# Hint Availability Slows Completion Times in Summer Work


Paul Salvador Inventado,
Peter Scupelli
School of Design
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, USA
paulsb@andrew.cmu.edu,
pgs@andrew.cmu.edu

Eric G. Van Inwegen,
Korinn S. Ostrow,
Neil Heffernan III
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA, USA
egvaninwegen@wpi.edu,
ksostrow@wpi.edu,
nth@wpi.edu

Jaclyn Ocumpaugh,
Ryan S. Baker,
Stefan Slater,
Mia Almeda
Teachers College, Columbia University
525 W. 120th Street,
New York, NY, USA
jo2424@tc.columbia.edu,
baker2@exchange.tc.columbia.edu,
slater.research@gmail.com,
victoriaalmeda@gmail.com



## ABSTRACT
On-demand help in intelligent learning environments is typically linked to better learning, but may lead to longer completion times. This present work provides an analysis of how students interacted with a summer learning assignment when on-demand help was available, compared to when it was not. When hints were available from the start, students were more likely to delay work, compared to students for whom step-wise hints were only available after the third problem. When hints were always available, participants took significantly more time to complete a mastery learning assignment. We interpret this difference in time to complete the assignment as an opportunity to re-engage in productive math learning.



## Categories and Subject Descriptors
H1.2 [**Information Systems**]: User/Machine Systems – human factors

## General Terms
Measurement, Design, Experimentation, Human Factors

## Keywords
Hints, completion time, randomized controlled trial, ASSISTments


## 1. INTRODUCTION
Help-functions—including on-demand help, contextualized hints, or supplementary learning materials [2]—are a major asset of modern intelligent learning environments. These functions have often been associated with better student learning outcomes ([1][9][25]), but not all help has proven equally effective, and even well-crafted hints may be used ineffectively by students who do not actually need them ([2][20]) . Research has shown cases in which help functions fail [1] and has sought to identify the contexts in which different types of help strategies are most effective ([12][22]).


Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.



*EDM '16*, June 29–July 2, 2016, Raleigh, NC, USA.



Copyright 2010 ACM 1-58113-000-0/00/0010 …$15.00.


Analysis of hint use serves many purposes and may be an obvious answer to *wheel-spinning*, where a student persists long past the point of productive effort [6]. It is also feasible to predict the problematic behaviors of hint misuse or hint abuse. Previous research has analyzed relationships between problem-related features (e.g., problem length, number of hints available, hint length) and student affect, behavior, and learning ([3][11][13][19]). Among other findings, hint length has been positively correlated with *gaming the system* [3], a behavior incorporating help abuse that is associated with poorer learning outcomes ([21][23]). Other research has indicated problems unrelated to the deliberate behavior of students. For example, poorly designed hints may lead to ineffective hint usage ([4][15]). Research also suggests that low-knowledge students, or those that need the most help, are the least likely to use it effectively ([2][3][18]).

In this paper, we present results from a randomized controlled trial (RCT) that examined how hint availability effected other aspects of student learning, including the time required for students to complete the assignment, presented using the ASSISTments online learning system [11]. To our surprise, we found that students who were given the option to request on-demand hints appeared to spend more time on tasks unrelated to the completion of the problem set (e.g., solve other problem sets, work on learning activities outside of ASSISTments, or engaged in activities external to the learning system). Specifically, these students took more time to complete the assignment even though they did not (a) spend significantly more time on task, (b) answer significantly more problems, or (c) make significantly more attempts per problem as compared to the control condition. The analyses presented herein explore this pattern more thoroughly, in order to contribute to the growing literature on help systems in online learning.

