ABSTRACT
Recent studies [10, 23] using US nationwide databases showed high school boys spent significantly less time doing homework than girls, based on their responses to questionnaires and surveys. To investigate gender differences in homework in middle school, in this paper, we analyzed computer log data and standardized test scores of more than 1,000 7th grade students who participated in a large-scale randomized controlled online homework efficacy study. Students used the ASSISTments platform to do their homework for a school year. Our results suggested no significant difference between the time the two genders spent on homework overall. There was a marginally significant difference in homework time between genders in the high performing group only. When examining the system-student interaction data, we found significant difference between boys and girls in their help-seeking behaviors. In addition, we found out that boys have benefited from the online homework intervention more than girls.

Keywords
Gender gap, homework, online homework intervention

1. INTRODUCTION
Studies have investigated gender differences in homework completion rates, learning habits, and technology use outside of school. The investigation into gender differences found that girls spend more time on homework [36], including math [28]. Further, research has also shown that girls are more likely to spend time regulating study habits (e.g., time management, engaging in emotion self-regulation while doing homework) [13, 16, 37, 38, 39]. This was especially true with girls receiving family help while doing homework [35, 36]. With regards to gender differences in technology out of school, research clearly indicates boys have an advantage over girls with using technology for more varied reasons (e.g., programming, gaming, or internet surfing) than girls (e.g., drawing) [33] and more frequently as well [12, 17, 22, 24, 27, 34]. This gender-based advantage extends to girls’ attitudes toward computer usage. Girls tend to exhibit lower self-efficacy beliefs about their use of computers [21, 33]. At the same time, studies also document parent support as a critical mitigating factor that can increase girls’ use of and experience with computers [21, 33].

More recently, two studies, [10] and [23] suggested that boys spend less time on homework than girls. Based on the PISA 2012 Database, [23] shows that around the globe, 15-year-old boys are overwhelmingly less likely than girls to spend time doing homework, which may in part explain why they are more likely to struggle academically. The study has been widely cited in recent press coverage (e.g. [26]). In the U.S., boys on average spend 1.8 hours less time per week doing homework than girls. When considering boys and girls who spend the same amount of time doing homework, the gender gap in mathematics achievement is wider. [10] examined data from American Time Use Survey (AUS) responses. They showed that high school girls spent significantly more time (17 minutes per day) on homework than male high schoolers, even after controlling for SES indicators, daily activities and other factors. Furthermore, the gap for time spent on homework is largest among high-achieving students.

These studies illustrate that achievement gaps between genders’ use of homework does exist. However, we noticed almost all studies on gender differences in homework use self-reported measures. PISA 2012 asked students to report how much time per week they spend doing homework by teachers. [10] used students’ non-school study time using time diary data from 2003-2013 waves of the AUS and transcript data from the Educational Longitudinal Study of 2002 (ELS). Our literature search shows that there is a serious need for rigorous homework research on homework in K-12 settings. The existing studies are mostly correlational survey studies with thousands of students that relate homework time, academic-, and non-academic outcomes.

Our online homework study, which is a rigorously designed, randomized controlled experiment, gives us a unique opportunity to study the gender gap using more objective data sources of homework—computer logs from an online platform that support middle school students doing math homework. In this paper, examine the difference between genders in middle school mathematics on

- homework time
- the amount of problems completed by each gender
- homework performance
- how each gender interacted with the system
- whether there was any difference in the outcome measure between the two genders
- which gender benefited more from the technology-based intervention

Proceedings of the 9th International Conference on Educational Data Mining, June 29-July 2, 2016, Raleigh, NC
2. BACKGROUND

2.1 Online homework study
Research has been conducted to study the role and practices of homework and its relationship with student learning, particularly for mathematics (e.g. [2, 3, 5, 8, 9, 19, 20, 28, 29]). The link between homework assignments and student achievement is far from clear across the board, as noted by Cooper and others [30]. Although some studies show that students—and especially struggling students—could benefit from middle school mathematics homework, they may not benefit under typical conditions. Technology-based learning environment, such as ASSISTments, provides ways to make homework more adaptive and productive for the students who could benefit most. These environments can also do some of the bookkeeping and help teachers to keep track of the progress of their students. They enable teachers to assign customized homework to their students. For example, while doing homework in ASSISTments, students receive support including immediate feedback on the accuracy of their answers, as well as extensive tutoring. With these supports in place, students may complete more homework and learn more while doing homework. Teachers may be freed from the tedium of grading homework and be able to instead focus their energies adjusting and differentiating instruction.

