Addendum to the Evaluation of the Expository Reading and Writing Course

Anthony B. Fong
Neal D. Finkelstein

May 2016

WestEd.org

This report was supported by the Office of Innovation and Improvement within the U.S. Department of Education through Grant #U411C110425. The opinions expressed here are those of the authors and do not represent the views of the U.S. Department of Education.

WestEd — a nonpartisan, nonprofit research, development, and service agency — works with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont and Georgia, to Illinois, Arizona and California, with headquarters in San Francisco. For more information about WestEd, visit http://www.WestEd.org; call 415.565.3000 or, toll-free, (877) 4-WestEd.

.

Table of Contents

List of Tables

Introduction

In 2015, Fong, Finkelstein, Jaeger, Diaz, and Broek reported the findings from an independent evaluation of the Expository Reading and Writing Course (ERWC) that was funded by an Investing in Innovation development grant. The evaluation used a quasi-experimental design that matched grade-12 students who enrolled in the ERWC with grade-12 comparison students who did not enroll in the ERWC; comparison students were referred to as “non-ERWC students” in the report. The outcome measure used to evaluate the course’s impact on student achievement was the English Placement Test, which is a standardized test given to students who matriculate to a California State University to determine their eligibility for enrollment in a credit-bearing college English course. The evaluation found positive and statistically significant effects of the ERWC on student achievement.

In the matching analysis of the evaluation, the following student-level variables were used to match ERWC and non-ERWC students: grade-11 English language arts (ELA) California Standards Test (CST) scale score, grade-11 Advanced Placement (AP) English course enrollment, average grade-11 English grade earned, gender, and ethnicity. Key aspects of the matching process follow: the Mahalanobis distance metric was used as the measure of the degree of similarity between ERWC and non-ERWC students; each ERWC student was matched to the four most similar non-ERWC students; and matching was conducted with replacement so that a non-ERWC student could be used as the match for multiple ERWC students. After the matching was completed, all matched students were included in an ordinary least squares (OLS) regression analysis that used as covariates the same variables used in the matching process.

Because each ERWC student was matched to four non-ERWC students, in the OLS regression model each non-ERWC student received a weight of 0.25 for each time he or she was used as a match. This weighting scheme ensured that, collectively, the four matched non-ERWC students had the same weight as the one ERWC student with whom they were matched. If, for instance, a non-ERWC student had been matched to three different ERWC students, then that non-ERWC student received a weight of 0.75 (= 0.25 * 3) in the OLS regression. Thus, after the weights were applied, the sample size of each group (ERWC students and non-ERWC students) in the OLS regression analysis was the same.

Since each non-ERWC student could be included in the OLS regression analysis multiple times, cluster-robust standard errors were used to allow for intragroup correlation at the student level. Clustering the standard errors on the classroom or the teacher was not performed; the rationale for this analytic decision was because matching makes the assumption of unconfoundedness/conditional independence. As described in Heinrich, Maffioli, and Vazquez (2010), the assumption is that after controlling for the covariates, "the treatment assignment is 'as good as random'" (p. 16). Similarly, Imbens and Rubin (2015) note that the unconfoundedness assumption "requires that conditional on the pre-treatment variables the assignment is effectively random" (p. 567). Lastly, as Firpo (2007) describes, "the relevant restriction is the assumption that selection to treatment is based on observable variables (exogeneity assumption). In other words, it is assumed that given a set of observed covariates, individuals are randomly assigned either to the treatment group or to the control group" (p. 261). As a result, under the assumption of unconfoundedness, matching at the student level is equivalent to randomizing at the student level. And when randomization and analysis occur at the student level, the clustering of standard errors is not required since the level of assignment matches the level of analysis (What Works Clearinghouse, 2014).[1] As a result of the necessary assumption of unconfoundedness in matching studies, the What Works Clearinghouse (2014) allows quasi-experimental design studies to receive a rating no higher than "Meets WWC Group Design Standards with Reservations": "Randomized controlled trials with high attrition and all quasi-experimental designs are not eligible to receive the highest rating because of a greater concern about the similarity of the intervention and comparison groups" (p. 15).

However, due to some concerns that the previously reported results in Fong et al. (2015) did not account for the possibility that students who were taught by the same teacher may have correlated error terms, an additional OLS regression analysis was conducted using cluster-robust standard errors that allow for clustering on the teacher. The results of this additional analysis are reported in this addendum. The previously reported OLS regression results had used cluster-robust standard errors that allowed for clustering on the student. The newly reported regression analysis in this addendum only clusters on the teacher, and not on both the teacher and the student, since previous literature has recommended only clustering at the highest level (see, for instance, Cameron & Miller, 2015; Bertrand, Duflo, & Mullainathan, 2004).

