance on students' math achievement in 5 out of 10 districts that participated in IES's Teacher Incentive Fund intervention (Wellington et al. 2016) and the 0.13 impact of being taught by a Teaching Fellows math teacher rather than a math teacher from a less selective alternative route into teaching (Clark et al. 2013). The minimum detectable effects for teacher observation scores (0.33 to 0.47) are within the range of impacts (0.29 to 0.61) for content-focused professional development interventions in several IES studies (Garet et al. 2008, 2016). The minimum detectable effects for teacher satisfaction (11 to 15 percentage points) are smaller than most of the impacts of departmentalization on teacher satisfaction found by Strohl et al. (2014). Finally, for teacher retention, our minimum detectable effects are smaller than the 11-percentage point impact that Clotfelter et al. (2008) found for a program providing small retention bonuses to North Carolina teachers in high-poverty schools. Our proposed sample sizes will be sufficient to detect impacts of these magnitudes.

Table B.1 displays minimum detectable effects for the full sample of schools as well as a 50 percent subsample. The full sample will include 200 schools, 100 in the treatment group and 100 in the control group. The study design will maximize power to detect impacts by matching schools with similar characteristics into blocks within each district. The characteristics used to match schools will include average school baseline performance and socioeconomic status of students (as measured by free or reduced-price lunch receipt). Since the fourth- and fifth-grade students in the study will have baseline test scores from third- and fourth-grade assessments, we will also use students' prior test scores as covariates in the impact analysis to increase statistical power.

The calculations in Table B.1 assume the following: (1) 80 percent power and a 5 percent significance level for a two-tailed test; (2) each school will have an average of 3 fourth grade teachers and 66 students; (3) 85 percent of teachers will respond to the survey and have classroom observation ratings; (4) the school-level intracluster correlation is 0.16 for student outcomes, 0.15 for teacher observation scores, 0.13 for teacher satisfaction, and 0.02 for teacher retention; (5) the percentages of the between-school and within-school variances explained by covariates are 80 and 40 percent for student test scores, 80 and 72 percent for classroom observation outcomes, 30 and 15 percent for teacher survey outcomes, and 20 and 10 percent for teacher retention; (6) 75 percent of control group teachers will feel satisfied with their jobs and 64 percent of these teachers will be retained across two years; and (7) reliability of classroom observations is 0.21. Assumptions on the clustering of outcomes and the explanatory power of covariates for the student analyses are based on data from five large random assignment education evaluations (Deke et al. 2010). Assumptions for the analyses of teacher observation scores come from studies with information on the reliability of the CLASS observation instrument (Raudenbush et al. 2011), the school-level intraclass correlation of classroom observation scores (Schochet 2011), and the explanatory power of covariates in a random assignment study of math professional development (Garet et al. 2016). Assumptions for the analyses of teacher satisfaction and teacher retention across two years come from a random assignment study of pay-for-performance for teachers (Wellington et al. 2016).

Table B.1. Minimum detectable effects with 200 study schools

		Minimum detectable effect	
Data source	Outcome	Full sample	50 percent sub-sample
District records	Students' reading and math test scores	0.08 SDs	0.11 SDs
Classroom observations	Teacher observation scores	0.33 SDs	0.47 SDs
Teacher survey	Percentage of teachers who felt satisfied about their jobs	11 percentage points	15 percentage points
District records	Teacher retention across 2 years	8 percentage points	11 percentage points

SD = standard deviation.

d. Unusual problems requiring specialized sampling procedures

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

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

In order to limit respondent burden as much as possible, we have carefully considered what the minimum amount of data is needed to answer the research questions and how to structure the data collection. For example, the teacher surveys, principal interviews, and classroom observations will be collected only once in spring 2019. In addition, we will request administrative data no more than once a year, and in some cases (for example, fall 2019) we will request multiple years of data within a single request to reduce the number of separate requests.

B3. Methods to maximize response rates and deal with nonresponse

The study will employ multiple strategies to maximize response rates while minimizing burden on respondents. These include: establishing positive relationships with respondents and school and district staff; sending letters to teachers to alert them to an upcoming request to complete the survey; providing survey instruments that are accessible in both web and mobile formats; scheduling calls with principals at a time that is convenient for them; accepting administrative data files in formats that are most convenient for districts; and establishing efficient and flexible scheduling for classroom observations. To reassure respondents on the confidentiality of the data they provide, we will include a statement on confidentiality and data collection requirements (Education Sciences Reform Act of 2002, Title I, Part E, Section 183) in all letters and data collection instruments. Finally, we will include a statement indicating that participation is voluntary, yet we will also emphasize the importance of each response for the study findings.