SRI International, in conjunction with the University of Maine and Worcester Polytechnic Institute (the developer of the ASSISTments platform) conducted a multiyear randomized controlled efficacy trial at the school level. The study was conducted in 44 schools in the state of Maine, where one-to-one computing has been well-established for over 10 years. This experiment tested the hypothesis that the ASSISTments homework support improves student mathematics outcomes and will also examine impacts for struggling students and other important demographic groups. Schools in the study were randomly assigned to treatment or control (i.e. “business as usual”) conditions. The intervention was implemented in 7th grade classrooms in treatment schools over 2 consecutive years. In the control condition, teachers and students continue with their existing homework practices. In the treatment condition, teachers received professional development and used ASSISTments in the first year to become proficient with the system and then teachers used ASSISTments with a new cohort of students in the second year which is considered the “experiment year”. At the end of the experiment year, students were administered the TerraNova Common Core math test to provide data on student achievement in mathematics. TerraNova is a norm-referenced achievement test that is nationally normed. It generates scaled scores (ranging from 400 to 900 points) and achievement-level information that include five levels of performance proficiency (1: Starting-out; 2: Progressing; 3: Near Proficient; 4: Proficient; 5: Advanced).

2.2 Key features of ASSISTments platform
ASSISTments [www.assistments.org] [6, 14] is an online tutoring system that provides “formative assessments that assist.” Teachers choose (or manually add) homework items in ASSISTments and students can complete their homework online. As students do homework in ASSISTments, they receive feedback on the accuracy of their answers. Some problem types provide hints to help students improve their answers, or help decompose multistep problems into parts (so-called “scaffolding questions”)

Figure 1. Screen shots of an 7th grade item in ASSISTments that provides correctness feedback and breaks the problem into steps. (see Figure 1). Teachers may choose to assign problem sets called “skill builders” that address individual math concepts and skills at grade level and are organized to promote mastery learning. Every night, ASSISTments servers generate customized, cognitive diagnostic reports. The reports show teachers homework completion rates, performance data for each student on every problem and each math skill covered in the assignment, which questions and/or skills were particularly challenging for, and what the common wrong answers were. The report is emailed to teachers early in the morning for their review. This data allows teachers to make real-time, informed decisions about what and how they teach, and it is ideally used to guide homework review practices in class.

The usage model of the online homework study specifies that teachers who used ASSISTments in the study were expected to assign approximately 20 minutes of homework in ASSISTments for a minimum of three nights per week (making adjustments as needed to accommodate district and school homework policy).

3. EXPLORING GENDER DIFFERENCES IN HOMEWORK TIME, BEHAVIORS, AND PERFORMANCE
The data used in this section includes ASSISTments system logs of 1033 7th grade students, including 514 boys and 519 girls, who participated the second year of the homework study in the treatment condition. Also included in the data are their TerraNova test scores including both scaled scores and their performance levels. These students used ASSISTments for homework for the whole school year.

3.1 Features
We started the data analysis by constructing features that represent student’s intensity of use, performance, and behaviors while working in ASSISTments. Below, we list all the features.

• mins_s: Total number of minutes students spent on homework in the year
• probs_c: Total number of problems completed
• perc: Average percent correct over all assignments
• hint_c: Average number of hint requests per problem

Marty surveyed 24 students and asked them to name their favorite fruit. The circle graph below shows the results of his survey.

Which fruit was the favorite of exactly 6 of the students?
• attempt_c: Average number of attempts per problem
• bottom_hint_c: Average number of bottom-out hint requests per problem
• comp_perc: % of homework assignments completed on time
• late_perc: % of assignments completed but late

Two features, mins_s and probs_c are measures of intensity of use of ASSISTments. perc is a measure of student’s performance on homework problems. Some other system features (hint_c, attempt_c, and bottom_hint_c) capture students’ interaction with the system while doing homework, including their help-seeking behaviors (hint_c and bottom_hint_c) and the number of attempts they made before getting a correct answer (attempt_c). The last two features show whether they complete their homework on time or late (comp_perc, late_perc) as opposed to not completing an assignment at all.

