Predicting Students' Standardized Test Scores Using Online Homework

Mingyu Feng
SRI International
333 Ravenswood Ave
Menlo Park, CA 94025 USA
mingyu.feng@sri.com

Jeremy Roschelle
SRI International
333 Ravenswood Ave
Menlo Park, CA 94025 USA
Jeremy.roschelle@sri.com

Abstract
How students do homework has been underresearched relative to classroom learning because it is more difficult to collect data on students' homework behaviors. Presumably, such data would have implications for students' achievement. To understand how students do homework and how homework performance and behaviors relate to end-of-year standardized test scores, we analyzed the system logs from an online homework support platform used by more than 1,500 seventh-grade students in Maine.

Author Keywords
Online math homework; log analysis, prediction

L@S 2016, April 25-26, 2016, Edinburgh, Scotland UK
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893417

ACM Classification Keywords
K.3.1 Computers and Education: Computer Uses in Education–Computer-assisted instruction

Introduction
Homework is a well-established practice in schools and has been stable since the mid-1980s [10, 12]. Yet public controversy has arisen from time to time about homework's impact on learning [e.g., 2, 11] and whether it is worthwhile [1]. Research has also addressed the role and practices of homework and its relationship with student learning, especially in mathematics [3-5, 8, 9, 13, 14, 15, 16].

Relative to classroom learning, however, homework has been underresearched because it is difficult to collect objective data (as opposed to self-reported data) on homework behaviors. Little is known about when students do homework, how long it takes them to complete it, how much time they spend on problems, and whether and where they struggle. As educational technologies such as Khan Academy, ALEKS, and IXL.com have gained wider use in schools and in the home, the opportunity exists to better leverage homework for learning and also better understand homework practices from the data collected through computer systems.

Design of the Maine Online Homework Efficacy Study

The Maine Online Homework Efficacy Study was funded by the Institute of Education Sciences, U.S. Department of Education. The study had a school-level, delayed-treatment, randomized experimental design. Implementation of Use of ASSISTments was staggered by cohort. During the 2012–15 school years, 44 schools in Maine joined the study in two cohorts. Each school stayed in the study for two consecutive years. Within each cohort, half the schools were randomly assigned to the treatment condition and half to the control condition. A total of 87 seventh-grade math teachers in the treatment schools used ASSISTments to assign and review homework. They were expected to assign approximately 25 minutes of homework in ASSISTments for a minimum of three nights per week. The specific days of assignment were determined by the teachers.

SRI International, with the University of Maine and Worcester Polytechnic Institute, conducted a large-scale efficacy study in 44 schools in Maine. The objective was to test the hypothesis that the ASSISTments homework support platform improves student mathematics outcomes. Teachers choose (or add) homework items in ASSISTments, and students complete the items online. Students receive immediate feedback on the correctness of their answers, hints messages, or help decomposing multistep problems into parts. Teachers receive reports with such information as the percentage of items students got correct on their nightly assignment and common errors among groups of students. Teachers are encouraged to use the information for more targeted homework reviews and to more generally adapt or differentiate their teaching. We've analyzed the outcome data from the study and found a significant treatment effect (effect size = .27).

Feng et al. [6] had established the assessment validity of ASSISTments using data on how students performed and interacted with the system during class time to predict their end-of-year state test scores. In the research reported here, we replicated that approach to explore the relationship between students' ASSISTments use on homework outside the classroom and their performance on a summative standardized test.

Data Source

The data were math homework log data from ASSISTments, TerraNova test scaled scores (range 400–900), and performance levels (1–5) of 1,555 seventh-grade students from Maine. The ASSISTments system collects data on student log-in, problems solved, problem-solving attempts, requests for help, response from the online tutor, and assignment completion status. All the actions are time stamped. TerraNova is a standardized paper test and was given to all students in control and treatment groups so as to have a common end-of-year measure. The test was nationally norm referenced, and both the scaled score and performance levels for each student were reported.

Analysis of Homework Log Data

We constructed metrics that represented students' use, performance, and behaviors in ASSISTments and calculated their correlation with students' scaled scores from the TerraNova test (Table 1). Among the features, *p*_count is a measure of intensity of use, and perc_avg and *p*_time_avg are performance measures. The remaining features captured students' behaviors: help-seeking (hint_avg and bhint_avg), frequency of attempts, time taken to respond to a problem on the first try, and whether homework was completed on time, late, or left unfinished (completed_perc, incomplete_perc, and late_perc). We found that perc_avg positively and strongly correlated with the scaled score (.479). We observed a weak negative relationship (.2 < r < .4) between TerraNova scaled scores and the behavioral variables hint_avg, bhint_avg, attempt_avg, and resp_time_avg.