Because we will develop an MOU with each district specifying in detail all data requirements, we anticipate full district participation for administrative records and their support for teacher participation. To further solidify administrators' cooperation, we will adhere to additional data collection requirements that districts may have such as preparing research applications and providing documentation of institutional review board (IRB) approvals. Reducing districts' burden in the submission of study data will facilitate attaining a response rate of at least 85 percent on student records and educator administrative data. Federal rules permit ED 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 maximize the response rate and minimize burden on schools and parents, we will follow these federal rules.

Based on Mathematica's experience conducting surveys with teachers, we expect at least an 85 percent response rate for the teacher survey. Because teachers will receive full information on the study (including the purpose of the requested information, how it will be used, and how we will maintain confidentiality of all data collected), we anticipate high levels of cooperation. To maximize completion of surveys, we will take the following steps. We will send teachers an invitation letter both by mail and email with a link to the web-based survey. In previous studies in similar settings, we have found that some teachers do not check school email accounts frequently. Therefore, we will include the web link to the survey in their invitation letters and also give teachers the option of completing a hard-copy survey, which will be mailed to them at their schools. Over a 12-week data collection period, we will send teachers email and mail reminders (see Appendix B). We plan to offer $30 to teachers who complete the teacher survey, which will take no more than 30 minutes to complete. We will also coordinate in-person school

visits with our field staff during the last four weeks of data collection to provide teachers with a hard-copy version of the teacher survey. This in-person connection has helped motivate teachers to participate in past surveys.

We have taken or will take a number of steps to ensure high-quality survey data. First, we have used many items that have been successfully used in other federal studies. Second, we have pretested the teacher survey instrument for clarity, accuracy, length, flow, and wording. Based on the pretest, the instrument is estimated to take under 30 minutes to complete. Third, the web-based survey will not allow respondents to enter out-of-range or inconsistent responses, and data entry programs will also check for these errors. Fourth, for surveys that are completed on paper, trained quality-control staff will identify item nonresponse and reporting errors by checking for complete and reasonable answers as soon as a hard-copy questionnaire is received and follow up with respondents if problems are identified. Finally, weekly reviews of web survey data will allow us to identify potential errors and follow up with respondents prior to the end of data collection. We will be courteous but persistent when following up with participants who do not respond quickly.

Videos of classroom instruction will be conducted in spring 2019 and an 85 percent response rate is expected. To maximize teacher cooperation, we will communicate the procedures to be used and obtain parent consent for students in their classrooms. Study field staff will visit the school with consent forms and work with teachers to develop student rosters for each class chosen to be recorded. The study team will use these rosters to accurately track receipt of parent consent. Forms will be returned to teachers, and the study plans to offer teachers $25 for collecting the consents and an additional $25 in active consent districts if they are able to get at least 85 percent of their students or parents to return forms (regardless of how students or parents respond).

Additionally, the study will work with district, school, and teacher schedules to avoid conflicts with testing and other planned activities. Appointments will be made for the video recordings, and well-trained videographers will record the classrooms. They will be instructed to set up the video equipment ahead of instructional time or during transition periods to minimize any disruption to student instruction. They will have a list of which students' parents provided permission to be recorded (and which did not) and they will be trained to seat children without permission outside of the view of the camera. We will communicate several key points to teachers and parents including (a) the purpose of the classroom video recordings and how they will be used for research purposes only, (b) the protections that are in place to ensure that the videos are only accessible to the study team, and (c) that the videos will be destroyed at the end of the study.

B4. Tests of procedures or methods to be undertaken

As much as possible, the data collection instruments for the study draw on surveys, forms, and protocols that have been used successfully in previous federal studies. For example, the teacher survey was modeled on instruments used in previous studies, such as the Impact Evaluation of Teacher Preparation Models and the Impact Evaluation of the Teacher Incentive Fund.

We pretested the teacher survey and principal interview protocol. The purpose of the pretests was to identify problems that study respondents might have providing the requested information and to confirm the level of burden. For example, the pretests assessed the content and wording of individual questions, organization and format of the instruments, the amount of time it took to respond, and potential sources of response error.