## 2. ASSISTMENTS
ASSISTments is an online learning system designed primarily for middle school mathematics. The platform allows teachers to easily create and assign their own problem sets (including questions, associated solutions, mistake messages, feedback) or to select from a set of *ASSISTments Certified Problems* (vetted by ASSISTment's expert team) ([11][22]). These problem sets simultaneously support student learning and serve as automated formative assessments that provide real-time data to teachers [11]. The platform is also used as a research tool to conduct RCTs

([8][16][26]). ASSISTments logs learning-related features at multiple granularities (e.g., problem text, problem type, student actions, timestamps, etc.). Figures 1 and 2 show screenshots of the types of ASSISTments problems used in the present work. Based on experimental condition, students were able to request hints, receive feedback messages, or simply answer the question.

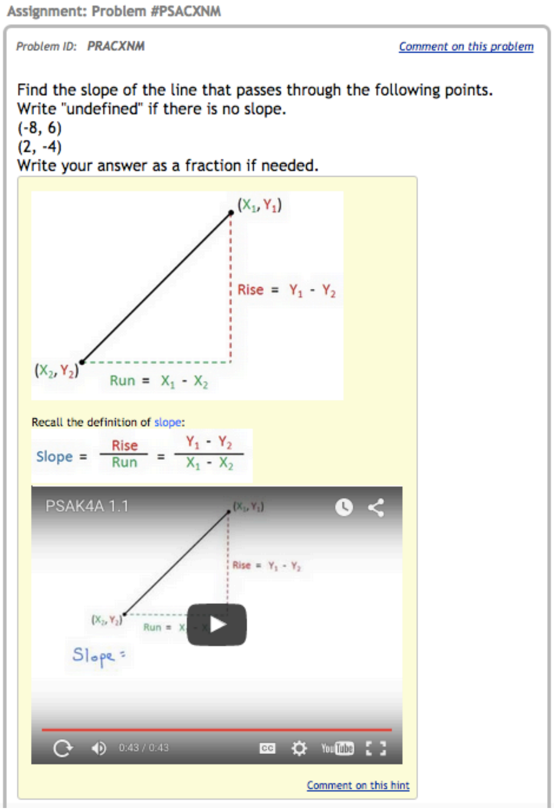


**Figure 1. An example question from the hints-early condition, presented with its associated hints.**

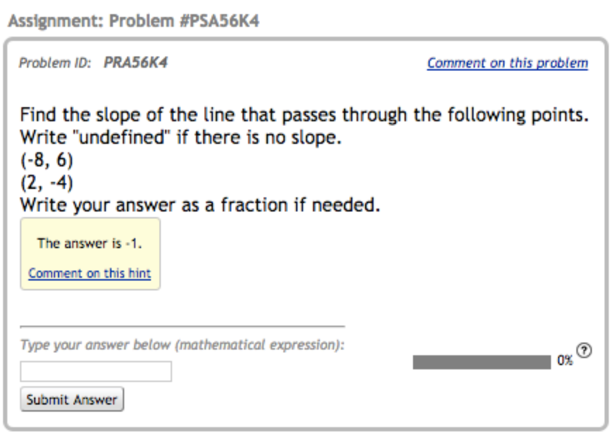


**Figure 2. The same example question as presented in the no-hints-early condition.**

## 3. METHODOLOGY

This study used an RCT design in which several linear presentations of a problem set were embedded within two conditions: a control condition with on-demand hints (*hints-early*, **HE**) or an experimental condition with on-demand hints only after the third problem (*no-hints-early*, **NHE**). The problem set for this study (available at [14]) was chosen from ASSISTments Certified content and was designed to address the 8th grade Common Core State Standard, "Finding Slope from Ordered Pairs," [17]. It was deployed within ASSISTments as a *Skill Builder*, a type of problem set requiring students to accurately answer three consecutive problems in order to complete the assignment.

Students were randomly assigned into one of 12 groups (6 control and 6 experimental) when they began the problem set. As depicted in Figure 3, students in each group saw the same 3 problems, but presentation order was randomized to minimize cheating (i.e., A-B-C, A-C-B, B-A-C, etc.). All students, regardless of condition, received immediate correctness feedback (e.g., "Sorry try again: '2' is not correct").