### 3.2 Visual exploration of homework time

Research has shown that spending more time doing homework is better for academic achievement [3, 28, 30, 32]. Additional research has also shown that homework time is associated with many factors that may have a positive effect on academic success such as motivation or academic interest [4, 15] and parent involvement [1, 25]. Therefore, we started with an exploratory analysis focusing on the time students have spent on doing homework in ASSISTments. We observed relatively weak positive relationships\(^1\)\((0.2 < \rho < 0.4)\) between students’ TerraNova scaled scores and system use and performance indicators (mins_s, probs_c, and perc), suggesting students who spent more time on homework and completed problems scored higher on the TerraNova test. When we examined the usage data closely, we found that students spent a wide range of time on homework in ASSISTments in the school year (ranging from 2 to 4,238 minutes, mean = 640, standard deviation = 784), and amount of use varies a lot by schools (65% of the variance in mins_s is accounted by schools). Although homework practice is expected to differ across teachers and schools, the large variance is to some extent surprising, as the research team has specified a desired use model and has expressed the expectations clearly to all teachers in the treatment schools. On the other hand, this result confirms our previous findings on implementation fidelity from the previous 2013-14 school year where adherence, exposure, and uptake of users varied by teachers [7].

Next, we further explored the relationship between homework time and students’ achievement outcomes. We found that higher-performing students tend to spend more time on homework. Girls seem to spend relatively more time on homework than boys do, except in the middle level of achievement. The difference is most notable in the “5: Advanced” level.

Then we compared the TerraNova performances of boys and girls who spend similar amounts of time on homework. We found that there are more girls than boys who spent a significant amount of time on homework (defined as over 3,200 minutes in the school year). Unlike [23], however, we didn’t see big gender gaps in mathematics achievement (Figure 3).

![Figure 2. Bar graph comparing the average homework time by students in each TerraNova performance level, split by gender](image)

![Figure 3. Plot comparing TerraNova performance of boys and girls who spend similar amount of time on homework](image)

### 3.3 Modeling gender difference on usage and homework performance

Table 1 shows the descriptive statistics of all the features by gender. We noticed that the mean difference between the two genders was high on the two features, mins_s and probs_c, yet standard deviations on those measures were also quite high. We understood the extent to which schools create variation in homework behaviors: differences in the amount of homework assigned between teachers and schools, possible variations in homework review processes, and differences in teachers’ completion policies. Since these factors could affect students’ performance and/or behavior when doing homework, we trained a series of 3-Level Hierarchical Linear Regression models (HLM) (students nested in classes and classes in schools) to account for the difference in schools’ and teachers’ homework assignment practices. We used each feature as a dependent variable and use gender of students as the predictor (male = 0, female = 1).

<table>
<thead>
<tr>
<th>Features</th>
<th>Male Mean</th>
<th>Male Stdev</th>
<th>Female Mean</th>
<th>Female Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>mins_s</td>
<td>820.337</td>
<td>759.742</td>
<td>874.755</td>
<td>807.623</td>
</tr>
<tr>
<td>probs_c</td>
<td>703.214</td>
<td>592.099</td>
<td>770.734</td>
<td>621.226</td>
</tr>
<tr>
<td>perc</td>
<td>0.740</td>
<td>0.115</td>
<td>0.744</td>
<td>0.117</td>
</tr>
<tr>
<td>hint_c</td>
<td>0.115</td>
<td>0.143</td>
<td>0.094</td>
<td>0.103</td>
</tr>
<tr>
<td>attempt_c</td>
<td>1.403</td>
<td>0.281</td>
<td>1.375</td>
<td>0.248</td>
</tr>
<tr>
<td>bottom_hint_c</td>
<td>0.072</td>
<td>0.074</td>
<td>0.061</td>
<td>0.068</td>
</tr>
<tr>
<td>comp_perc</td>
<td>0.614</td>
<td>0.284</td>
<td>0.646</td>
<td>0.259</td>
</tr>
<tr>
<td>late_perc</td>
<td>0.129</td>
<td>0.12</td>
<td>0.14</td>
<td>0.127</td>
</tr>
</tbody>
</table>

As shown in Table 2, the results suggest that overall there is no significant difference between girls and boys in terms of the amount of time they spent on homework or the number of problems they completed. Furthermore, there is no difference between the two genders in their rates of correctly answered

---

1 The system doesn’t limit the number of answers a student could attempt on a problem.
2 When using ASSISTments in the practice and learning modes (as opposed to testing mode), the system requires that every problem has to be answered correctly in order for students to move to the next one. Bottom-out hints in ASSISTments reveal the correct answer to students so that they won’t get stuck.
3 No other correlations were noticed.
problems in ASSISTments. Girls tend to complete more assignments on time than boys, but the difference is only marginally significant ($p = .086$). However, interestingly, girls and boys interacted with the system differently; girls made fewer hint requests and fewer attempts on problems, and they also requested fewer bottom-out hints as compared to boys in the same classes.