The teacher survey was pretested with eight fourth-grade elementary school teachers across eight districts, including teachers using either departmentalized or self-contained instruction. We sent a full survey packet to pretest respondents and asked them to complete the survey and to return completed forms by mail. The study team reviewed the completed surveys and conducted debriefing interviews, by phone, with each respondent to review problems teachers may have encountered. Interviewers followed a protocol to probe on a number of items to be sure the survey questions were communicated clearly and collected accurate information. Respondent burden to complete the survey averaged 25 minutes and ranged from 14 to 30 minutes as reported by pretest respondents. The results of the pretest were used to revise and improve the survey instrument.

The principal interview protocol was pretested, by phone, with six principals across five districts. Some principals were from schools that use departmentalized instruction, and some were from schools that use self-contained instruction. We first asked the principal the interview questions. After completing the interviews, we immediately conducted a debrief interview with the principal to probe on the clarity and relevance of the interview questions. Respondent burden to complete the interview averaged 33 minutes, ranging from 24 to 33 minutes and with a median of 31.5 minutes. The principal interview pretest findings were used to revise and streamline the interview protocol.

We did not pretest the school records data request or parent consent forms as both were closely modeled on forms that have been effectively used for other studies, such as the Impact of Teacher Feedback using Classroom Videos and the Impact Evaluation of the Teacher Incentive Fund. The classroom observation rubric (CLASS) was also not pretested in light of its established validity and reliability (Kane and Staiger 2012; Pianta et al. 2012).

The study will provide a help desk for questions, and our field staff will be available to answer questions throughout the data collection period. Staff will be trained to respond to frequently asked questions about the study and individual forms, so they can provide technical assistance and report any issues that come up in the field.

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

The following individuals were consulted on the statistical aspects of the study:

Table B.2. Individuals consulted on statistical design

Name	Title	Telephone Number
Alison Wellington	Senior Researcher, Mathematica	202-484-4696
Hanley Chiang	Senior Researcher, Mathematica	617-674-8374

Name	Title	Telephone Number
Melissa Clark	Senior Researcher and Deputy Director of Education Research, Mathematica	609-750-3193
Mariesa Herrmann	Senior Researcher, Mathematica	609-716-4544
Paul Burkander	Researcher, Mathematica	609-945-6625
Susanne James-Burdumy	Senior Fellow and Director of Education Research, Mathematica	609-275-2248

The following individuals will be responsible for data collection and analysis:

Table B.3. Individuals responsible for data collection and analysis

Name	Title	Telephone number
Tim Bruursema	Survey Researcher, Mathematica	202-484-3097
Paul Burkander	Researcher, Mathematica	609-945-6625
Florence Chang	Director of Planning and Evaluation, Jefferson County Public Schools	502-485-3278
Hanley Chiang	Senior Researcher, Mathematica	617-674-8374
Melissa Clark	Senior Researcher and Deputy Director of Education Research, Mathematica	609-750-3193
Megan Davis-Christianson	Lead Program Analyst, Mathematica	609-275-2361
Sarah Crissey	Survey Researcher, Mathematica	510-285-4640
Sheila Heaviside	Senior Survey Researcher, Mathematica	202-484-3096
Mariesa Herrmann	Senior Researcher, Mathematica	609-716-4544
Libby Makowsky	Researcher, Mathematica	734-794-8026
Catherine McClellan	Principal Scientist, Clowder Consulting	609-915-6676
Bryce Onaran	Survey Researcher, Mathematica	202-484-4524
Alison Wellington	Senior Researcher, Mathematica	202-484-4696

REFERENCES

Chan, Tak Cheung, and Delbert Jarman. "Departmentalize Elementary Schools." *Principal*, vol. 84, no. 1, 2004, p. 70.

Clark, Melissa, Hanley Chiang, Tim Silva, Sheena McConnell, Anastasia Erbe, Michael Puma, and Kathy Sonnenfeld. *The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs.* (NCEE 2013-4015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2013.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina." *Journal of Public Economics*, vol. 92, nos. 5–6, 2008, pp. 1352–1370.

Condie, Scott, Lars Lefgren, and David Sims. "Teacher Heterogeneity, Value-Added and Education Policy." *Economics of Education Review*, vol. 40, 2014, pp. 76–92.

