



The Impact of TenMarks Math Premium: Evidence from Wisconsin

March 2017



Executive Summary

Purpose: This study evaluates the impact of providing TenMarks Math Premium on student achievement using evidence from the 2016 Wisconsin Forward Exam in mathematics.

Findings: Providing students with TenMarks Math Premium is associated with an average improvement of 5.4 percentage points in the proficiency rate on the 2016 Forward Exam in mathematics. The improvement in proficiency is equivalent to moving a group of students from the 53rd to the 64th percentile of classes in the state. The study also finds that providing students with TenMarks Math Premium is associated with an average improvement of 4.1 percentage points in the mastery rate. The improvement in mastery is equivalent to moving a group of students from the 69th to the 80th percentile of classes in the state. These findings are statistically significant and are based on a matched comparison research design that accounts for two years of prior achievement and a rich set of school-level characteristics (Figure 1). The study sample consists of 180 classes, representing approximately 7,400 students in grades 3 through 8.¹

Estimated Impact of TenMarks Math Premium

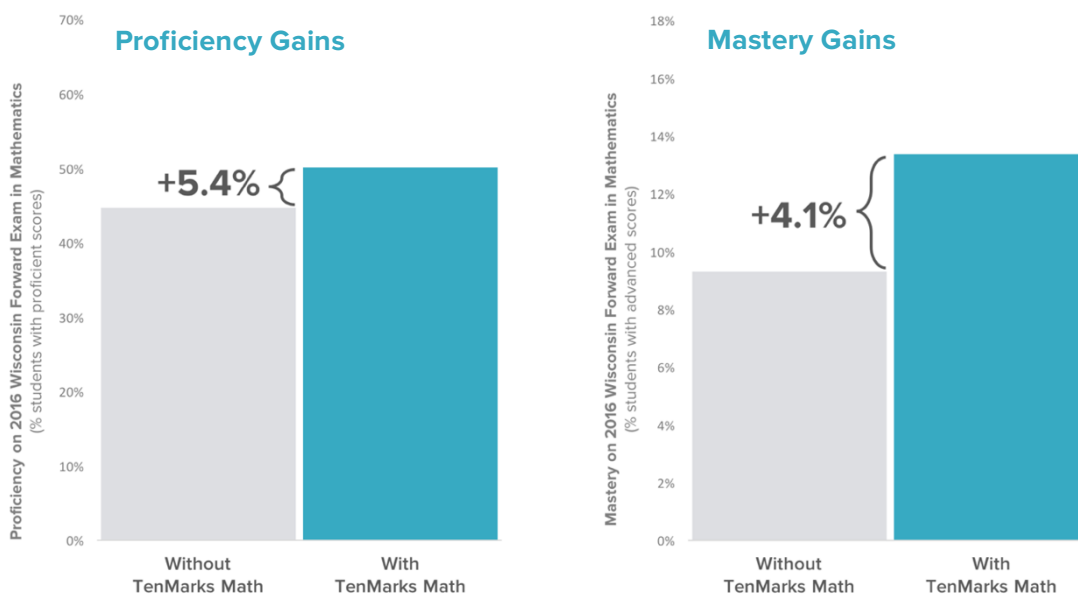


Figure 1. Estimated Impact of TenMarks Math Premium on Proficiency (left panel) and Mastery (right panel). The bars represent the percentage of students in a class reaching proficiency (left panel) or mastery (right panel) on the 2016 Forward Exam in mathematics. Within each panel, the blue bar on the right represents the TenMarks Math Premium group and the gray bar on the left represents the matched comparison group, statistically adjusted at the means of the TenMarks Math Premium group for all baseline characteristics. The difference between the bars in each panel (+5.4% for proficiency and +4.1% for mastery) represents the estimated impact, i.e., the average gains in achievement associated with providing TenMarks Math Premium during the 2015–16 school year. The impact estimates are statistically significant and are based on an analysis that accounts for two years of prior achievement in math and ELA as well as for school-level characteristics such as school size, student demographics, and the percentage of students eligible for a free or reduced-price lunch.

I. Research Design

This study evaluates the impact of providing TenMarks Math Premium on student achievement based on evidence from Wisconsin’s state assessments in mathematics and a matched comparison research design. The study uses a technique called *propensity score matching* to compare math achievement on the 2016 Forward Exam between classes provided with TenMarks Math Premium during the 2015–16 school year and a matched set of classes not provided with TenMarks Math Premium.¹ The study sample consists of 180 classes, representing approximately 7,400 students in grades 3 through 8.