Table 2. HLM Results Overall – Predictor: Female

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Difference</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mins s</td>
<td>22.482</td>
<td>0.350</td>
</tr>
<tr>
<td>probs c</td>
<td>27.219</td>
<td>0.177</td>
</tr>
<tr>
<td>perc</td>
<td>0.006</td>
<td>0.351</td>
</tr>
<tr>
<td>hint c</td>
<td>-0.018</td>
<td>0.005 **</td>
</tr>
<tr>
<td>attempt c</td>
<td>-0.039</td>
<td>0.005 **</td>
</tr>
<tr>
<td>bottom hint c</td>
<td>-0.011</td>
<td>0.002 *</td>
</tr>
<tr>
<td>comp perc</td>
<td>0.015</td>
<td>0.086</td>
</tr>
<tr>
<td>late perc</td>
<td>-0.000</td>
<td>0.907</td>
</tr>
</tbody>
</table>

Inspired by Gershenson & Holt (2015) and Figure 3 shown above, we were interested to see whether there was any interaction effect between gender and students’ performance levels. Thus, we split the students into 3 groups based on their performance level on the TerraNova test. We then trained similar HLM models within each group of students, and the results are shown in Table 3.

• **Progressing or Below**: performance levels = 1 or 2; N= 328 (male: 145, female: 183)
• **Near Proficient**: performance levels = 3; N = 368 (male: 165, female: 203)
• **Proficient or Above**: performance levels = 4 or 5); N = 337 (male: 166, female: 171)

We observed trends with regard to how students interact with the system in both the **Near Proficient** and **Proficient or Above** groups. The trends are consistent with the overall trend: girls requested significantly fewer hints or bottom-out hints, and made fewer attempts on problems. Results regarding completion assignment status are mixed. Low-performing girls completed fewer assignments after they were due than low-performing boys did; yet in the **Near Proficient** group, girls completed more assignments later than boys did. In the **Proficient or Above** group, girls were more likely to complete assignments on time. Interestingly, we noticed a marginally significant difference in **mins s** in the **Proficient or Above** group, suggesting high-performing girls spent more time on homework than high performing boys. This result is in consistent with [10], but the latitude of difference is not as big.

4. WHICH GENDER BENEFITED MORE FROM TECHNOLOGY-BASED HOMEWORK INTERVENTION?

One of the research questions of the online homework study is to investigate whether the impact of the ASSISTments on learning outcomes differ by student demographic characteristics. Here we present the analysis that was conducted to examine which gender benefited more from online homework intervention. A different dataset was used for this analysis. Students from the control condition were included in this dataset in order to detect the interaction between intervention and gender, which increased the total number of students to 2,756 from 44 schools. Only students’ assigned condition, gender, their 6th grade state test scores, and their TerraNova scaled scores were included in this dataset. TerraNova scores were used as dependent variable. 3-level HLM models were employed in all the analysis.

We first ran a basic model that includes prior achievement (6th grade math state test scores) and gender (male=1, female=0) as predictors of TerraNova scaled scores to examine effects of gender. The HLM model for the analysis of effect of gender is illustrated below.

Level-1 model: \[ TScore = \beta_0 + \beta_1*(PriorMath) + \beta_2*(Male) + r \]

Level-2 model:
\[
\begin{align*}
\beta_0 &= \gamma_{00} + \gamma_{01}*(Trx) + u, \\
\beta_1 &= \gamma_{10}, \\
\beta_2 &= \gamma_{20}
\end{align*}
\]

In this model, TScore is the student’s scaled score from the TerraNova test. Trx is a school-level indicator of the school being in the treatment condition (0=Control, 1=Treatment). Student-level variables, PriorMath is a student-level variable, representing the student’s 6th grade math state test score. Male is a student-level variable, indicating the student’s gender (0=Female, 1=Male). The model showed that students in the treatment condition scored 10.26 points higher than control students and males scored 5.21 points lower than females. Both effects are statistically significant ($p < .001$). To help understand the difference, we referred to TerraNova technical norms published by CTB. The norms showed that the average yearly growth from 7th to 8th grade is about 10 points in scaled score.

Table 4. HLM Results on Intervention and Gender Effect

<table>
<thead>
<tr>
<th>Gender</th>
<th>Control</th>
<th>Treatment</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>683.21</td>
<td>693.46</td>
<td>10.26</td>
</tr>
<tr>
<td>Males</td>
<td>677.99</td>
<td>688.25</td>
<td>10.26</td>
</tr>
<tr>
<td>Difference</td>
<td>-5.21</td>
<td>-5.21</td>
<td></td>
</tr>
</tbody>
</table>

Then we augmented the basic model by adding an interaction term between treatment and gender. The augmented model is illustrated below.