Deke, John, Lisa Dragoset, and Ravaris Moore. *Precision Gains from Publically Available School Proficiency Measures Compared to Study-Collected Test Scores in Education Cluster-Randomized Trials.* (NCEE 2010-4003). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2010.

Duncan, Greg J., and Katherine Magnuson. "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances,* edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation, 2011.

Fox, Lindsay. "Playing to Teachers' Strengths: Using Multiple Measures of Teacher Effectiveness to Improve Teacher Assignments." *Education Finance and Policy*, vol. 11, no. 1, 2016, pp. 70–96.

Garet, Michael S., Stephanie Cronen, Marian Eaton, Anja Kurki, Meredith Ludwig, Wehmah Jones, Kazuaki Uekawa, Audrey Falk, Howard Bloom, Fred Doolittle, Pei Zhu, and Laura Sztejnberg. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement.* (NCEE 2008-4030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2008.

Garet, Michael S., Jessica B. Heppen, Kirk Walters, Julia Parkinson, Toni M. Smith, Mengli Song, Rachel Garrett, Rui Yang, and Geoffrey D. Borman. *Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development.* (NCEE 2016-4010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2016.

Goldhaber, Dan, James Cowan, and Joe Walch. "Is a Good Elementary Teacher Always Good? Assessing Teacher Performance Estimates Across Subjects." *Economics of Education Review*, vol. 36, 2013, pp. 216–228.

Goldring, Rebecca, Lucinda Gray, and Amy Bitterman. *Characteristics of Public and Private Elementary and Secondary School Teachers in the United States: Results from the 2011–12 Schools and Staffing Survey.* (NCES 2013-314). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2013.

Kane, Thomas J., and Douglas O. Staiger. "Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains." Research Paper. MET Project. Seattle, WA: Bill & Melinda Gates Foundation, 2012.

McPartland, James, and Jomills Henry Braddock II. "A Conceptual Framework on Learning Environments and Student Motivation for Language Minority and Other Underserved Populations." *Proceedings of the Third National Research Symposium on Limited English Proficient Student Issues.* Washington, DC: National Clearinghouse for Bilingual Education, George Washington University, 1993.

National Council on Teacher Quality. *Running in Place: How New Teacher Evaluations Fail to Live Up to Promises.* Washington, DC: National Council on Teacher Quality, January 2017.

Pianta, Robert C., Bridget K. Hamre, and Susan Mintz. *Upper Elementary and Secondary CLASS Technical Manual*. Charlottesville, VA: Teachstone, 2012.

Raudenbush, Stephen W., Andres Martinez, Howard Bloom, Pei Zhu, and Fen Lin. "Studying the Reliability of Group-Level Measures with Implications for Statistical Power: A Six-Step Paradigm." Unpublished manuscript. Chicago: University of Chicago, 2011.

Schochet, Peter Z. "Do Typical RCTs of Education Interventions Have Sufficient Statistical Power for Linking Impacts on Teacher Practice and Student Achievement Outcomes?" *Journal of Educational and Behavioral Statistics*, vol. 36, no. 4, 2011, pp. 441–471.

Strohl, Alecia, Lorraine Schmertzing, Richard Schmertzing, and E-Ling Hsiao. "Comparison of Self-Contained and Departmentalized Elementary Teachers' Perceptions of Classroom Structure and Job Satisfaction." *Journal of Studies in Education*, vol. 4, no. 1, 2014, pp. 109–127.

Troppe, Patricia, Anthony T. Milanowski, Camilla Heid, Brian Gill, and Christine Ross. *Implementation of Title I and Title II-A Program Initiatives: Results from 2013–14.* (NCEE 2017-4014). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2017.

U.S. Department of Education. *Schools and Staffing Survey (SASS): Number and Percentage of Regular Full-Time Public School Teachers of Grades 1 through 4, by Classroom Organization: Selected Years 1987–88 through 2007–08.* Washington, DC: U.S. Department of Education, 2009. Available at: https://nces.ed.gov/surveys/sass/tables/sass0708_042_t1n.asp. Accessed June 19, 2017.

Wellington, Alison, Hanley Chiang, Kristin Hallgren, Cecilia Speroni, Mariesa Herrmann, and Paul Burkander. *Evaluation of the Teacher Incentive Fund: Implementation and Impacts of Pay-for-Performance After Three Years.* (NCEE 2016-4004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2016.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research