Metrics	Description	Cor.
*p*_count	Total number of problems completed	.226
perc_avg	Avg % correct across assignments	.479
*p*_time_avg	Avg numbers of times needed to complete a problem	-.055
hint_avg	Avg number of hint requests per problem	-.340

Group	Avg *p*_count	*t/p*
I	620	I vs. II: t = -9.55, $p < .001$;
II	919	
III	1,097	II vs. III: t = -2.31, p = .02

Table 2. Number of problems completed by groups

Metrics	Description	Cor.
bhint_avg	Avg number of bottom-out hint (revealing answer) requests	-.353
attempt_avg	Avg number of attempts	-.242
resp_time_avg	Avg response time	-.307
completed_perc	% of assignments completed on time	.214
late_perc	% of assignments completed but late	-.160
incomplete_perc	% of assignments started but not completed	-.153

Table 1. Data features and their correlation with students' scaled scores from the TerraNova test

We then split the students into three groups based on their performance on the TerraNova test: I, performance levels 1 or 2; II, levels 3 or 4; and III, level 5. There was a significant difference in the *p*_count among the three groups (see Table 2). Similar trends were evident across the three groups for other features. For example, students at a higher performance level finished more assignments on time, used a significantly fewer bottom-out hints, and had significantly fewer incomplete assignments.

After the exploratory analysis, we built a series of predictive models (Table 3) using students' TerraNova scores as the dependent variable and various combinations of the homework features as predictors (after they were normalized) and used R^2 and the Bayesian Information Criterion (BIC) to compare models. We started with a baseline linear regression model with no predictors (model.0), added perc_avg (model.1), and then added all other features in Table 1 as independent variables. We used a two-direction stepwise model training and variable selection process, using BIC as the selection criterion (model.2).

Of all the models, model.3 had the lowest BIC and the highest R^2, 0.41. It was a two-level mixed-effects model, including perc_avg, resp_time_avg, and *p*_count (the selected variables from model.2) as fixed effects and a school-level random effect. This suggested that students doing more homework problems in less time would have better achievement outcomes.

Models	Independent V.	BIC	R^2
model.0	None	15184	0
model.1	perc_avg	14817	0.22
model.2	perc_avg + p_count + resp_time	14743	0.26
model.3	variables in model.2 as fixed effects + school level random effect	**14489**	0.41
model.4	All features as fixed effects + school level random effect	14511	0.41

Table 3. Predictive models, BICs and R^2

Conclusion and Future Work

The predictive model using student homework logs from an online support system did not show results as impressive as those reported [6] when classroom use data were used to predict end-of-year standardized test scores ($R^2 = 0.73$). Yet considering that homework (a) takes only a limited percentage of time in students' overall learning, (b) is much more distributed and casual than classroom practice, and (c) is largely self-monitored and student controlled, we believe this work contributes to our understanding of how students do homework and thus can help teachers and students better leverage it to improve learning outcomes. We foresee that using technology to drive homework improvement is an important opportunity for learning at scale. An immediate next step following this work is to examine whether a relationship exists between students' use of ASSISTments for homework and any change in their performance on summative standardized tests, controlling for their incoming knowledge.

Given that we have identified variables linked to student performance, adaptations could be made so that ASSISTments better supports students. For example, detectors could be built in to sense and alert teachers to changes in homework completion patterns or high rates of bottom-out hints. Teachers could then provide interventions to promote more effective use of ASSISTments for homework support.

References

[1] Abramian, A. Fourth-grade twins, one with homework, one without. The Motherlode blog. *New York Times* (2015).

[2] Bennett, S. and Kalish, N. *The Case Against Homework: How Homework Is Hurting Our Children and What We Can Do About It*. Crown Publishers, New York, NY, USA, 2006.

[3] Cooper, H., Robinson, J.C. and Patall, E.A. Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research, 76* (2006), 1–62.

[4] Cooper, H. *The Battle Over Homework* (3rd Ed.). Corwin Press, Thousand Oaks, CA, USA, 2007.

[5] Eren, O. and Henderson, D. Are we wasting our children's time by giving them more homework? *Economics of Education Review*, *30*, 5 (2011), 950-961.

[6] Feng, M., Heffernan, N.T. and Koedinger, K.R. Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*, 3 (2009), 243-266.

[7] Feng, M., Roschelle, R., Murphy, R. and Heffernan, N. Using analytics for improving implementation fidelity in a large scale efficacy trial. In *Proc. ICLS 2014,* International Society of the Learning Sciences (2014), 527-534.

[8] Fernández-Alonso, R., Suárez-Álvarez, J. and Muñiz, J. Adolescents' homework performance in mathematics and science: Personal factors and teaching practices. *Journal of Educational Psychology,* March 16 (2015).

[9] Galloway, M. K. and Pope, D. Hazardous homework? The relationship between homework, goal orientation, and well-being in adolescence. *Encounter*, 20 (2007), 25–31.

[10] Gill, B. and Schlossman S. A nation at rest: The American way of homework. *Educational Evaluation and Policy Analysis*, *25*, 3 (2003).

[11] Kohn, A. *The Homework Myth: Why Our Kids Get Too Much of a Bad Thing*. Da Capo Press, Cambridge, MA, USA, 2006.

[12] Loveless, T. *The Brown Center Report on American Education*. The Brookings Institution, Washington, DC, USA, 2014.

[13] Maltese, A.V., Robert, H.T. and Fan, X. When is homework worth the time? Evaluating the association between homework and achievement in high school science and math. *The High School Journal*, October/November (2012), 52-72.

[14] Marzano, R. J. and Pickering, D. J. The case for and against homework. *Educational Leadership*, *64* (2007), 74–79.

[15] Trautwein, Ulrich, et al. Do homework assignments enhance achievement? A multilevel analysis in 7th-grade mathematics. *Contemporary Educational Psychology*, 27.1 (2002), 26-50.

[16] Trautwein, U. The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction*, *17* (2007), 372–388.