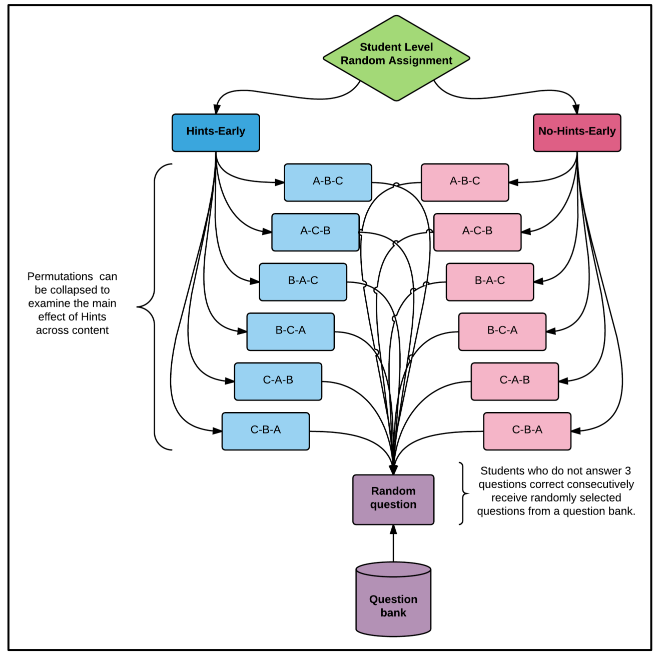


**Figure 3. Research Design depicted as a flow chart.**

In a Skill Builder, students are able to attempt each problem multiple times, but (in line with common practice) problem accuracy is calculated using binary correctness on the student's first attempt (1=Right, 0=Wrong) [10]. Students who did not answer the first three problems correctly were assigned additional problems randomly selected from a skill bank. In order to provide all students with adequate learning support, all students were permitted on-demand hints—regardless of condition—upon reaching these additional problems.

Figures 1 and 2 demonstrate how the interface differed by condition. In the HE condition, students could access hints at any time by clicking on a button in the lower right corner of their screen. The problem remained on the screen while video tutorials and text-based hints were simultaneously delivered (text-based hints ensured access when school firewalls or connectivity issues may have limited access to YouTube). In contrast, the NHE condition only offered a *Show Answer* button in the lower right corner of the screen during the first three problems (a design seen in early intelligent tutors [24]) allowing students who were stuck to move on to the next problem and eventually complete the assignment.

### 3.1 Student Populations

To help retain students' math skills, the Skill Builder in this study was one of many assigned as summer work at two suburban high schools (henceforth Schools A and B) in the Northeastern U.S. **School A** was an agricultural/vocational high school that assigned this Skill Builder to 113 9th graders and 95 10th graders, along with numerous other Skill Builders (32 in 9th grade, 36 in 10th). **School B** was a high school without a known specialization; it assigned this Skill Builder (as well as 45 others) to 204 9th graders. Students worked on these assessments throughout the summer (Jun-Sept 2015) and data was harvested six months later.

Condition distributions were well matched for student gender (HE: 101 f., 86 m., 29 unknown vs. NHE: 93 f., 89 m., 14 unknown), school, grade level, and classroom section. Students in both conditions had the same **prior Skill Builder completion rate** (HE: *M*=0.91, *Mdn*=1.0; NHE: *M*=0.91, *Mdn*=1.0, *p*=.463), which was computed by dividing the sum of prior Skill Builders started by the number of prior Skill Builders completed (amongst all ASSISTments assignments experienced by students in the sample) . Analysis using Mann-Whitney U tests (which are robust to skew) with a Benjamini-Hochberg false-discovery rate post-hoc correction for multiple tests (p<.05) [7], yielded no significant differences between the two conditions on several measures including total number of problems solved, time per problem, and number of attempts.