Level-1 model: \[ TScore = \beta_0 + \beta_1*(PriorMath) + \beta_2*(Male) + r \]

Level-2 model:
\[
\begin{align*}
\beta_0 &= \gamma_{00} + \gamma_{01}*(Trx) + u, \\
\beta_1 &= \gamma_{10}, \\
\beta_2 &= \gamma_{20} + \gamma_{21}*(Trx) + u
\end{align*}
\]

Proceedings of the 9th International Conference on Educational Data Mining, June 29-July 2, 2016, Raleigh, NC 367
In this model, the effect of ASSISTments intervention was found to vary by gender ($\chi^2=7.476$, $t(42)=2.232$, $p = 0.031$). As shown in Table 5, boys in the control group scored 9.61 points lower than girls in the control group, but boys in the treatment condition scored only 2.13 points lower than girls in the same group. Girls in the treatment group scored 6.73 points higher than those in the control group (which was not significant after adding in the interaction), while boys in the treatment group scored 14.21 points higher than those in the control group. In essence, the intervention helped close the gender gap between girls and boys for standardized test achievement and boys have benefitted more from the intervention than girls.

Table 5. HLM Results on Intervention and Gender Interaction Effect

<table>
<thead>
<tr>
<th>Gender</th>
<th>Control</th>
<th>Treatment</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>685.20</td>
<td>691.93</td>
<td>6.73</td>
</tr>
<tr>
<td>Males</td>
<td>673.59</td>
<td>689.79</td>
<td>14.21</td>
</tr>
<tr>
<td>Difference</td>
<td>-9.61</td>
<td>-2.13</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSIONS AND FUTURE STUDIES

In this paper, we examined the difference between genders in middle school mathematics on homework time, the amount of problems completed by each gender, homework performance, and how each gender interacted with the system, using computer system log data from an online homework intervention. We also answered two research questions regarding which gender benefited more from a technology-based intervention supporting homework. Our results suggested no significant difference between the time the two genders spent homework overall. Among students who performed proficiently or above on the end-of-year standardized test, girls have spent more time on homework than boys, and the difference was marginally significant. We also found out that when using ASSISTments for homework, girls and boys differed in their help-seeking and problem Attempting behaviors. Girls requested less hints, made less number of attempts on problems, and they also requested less amount of bottom-out hints that would reveal the correct answers to problems. Our findings suggested that the intervention closed gender gaps in mathematics achievement in 7th grade and boys benefited from the online homework intervention more than girls.

We speculated on the reasons why boys have benefited more from the technology-based intervention. One reason could be boys in the study were more comfortable with using technologies, similar to what has been reported in earlier research. We also checked to see if there was any difference between the two genders in prior achievement. Using a simple t-test, there was no gender difference in 6th grade state math test scores (Female average score = 645, Male average score = 644, $p = 0.252$).

Researchers have been able to identify factors that impact this relationship between time spent doing homework and academic achievement. It was found that the quality of time spent on a task, i.e., homework, is a more critical predictor of student learning than the total number of minutes spent on the task. For instance, time on task or perseverance manifested with low distraction rates is positively correlated with achievement [30]. Other factors, especially the effort students put into homework and how frequently they do homework are far more reliable and positive predictors of student achievement [28, 30, 32]. As a follow-up study, we plan to look at student behaviors when working in the system more closely, taking sequence and time into account. We plan to study help-seeking and problem-Attempting behaviors at action level and to see whether there are any the sequential pattern of actions taken, and whether there is between girls and boys. For instance, did boys ask for hints/bottom-out hints right away, while girls took time to persevere through challenging homework problems before requesting for assistance? We also plan to build a dataset including students’ frequency of logging in each day and each week and the duration of the working sessions by gender, and see how such features predict student learning. Such studies will help the field better understand gender differences in STEM learning, esp. in out-of-classroom settings. The findings can be informative for the development of behavior detectors in online learning systems like ASSISTments so that the systems can provide interventions to improve learning outcomes and close gender gaps.

We recognize the limitations in our study. We have no access to information, such as parent involvement, their extra-curricular activities, etc. that may affect student’s homework completion rates, their behaviors when doing homework, or their performance. Nor do we have access to student’s attitudes towards mathematics, technology or homework. All of these limit our ability to explain the differences we’ve discovered. The results presented in this paper were based on data from 7th grade students who are younger than the high school students who have been the focus of [23] and [10]. It would be a reasonable next step to extend such kind of study to elementary students and see if there might exist a trajectory in the gender differences in homework.

6. ACKNOWLEDGMENTS

This material is based upon work supported by the Institute of Educational Sciences (IES) of U.S. Department of Education under Grant Number R305A120125. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

7. REFERENCES