---

[1] In March 2016, the What Works Clearinghouse released further guidance relating to cluster design standards (What Works Clearinghouse, 2016). That document explains that if 1) the unit of assignment is a cluster and 2) the data for the analysis are based on individuals within those clusters, then the design is considered a cluster-design study. If the study does not meet either of the above two criteria (e.g., a design where the unit of assignment is at the student level), then the design is an individual-level design.

Results of the Additional Analysis

This section presents the results of the OLS regression analysis that uses cluster-robust standard errors that allows for clustering on the teacher (table 1). For ease in comparing the previously reported results in Fong et al. (2015) with the results from this new analysis, the previously reported results are included in table 1 as Model 1. The new OLS regression analysis that uses cluster-robust standard errors that allows for clustering on the teacher is presented as Model 2. Other than the way in which the OLS regression model accounted for clustering, Model 1 and Model 2 are exactly the same. In other words, both OLS regression models use the same independent and dependent variables, the same sample of students, and the same frequency weights for the students.

Table 1. Additional Analysis that Accounts for Clustering on the Teacher

Characteristic	Model 1: Previously Reported Results (Standard Errors Clustered on the Student)	Model 2: Additional Analysis (Standard Errors Clustered on the Teacher)
ERWC enrollment	1.221*** (0.239)	1.221*** (0.403)
Female	-0.319 (0.240)	-0.319 (0.279)
Asian	0.965 (0.843)	0.965 (0.749)
Hispanic	0.032 (0.813)	0.032 (0.700)
White	1.516* (0.844)	1.516** (0.688)
Grade-11 ELA CST scale score	0.136*** (0.003)	0.136*** (0.003)
Average grade-11 English grade earned	1.365*** (0.130)	1.365*** (0.139)
Grade-11 AP English enrollment	2.533*** (0.287)	2.533*** (0.339)
Intercept	87.743*** (1.181)	87.743*** (1.115)

* denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

Notes: The estimated coefficients are provided, with the cluster-robust standard errors shown in parentheses. With respect to ethnicity, African Americans were the omitted category as a result of being the first group alphabetically.
Observations = 6,618
Sources: English Placement Test (spring 2014) data and student records data collected from the nine school districts in the study sample. See Fong et al. (2015) for additional details.

As shown in table 1, in both Model 1 and Model 2 the estimated impact of enrollment in the ERWC is positive and statistically significant at the 1 percent level. In other words, the results remain unchanged when the standard errors allow for clustering on the teacher as opposed to on the student. The coefficient on the ERWC enrollment is unchanged at 1.221, because only the cluster-robust standard error calculation differs between Model 1 and Model 2. With respect to the standard errors on the ERWC enrollment variable, the cluster-robust standard error in Model 1 that allows for clustering on the student is calculated to be 0.239; in comparison, the cluster-robust standard error in Model 2 that allows for clustering on the teacher is calculated to be 0.403. These results suggest that, whether clustering is accounted for on the student or on the teacher, the ERWC had a positive and statistically significant impact on student achievement. For additional details about the intervention, study design, analytic sample, and other results not reported here, the reader is referred to Fong et al. (2015).

References

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*(1), 249–275.

Cameron, A., & Miller, D. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica, 75*(1), 259–276.

Fong, A., Finkelstein, N., Jaeger, L., Diaz, R., & Broek, M. (2015). *Evaluation of the Expository Reading and Writing Course: Findings from the Investing in Innovation development grant.* San Francisco, CA: WestEd.

Heinrich, C., Maffioli, A., & Vazquez, G. (2010). *A primer for applying propensity score matching.* Office of Strategic Planning and Development Effectiveness, Impact-Evaluation Guidelines, Technical Notes, No. IDB-TN-161. Washington, DC: Inter-American Development Bank. Retrieved from http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35320229

Imbens. G., & Rubin, D. (2015). *Causal inference in statistics, social, and biomedical sciences.* New York, NY: Cambridge University Press.

What Works Clearinghouse. (2014). *WWC procedures and standards handbook: Version 3.0.* Washington, DC: United States Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

What Works Clearinghouse. (2016). *WWC cluster design standards*. Washington, DC: United States Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_cluster_standards_030416.pdf