## 3.2 Measures Considered

This study considered several measures pertaining to students' answers and hint patterns. As noted above, students only completed the Skill Builder when they correctly answered three consecutive questions using first attempts. However, students were able to attempt problems multiple times. Students wishing to advance to the next problem but unable to generate the correct answer were able to request a bottom-out hint. When hints were available, students had to view between 1 and 3 regular hints before they were able to obtain the bottom-out hint, which provided the correct answer. In the first three problems of NHE condition, students could select *Show Answer*, which displayed only the bottom-out hint, but no additional assistance.

Several measures based on these behavioral patterns were considered, including: **number of problems solved (PS), mean answer-attempts per problem (MAA)**, **total answer attempts (TAA)**, **total hint requests (THR)** and **mean hint requests per problem (MHR)**. Spanning conditions, participants required 9.12 problems on average (*Mdn*=9.0, *SD*=3.32) to complete the assignment. Spanning conditions and problems, students averaged 16.14 total answer attempts (*Mdn*=14.0, *SD*=10.72), or 1.72 answer attempts (*Mdn*=1.71, *SD*=0.78) per problem. On average, students requested approximately one hint per nine problems (*Mdn*=0.0 , *SD*=2.23) throughout the Skill Builder. There were no significant differences in the aforementioned measures by condition according to Mann-Whitney U tests conducted with false discovery rate post-hoc corrections.

Next, we assessed several time-based measures to determine how hints were affecting students' completion rates. Basic measures including the number of **days** and **weeks** it took for a student to finish the Skill Builder were considered. These measures were analyzed both by completion time and by week of completion. As the data was heavily skewed (most students finished in week 1), a Mann-Whitney U test was used to analyze completion time. Six months after beginning the study, when data was harvested, seventy-two students (18%) had not completed the Skill Builder. Students who completed the Skill Builder were grouped according to whether it had taken them *1*, *2*, *3*, or *4 or more* weeks to complete, while those who never finished the Skill Builder were labeled as *incomplete*. We also considered, **Completion time** (**CT**, in seconds), or the total time it took students to complete the assignment, which was calculated by subtracting the start time of the first problem from the end time of the last problem solved.

Because the time students spent solving a problem was skewed, with a median of 1.1 minutes (*M*=16.22 hr, *SD*=4.69 days, *Min*=2 sec, *Max*=74.96 days), this value was *winsorized* to 15 minutes (900 sec) in order to exclude irrelevant conditions (e.g., disconnection from the network, shifts between learning activities, off-task behavior). The fifteen-minute time frame accounted for 93% of the data.

The winsorized measures were used to calculate **time-on-problem (TOP,** in seconds**)** for each problem in the Skill Builder that the student attempted to solve (i.e., end time minus start time for each problem). This measure was subsequently used to generate several others, including **mean time-per-problem (MTPP),** which showed a mean of 2.62 min (*Mdn*=2.35 min, *SD*=1.78 min) across all students. For each student, **TOP** was also **totaled** across all attempted problems (**TOP-total**), resulting in a mean of 23.42 minutes (*Mdn*=20.72 min, *SD*=16.93 min) across all students. Finally, **total time-between-problems (TTBP)**, was calculated by subtracting TOP-total from each students' completion time. Readers should note that because students were allowed to return to this assignment over the course of the summer, these values were comparatively large (*M*=6.73 days, *Mdn*=43 sec, *SD*=14.49 days). However, as Table 1 shows, variation among students who took more than one week was minimal at the problem level.