The goal of the matching design is to make an apples-to-apples comparison between the TenMarks Math Premium and comparison groups. It does so by ensuring the two groups have “baseline equivalence” of observable characteristics, including two years of prior achievement (see the What Works Clearinghouse (2014) and the Technical Appendix of this document). Crucially, the matching design in this study ensures baseline equivalence in math and ELA performance for the two years prior to the study period. The matching design also ensures that the TenMarks Math Premium and comparison groups are apples-to-apples in terms of school characteristics, including total enrollment, student-teacher ratio, charter school status, magnet school status, percentage of students eligible for a free or reduced-price lunch, percentage of African American students, percentage of Hispanic students, and percentage of Asian students. A regression model was used to account for residual imbalance between the two groups and to increase the precision of the estimates. See Ho, Imai, King, & Stuart (2007), Rosenbaum and Rubin (1985), and the Technical Appendix.

The main limitation of the matching design is that it does not guarantee equivalence of unobserved characteristics such as curricular decisions or student motivation. Nevertheless, researchers have demonstrated that matched comparison designs such as the one employed in this study can provide unbiased, causal estimates of a program’s impact, especially when they take prior performance into account (e.g., Dehejia and Wahba 1999).

II. Findings

Impact on Proficiency

The estimated impact on proficiency associated with providing TenMarks Math Premium during the 2015–16 school year is a 5.4 percentage-point gain in the number of proficient students on the 2016 Forward Exam in math. The estimate is statistically significant at the .05 level and is based on a matched comparison of 180 classes, representing approximately 7,400 students (Table 1, Column 1).

The left panel of Figure 1 displays the impact estimate and provides a simple representation of the research design. The bars represent the average level of math proficiency on the 2016 Forward Exam for the TenMarks Math Premium group (50.1% proficient) and the matched comparison group (44.7% proficient), statistically adjusted at the means of the TenMarks Math Premium group for all baseline characteristics. Given that the two groups started from an equivalent baseline, including math performance during the prior two years, and given that the gains are statistically adjusted for a

¹ The study focuses on a paid version of TenMarks Math Premium that includes differentiated student assignments, assessments, and just-in-time interventions embedded into assignments. The unit of analysis in the study is a “class,” defined as the group of students in the same grade within a particular school, e.g., the 4th grade students at XYZ Elementary, the 5th grade students at XYZ Elementary, and so on.

substantial number of potentially confounding variables, the differential gains between the two groups (+5.4%) are interpreted as the impact of providing TenMarks Math Premium. That is, it's the additional improvement in the proficiency rate associated with providing TenMarks Math Premium.

How significant is a 5.4 percentage-point gain in proficiency? Imagine all classes in the state (the 4th grade at XYZ Elementary, the 7th grade at ABC Middle, and so on) lined up by math proficiency on the 2016 Forward Exam. The group of students in the middle of that line (the class at the 50th percentile in the state) had 42.9% of students demonstrating proficiency. The proficiency improvement associated with providing TenMarks Math Premium is about equivalent to moving a group of students from the 53rd to the 64th percentile in the state, ahead of the next 11% of classes in Wisconsin.

| | <u>Proficiency</u> | <u>Mastery</u> |
|-------------------------------------|--------------------|--------------------|
| Impact estimate (standard error) | 5.42 (2.27)* | 4.06 (1.51)** |
| Tested students (classes) | 7,400 (N = 180) | 6,900 (N = 176) |
| Grade levels | 3, 4, 5, 6, 7, 8 | 3, 4, 5, 6, 7, 8 |

Table 1. Impact Estimates of TenMarks Math Premium on Class Proficiency and Mastery Rates Notes: * significant at the .05 level; ** significant at the .01 level. Standard errors are cluster-adjusted to account for within-school correlations. Sample sizes include both the classes provided with TenMarks Math Premium and the matched classes. The number of tested students is rounded to the nearest hundred. All grade levels with sufficient data were included in the analysis.

Impact on Mastery

This section presents estimates for the impact of providing TenMarks Math Premium on the percentage of students scoring at the advanced level on the 2016 Forward Exam in mathematics. An “advanced” designation corresponds to the highest level of mastery of the Wisconsin State Standards. See Table 2 in the Technical Appendix for a description of the four mastery levels defined by the state.

