

REPORT

Impact Evaluation of Departmentalized
Instruction in Elementary Schools

Part B: Collection of Information Employing
Statistical Methods

March 24, 2021

CONTENTS

PART B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS.....	1
B1. Respondent universe and sample design.....	1
B2. Information collection procedures.....	1
B3. Methods for maximizing the response rate.....	5
B4. Tests of procedures.....	5
B5. Individuals consulted on statistical aspects of design.....	5
REFERENCES.....	7

APPENDICES

APPENDIX A: ADMINISTRATIVE RECORDS DATA REQUEST

APPENDIX B: CONFIDENTIALITY PLEDGE FOR MATHEMATICA EMPLOYEES

TABLES

B.1 Minimum detectable effects.....	4
B.2 Individuals consulted on statistical design.....	6

PART B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

Finding creative ways to redeploy existing teachers in the classroom may yield academic benefits to students at little cost. One such strategy is departmentalized instruction, where each teacher specializes in teaching certain subjects to multiple classes of students instead of teaching all subjects to a single class of students (self-contained instruction). While nearly ubiquitous in secondary schools, departmentalization has only recently become more popular in upper elementary grades and is an improvement strategy that low-performing elementary schools identified under the Every Student Succeeds Act (ESSA) may consider adopting. The Department, through its Institute of Education Sciences (IES), received OMB clearance in 2018 to collect information for an evaluation that will provide valuable evidence on the implementation and outcomes of teachers and students as they departmentalize in fourth and fifth grades.¹ However, the coronavirus pandemic during the 2019-20 and 2020-21 school years created substantial delays in data availability. The purpose of this new package is to request an extension of the original approved timeline so that it is possible to finish collecting the district administrative records needed for the evaluation.

The justification for this evaluation and the associated data collection was detailed in the [original approved Supporting Statement](#), and remains the same. The details of the statistical methods were also included in the original Supporting Statement Part B. A streamlined version is provided in this submission below, including updates to reflect the fact that the study sample has already been established.

B1. Respondent universe and sample design

The evaluation recruited a purposive sample of 90 elementary schools from 12 school districts from across the United States. The study did not statistically sample districts or schools, and thus will not make statements that generalize beyond the districts and schools in the study. The 90 elementary schools had self-contained classrooms in grades 4 and 5 at the time of recruitment. Approximately half of the schools implemented departmentalized instruction in grades 4 and 5 (the treatment group), while the other half continued to use self-contained classrooms in these grades (the comparison group) for the two-year duration of the evaluation.

B2. Information collection procedures

a. Statistical methods for sample selection

The study recruited a purposive sample of 12 districts that together include 90 schools that were eligible for and willing to participate in the study. Schools were eligible for the study if they contained fourth and fifth grades that were not departmentalized at the time of recruitment. All fourth and fifth grade students enrolled in the schools participating in the study were included in the evaluation. The study team will be collecting administrative data on student characteristics and test scores, enabled through a Memorandum of Understanding (MOU) that was established with each participating district.

¹ OMB Control Number 1850-0942 (https://www.reginfo.gov/public/do/PRAViewICR?ref_nbr=201801-1850-001)

b. Data collection

This study includes multiple data collection activities, which were summarized in detail in the [original approved Supporting Statement](#). Importantly, this new package is requesting an extension to complete the collection of one key aspect of the data: district administrative records. These records will include information on students and their teachers (Appendix A). To estimate the outcomes related to departmentalized instruction, records will be collected on students' test scores in reading and math, as well as data on student attendance and disciplinary incidents. Similarly, district administrative data on teachers' school assignments will be used to examine teacher retention outcomes related to departmentalized instruction.

c. Estimation procedures

The evaluation will include three broad sets of analyses: (1) outcomes analyses comparing students and teachers using departmentalized instruction vs. self-contained instruction; (2) analogous outcomes analyses for key subgroups of interest; and (3) implementation analyses to learn about study schools' experiences and challenges implementing departmentalized instruction.