**Table 1. Mean values of time-based measures according to completion-time categories (weeks).**

| Week | PS | TOP-total | MTPP | TTBP | CT |
|---|---|---|---|---|---|
| 1 | 9.15 | 20.2 m | 2.2 m | 0.48 d | 0.49 d |
| 2 | 10.04 | 35.9 m | 3.7 m | 10.1 d | 10.1 d |
| 3 | 9.00 | 38.2 m | 4.4 m | 18.5 d | 18.5 d |
| ≥ 4 | 11.81 | 38.6 m | 3.3 m | 40.8 d | 40.8 d |
| Incomplete | 5.55 | 16.9 m | 3.6 m | 5.4 d | N/A |

***Note.*** **PS – problems solved; TOP-total – total time on problem; MTPP – mean time per problem; TTBP – total time between problems; CT – completion time, m = minutes, d = days**

# 4. RESULTS

ASSISTments automatically logged data in analyzable form. The following subsections present the results on hint usage, problem attempts, skill builder completion, and time-on-problem.

## 4.1 Hint Usage and Problem Attempts

This study used four primary measures of student actions, including total answer attempts, mean answer attempts, total hint requests, and mean hint requests per problem. Because the two conditions in this study only applied to the first three problems (after which, students in the no-early-hints condition also had access to regular hints), we report on values for the first three problems and those that follow separately.

Table 2 presents significant differences both between and within-conditions. There were no significant differences between conditions with respect to the number of attempts per problem or the total number of attempts used in solving the first three problems of the Skill Builder. That is, the availability of hints in the first three problems did not effect the number of attempts used or the number of hints requested over the course of the experiment. Likewise, the significant differences observed within condition all trended in the same direction, suggesting little to no effect.

**Table 2. Significant differences in answer attempts and hint requests by condition and within condition (p<.05).**

| | HE vs. NHE | | 1st 3 vs. Other problems | |
|---|---|---|---|---|
| **Measure** | **1st 3** | **Others** | **HE** | **NHE** |
| TAA | *NS* | *NS* | **Others > 1st3** | **Others > 1st3** |
| MAA | *NS* | *NS* | *NS* | **Others > 1st3** |
| THR | N/A | *NS* | **1st3 > Others** | N/A |
| MHR | N/A | *NS* | **1st3 > Others** | N/A |

***Note.*** **TAA – total answer attempts; MAA – mean answer attempts; THR – total hints requests; MHR – mean hint requests; HE – hints-early; NHE – no-hints-early; NS – not significant**

## 4.2 Hint Usage and Skill Builder Completion

One of the most important measures in this study was whether or not students were eventually able to demonstrate skill mastery by consecutively answering three of the Skill Builder questions accurately. Chi Squared tests revealed no significant difference between conditions in the proportion of students who did not complete the Skill Builder ($X^2$(1, N=412)=0.714, $p$=.398).

Non-completion in both conditions was associated with lower prior Skill Builder completion rates, suggesting that students' inability to master this Skill Builder was indicative of larger issues in completing their mathematics assignments (HE: $U$=1115.5, $p$<.001, NHE: $U$=471, $p$<.001). Non-completion was also associated with higher numbers of hint requests and answer attempts, both of which occurred across significantly fewer problems than worked by students who were able to complete the Skill Builder. Finally, non-completion was associated with significantly longer time worked across problems (**TOP-total**).

Despite nearly identical Skill Builder completion rates, the two conditions differed significantly in the time it took students to complete the problem set (HE: $M$=208.23 hrs, $Mdn$=38.55 min, NHE: $M$=67.52 hrs, $Mdn$=20.9 min, $U$=16835, $p$=.008). Specifically, as shown in Table 3, students in the no-hints-early condition completed the Skill Builder faster than those in the hints-early condition. These results were complemented by Chi Squared results that analyzed the distribution of students completing the assignment over several weeks, $X^2$(4, $N$=411)=8.981, $p$=.062. Again, this might seem obvious, as students who access hints tend to take longer to digest problem and feedback content, but further analysis suggests other factors should also be considered.