The estimated impact on mastery associated with providing TenMarks Math Premium during the 2015–16 school year is a 4.1 percentage-point gain in the number of advanced students on the 2016 Forward Exam in math. The estimate is statistically significant at the .01 level and is based on a matched comparison of 176 classes, representing approximately 6,900 students (Table 1, Column 2).

The right panel of Figure 1 displays the impact estimate and provides a simple representation of the research design. The bars represent the average level of math mastery on the 2016 Forward Exam for the TenMarks Math Premium group (13.4% advanced) and the matched comparison group (9.3% advanced), statistically adjusted at the means of the TenMarks Math Premium group for all baseline characteristics. As above, the differential gains between the two groups (+4.1%) are interpreted as the impact of providing TenMarks Math Premium. That is, it's the additional improvement in the mastery rate associated with providing TenMarks Math Premium.

How significant is a 4.1 percentage-point gain in mastery? Imagine all classes in the state lined up, this time by math mastery on the 2016 Forward Exam. The group of students in the middle of that line (the class at the 50th percentile in the state in terms of mastery) had 5.0% of students demonstrating mastery (reaching the Advanced level of performance). The mastery improvement associated with providing TenMarks Math Premium is about equivalent to moving a group of students from the 69th to the 80th percentile in the state, ahead of the next 11% of classes in Wisconsin.

Assessing Baseline Equivalence

The What Works Clearinghouse (WWC) considers comparison groups as meeting baseline equivalence if the standardized mean difference (SMD) for pre-test measures is within 0.25 standard deviations and a statistical adjustment is applied in the analysis (WWC Procedures and Standards Handbook Version 3.0; WWC Review Protocol for Primary Mathematics Version 3.1). This study's analytic sample meets the baseline equivalence standard by ensuring the SMD for pre-test measures are well within 0.25 standard deviations. The study exceeds the WWC standard by also ensuring that the SMD for *every* school-level characteristic considered in the model is within 0.5 standard deviations: total enrollment, the student-teacher ratio, charter school status, magnet school status, percentage of students eligible for a free or reduced-price lunch, percentage of African American students, percentage of Hispanic students, and percentage of Asian students at the school level. Baseline equivalence was assessed separately for the proficiency and mastery analyses. Table 3 in the Technical Appendix provides the SMDs for two years of prior achievement in math and ELA as well as for the propensity scores.

III. Technical Appendix

A. Data Sources and Definitions

Wisconsin Forward Exam

The state of Wisconsin began administering the Wisconsin Forward Exam in Spring 2016 in grades 3–8 to assess student mastery of the Wisconsin State Standards in mathematics. The Forward Exam replaced the Badger Exam, a Smarter Balanced (SBAC) assessment, which was administered just once in Spring 2015. Prior to the Badger Exam, the state administered the Wisconsin Knowledge and Concepts Examination (WKCE).²

The Wisconsin Department of Public Instruction (WDPI) publishes data on student performance on the Forward Exam, the Badger Exam, and the WKCE at the class level. That is, the state reports student performance for each grade, within every school in the state (e.g., 4th grade students at XYZ Elementary, 5th grade students at XYZ Elementary, and so on). These data constitute the primary measure of student performance in the study's analyses.³

The WDPI defines four achievement levels, and considers proficiency as a passing score (see Table 2). A student's achievement level is determined by her scale score on the Forward Exam and a set of cut points set by the state. Consistent with the state's achievement levels, the measure of student

² <https://dpi.wi.gov/assessment/historical>

³ <http://dpi.wi.gov/wisedash/download-files>

proficiency used in this study is the percentage of proficient students in each class, and the measure of student mastery is the percentage of advanced students in each class.

| Below Basic | Basic | Proficient | Advanced |
|---|---|--|--|
| Student demonstrates minimal understanding of and ability to apply the knowledge and skills for his/her grade level that are associated with college content-readiness. | Student demonstrates partial understanding of and ability to apply the knowledge and skills for his/her grade level that are associated with college content-readiness. | Student demonstrates adequate understanding of and ability to apply the knowledge and skills for his/her grade level that are associated with college content-readiness. | Student demonstrates thorough understanding of and ability to apply the knowledge and skills for his/her grade level that are associated with college content-readiness. |

Table 2. Performance levels and descriptions for the Wisconsin Forward Exam. Student performance falls into one of four performance levels, based on score ranges determined by the state. Source (and additional information on the score ranges) at <http://dpi.wi.gov/assessment/forward/data#scores>.