Key outcomes of interest from the administrative records data include:

- Students' reading and math achievement
- Student behaviors, including attendance and disciplinary incidents
- Teacher retention (overall and for higher- and lower-performing teachers)

The study will use the following general regression model:

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

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 blocks (matched pairs of schools); T_{isbd} indicates whether the school was assigned to departmentalize instruction; $\dot{\epsilon}_{isbd}$ 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, the study team 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. For teacher-level outcomes, a similar model will be used at the teacher level. In each analysis, schools will be weighted equally and standard errors clustered 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 evidence on 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 evidence on departmentalization varies for different types of students. For example, if departmentalization seems to have better outcomes for high-achieving students, but not low-achieving students, schools with many low-achieving students may decide not to implement departmentalization.

The study team will estimate these outcomes 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 outcomes analyses.

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

d. Degree of accuracy needed

The study team estimates that the realized sample size for the study (90 schools and 12 districts) will achieve an approximate minimum detectable effect size of 0.11 standard deviations on student achievement and 11 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 can achieve approximate minimum detectable effects of 0.16 standard deviations on student achievement and 16 percentage points on teacher retention.

These estimated minimum detectable effects represent meaningful and realistic impacts that balance policy relevance against the costs of data collection. The estimated minimum detectable effects for student achievement (0.11) are similar to 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). For teacher retention, the estimated minimum detectable effects are similar to 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.

Table B.1 displays minimum detectable effects for the full sample of schools as well as a 50 percent subsample. The study design will maximize validity and 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, the study team 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 has an average of 3 fourth grade teachers and 66 students; (3) the school-level intraclass correlation is 0.16 for student outcomes and 0.02 for teacher retention; (4) the percentages of the between-school and within-school variances explained by covariates are 80 and 40 percent for student test scores, and 20 and 10 percent for teacher retention; (5) 64 percent of comparison school teachers will be retained across two years. 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 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

Data source	Outcome	Minimum detectable effect	
		Full sample	50 percent sub-sample
District records	Students’ reading and math test scores	0.11 SDs	0.16 SDs
District records	Teacher retention across 2 years	11 percentage points	16 percentage points

SD = standard deviation.

e. Unusual problems requiring specialized sampling procedures

There were no unusual problems requiring specialized sampling procedures.

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

To limit respondent burden as much as possible, the evaluation has carefully considered what is the minimum amount of data needed to answer the research questions and how to structure the data collection. For example, the evaluation will request administrative data no more than once a year, and whenever possible, will request multiple years of data within a single request to reduce the number of separate requests.

B3. Methods for maximizing the response rate

The study will employ multiple strategies to maximize response rates while minimizing burden on respondents, including establishing positive relationships with district staff and accepting administrative data files in formats that are most convenient for districts. To reassure respondents on the confidentiality of the data they provide, the study team 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. All Mathematica staff have to sign a confidentiality pledge as well, to further protect respondent confidentiality (see Appendix B).

Because an MOU was signed with each district as part of the agreement to participate in the evaluation and because this MOU specified in detail all data requirements, the study team anticipates full district participation for administrative records. To further solidify administrators' cooperation, the study team adhered to any additional data collection requirements that districts may have had, 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 the Department 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, the study team will follow these federal rules.

B4. Tests of procedures

As much as possible, the data collection instruments for the study draw on forms that have been used successfully in previous federal studies. The study team did not pretest the district administrative records request, as the form was 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 study will provide a help desk for 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.

B5. Individuals consulted on statistical aspects of design

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

Table B.2. Individuals consulted on statistical design

Name	Title	Telephone Number
Alison Wellington	Principal Researcher, Mathematica	202-484-4696
Hanley Chiang	Principal Researcher, Mathematica	617-674-8374
Melissa Clark	Principal Researcher, Mathematica	609-750-3193
Mariesa Herrmann	Senior Researcher, Mathematica	609-716-4544
Paul Burkander	Researcher, Mathematica	609-945-6625
Susanne James-Burdumy	Vice President, Mathematica	609-275-2248

REFERENCES

- 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.
- 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.
- 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.