**Table 3. Number of students per condition who completed the Skill Builder each week**

| Weeks | HE ($N$=215) | NHE ($N$=196) |
|---|---|---|
| 1 | 125 (58%) | 137 (70%) |
| 2 | 15 (7%) | 13 (7%) |
| 3 | 5 (2%) | 3 (1%) |
| ≥ 4 | 30 (14%) | 13 (7%) |
| Incomplete | 40 (19%) | 31 (16%) |

***Note.*** **HE – hints-early; NHE – no-hints-early**

## 4.3 Hint Usage and Time-on-Problem

Hint availability could effect time-on-problem (**TOP**) in more than one way, even when students use hints effectively. Students who need hints may be expected to answer more slowly than their peers, but powerful hints may actually reduce the time that a struggling student takes to complete a problem (compared to a situation in which the same student did not have access to hints).

Table 4 (calculated with the Benjamini-Hochberg correction) shows a complex interaction between time-per-problem and hint use, but overall there were few differences between conditions. On the whole, the use of (regular) hints lead to longer time on problem (**TOP**) measures, but the effect of bottom-out hints differed by condition. In both conditions, students who used bottom out hints took longer to complete problems than those who did not use them. However, those who used bottom-out hints in the HE condition took less time per problem than those who only requested one (regular) hint. The latter pattern could be indicative of *gaming* behavior, and this warrants further investigation, but it is also possible that students who quickly realized their mistakes clicked through to the bottom-out hint in order to start work on the next problem.

Results further indicated that differences were driven by hint use effects in the first three problems, where students who did not have access to hints (the NHE condition) were significantly slower at answering than those who did (HE) ($M$=1.92 min, $Mdn$=1.80 min vs $M$=1.65 min, $Mdn$=1.37 min). This was a predictable difference, as struggling students in the HE condition could ask for hints, thereby removing themselves from this calculation, while struggling students in the NHE condition could only remove themselves from this calculation by requesting a bottom-out hint.

**Table 4. Time-on-problem comparison by condition (in minutes)**

| | **Mean (SD)** | | | | | | | | **Median** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Regular Hints Requested* | | | | | | | *Bottom-out Hint* | *Regular Hints Requested* | | | *Bottom-out Hint* |
| **Condition** | N | 0 Hints | N | 1 Hint | N | 2 Hints | N | | 0 Hints | 1 Hint | 2 Hints | |
| **First 3 Problems** | 373 | 1.78 (1.15) | 103 | 3.62 (1.33) | 0 | N/A | 167 | 2.98 (1.45) | 1.48 | 3.85 | N/A | 2.95 |
| HE | 191 | 1.65* (1.17) | 103 | 3.62 (1.33) | 0 | N/A | 81 | 3.47* (1.33) | 1.37* | 3.85 | N/A | 3.43* |
| NHE | 182 | 1.92* (1.13) | 0 | N/A | 0 | N/A | 86 | 2.55* (1.43) | 1.80* | N/A | N/A | 2.53* |
| **Other Problems** | 366 | 1.52 (0.92) | 22 | 2.52 (1.93) | 59 | 3.27 (1.23) | 56 | 3.27 (1.23) | 1.33 | 1.78 | 3.02 | 2.98 |
| HE | 190 | 1.50 (0.87) | 13 | 2.02 (1.92) | 30 | 3.20 (1.33) | 29 | 3.23 (1.33) | 1.33 | 1.65 | 2.92 | 2.93 |
| NHE | 176 | 1.53 (0.98) | 9 | 3.25 (1.82) | 29 | 3.37 (1.15) | 27 | 3.28 (1.15) | 1.42 | 3.65 | 3.47 | 3.02 |
| **All Problems** | 377 | 1.58 (0.78) | 113 | 3.50 (1.37) | 58 | 3.32 (1.17) | 174 | 3.05 (1.35) | 1.53 | 3.65 | 3.22 | 3.03 |
| HE | 195 | 1.50 (0.73) | 104 | 3.53 (1.33) | 29 | 3.28 (1.20) | 87 | 3.45* (1.27) | 1.52 | 3.63 | 2.93 | 3.40* |
| NHE | 182 | 1.67 (0.83) | 9 | 3.25 (1.82) | 29 | 3.37 (1.15) | 87 | 2.65* (1.33) | 1.57 | 3.65 | 3.47 | 2.70 |