TenMarks Math Premium

Consistent with the state’s reporting of student performance at the class level, the analyses in this study define TenMarks Math Premium provision at the class as well (e.g., for the 4th grade at XYZ Elementary, the 5th grade at XYZ elementary, and so on). A class is considered to have been provided with TenMarks Math Premium if TenMarks Math Premium licenses were provided to at least 85% of tested students in that class. On average, the TenMarks Math Premium students in the TenMarks Math Premium classes completed 2.0 assignments per week during the 2015–16 school year.

School-level Covariates

The analysis accounts for institutional, economic, and demographic factors at the school level that are potentially correlated with both TenMarks Math Premium provision and student achievement. School-level data are drawn from the most recent Public Elementary/Secondary School Universe Survey (2013–14) provided by the National Center for Education Statistics (NCES) Common Core of Data (CCD).⁴ The study uses the following school-level variables from the NCES CCD data set:

- **Institutional characteristics:** total enrollment, student-teacher ratio, charter school status, magnet school status
- **Economic characteristics:** percentage of students eligible for a free lunch, percentage of students eligible for a reduced-price lunch
- **Demographic characteristics:** percentage of African American students, percentage of Hispanic or Latino students, percentage of Asian students

B. Matching Design and Statistical Model

The study uses caliper matching that combines propensity score and Mahalanobis distance metrics, following Rosenbaum and Rubin (1985) and Rubin and Thomas (2000). The propensity score was estimated using a logistic regression on the full set of covariates, including both linear and quadratic

⁴ <https://nces.ed.gov/ccd/pubschuniv.asp> includes the datasets and variable definitions.

terms for the two years of prior achievement in math and ELA to allow for potential non-linear trends in achievement. Potential matches were identified as having an exact match by grade level and falling within a caliper of 0.25 standard deviations on the propensity score. Matches were then created using Mahalanobis distance matching on math achievement in 2014 and 2015 in order to improve balance on these covariates in particular, since they are considered prognostic (highly correlated with the outcome measure). Prior achievement is taken as the class proficiency for the same class the prior year (class mastery for the mastery analysis). That is, the analysis is of repeated cross-sections rather than of student cohorts. Matching was done 1:1 (one comparison unit for each TenMarks Math Premium unit), with replacement (the same comparison unit can be used more than once) and with ties broken deterministically, using the Matching package in R (Sekhon 2011). The balance statistics are reported in Table 3.

| | <u>Proficiency</u> | <u>Mastery</u> |
|--|--------------------|----------------|
| Prior math achievement (% proficient/advanced 2015) | -0.01 | 0.04 |
| Prior math achievement (% proficient/advanced 2014) | -0.01 | 0.01 |
| Prior ELA achievement (% proficient/advanced 2015) | -0.05 | 0.22 |
| Prior ELA achievement (% proficient/advanced 2014) | 0.13 | 0.12 |
| Propensity score | 0.04 | 0.01 |

Table 3. Balance Statistics for Prior Achievement and Propensity Scores. Balance statistics are reported as the standardized mean difference (SMD) between TenMarks Math Premium and the comparison group. The What Works Clearinghouse (WWC) considers comparison groups as meeting baseline equivalence if the standardized mean difference (SMD) for pre-test measures is within 0.25 standard deviations and a statistical adjustment is applied in the analysis (WWC Procedures and Standards Handbook Version 3.0; WWC Review Protocol for Primary Mathematics Version 3.1).

Impact on proficiency and mastery rates was estimated via an OLS regression using the covariates listed above to provide additional bias adjustments and improve the precision of the estimates. Standard errors were clustered at the school level to account for within-school correlation. Because the matching and regression-based adjustments account for two years of prior achievement, the impact estimates for achievement levels in 2016 can also be interpreted as growth estimates from the 2015 baseline or as estimates of acceleration from the historical trend starting in 2014. It should be emphasized, however, that prior achievement is based on a different set of assessments: the Badger Exam in 2015 and the WKCE in 2014.

Works Cited

- Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*(448), 1053–1062.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199–236.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38.
- Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*(450), 573–585.
- Sekhon, J. S. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R. *Journal of Statistical Software*. *42*(7): 1–52.
- What Works Clearinghouse, Procedures and Standards Handbook Version 3.0. (2014) Washington, DC: Institute for Education Sciences, U.S. Department of Education.
- What Works Clearinghouse, Review Protocol for Primary Mathematics, Version 3.1. (2015) Washington, DC: Institute for Education Sciences, U.S. Department of Education.