***Note.*** **Units are in minutes. *p<.05. N – number of students; HE – hints-early; NHE – no-hints-early.**

Significant differences within and between conditions (summarized in Table 5) showed trends that suggested that behavior in the first three problems was driving the differences between the two conditions, where hint-access was restricted to the students in the HE condition. Interestingly, in the first three problems the mean time per problem was statistically similar. That is, for the first three problems, the HE and NHE condition did not differ overall, which suggests the need for understanding individual differences, such as those highlighted in Table 4. The significant differences between conditions emerged primarily in total time between problems (**TTBP**) and in the total completion time (**CT**), with students in the hints-early condition showing larger values for both measures.

**Table 5. Time Measures per Condition (p<.05).**

| | HE vs. NHE | | 1st 3 vs. Other problems | |
|---|---|---|---|---|
| | 1st 3 | Others | HE | NHE |
| MTPP | NS | NS | **1st3 > Others** | NS |
| TTBP | **HE > NHE** | NS | NS | **Others > 1st3** |
| CT | **HE > NHE** | NS | NS | **Others > 1st3** |

***Note.*** **MTPP – mean time-per-problem; TTBP – total time between problems; CT – completion time; HE – hints-early; NHE – no-hints-early; NS – not significant**

Further analyses revealed complementary patterns in within-condition differences. Students in the hints-early condition had significantly higher mean time-per-problem (MTPP) on the first three problems than they did on later problems ($M$=3.67 min, $Mdn$=2.63 min vs. $M$=2.17 min, $Mdn$=1.98 min, U=13281, $p<.001$), suggesting that those who effectively used these hints in the first three problems were learning the material well enough to complete later problems more efficiently. There were no significant differences in this group for other time-based measures (**TTBP** or **CT**). In contrast, students in the no-hints-early condition showed no significant differences for **MTPP**, but had longer **TTBP** and **CT** patterns for later problems than for the first three problems.

# 5. DISCUSSION

The present experiment was designed to explore the effects of ASSISTments' on-demand hints system. For ethical reasons, we limited differences between the control condition (providing hints) and the experimental condition (withholding hints) to the first three problems. All students had access to hints following the third problem to retain overall learning. However, effects could be seen even after students had moved past these first three problems.

The data used in the study was collected from one of many Skill Builders assigned to students for summer work. We explored the data using several different measures, extracting information about the number of attempts each student made, the number of hints (regular or bottom-out) they requested, and the length of time needed to complete the assignment.

Some findings were quite predictable, as reading hints would take more time than simply answering problems, assuming students were assigned problems that matched their current ability. However, other findings were more surprising. Even though students made the same number of attempts per problem and per assignment, those in the HE condition took significantly longer to complete the Skill Builder.

Students in the HE condition also spent relatively more time between problems compared to those in the no-hints-early condition, but only during the first three problems, where conditions were truly distinct. One interpretation of this finding is that students in the HE condition were taking more time between problems to process the new material they were learning. An alternative explanation is that students were procrastinating—deliberately putting off working on the Skill Builder out of difficulty or apathy (as summer work is highly self-regulated). These students could have been seeking out an easier Skill Builder to work on or may have spent their time doing something completely unrelated. Still, this latter interpretation may not be detrimental if students were using the time to work on other assignments. As Baker and colleagues have suggested [5], a student that goes off task and is able to re-engage afterwards may be more productive in the long run than those who persist at all costs.

# 6. CONCLUSION

This work presented an investigation of how students completing summer work responded to having or not having hints available on the first three problems of a Skill Builder assignment within the ASSISTments online learning system. When hints were available from the start, students were more likely to delay work in comparison to students for whom step-wise hints were only available after the third problem. When hints were always available, participants took significantly more time to complete the Skill Builder. We interpreted the difference in completion times as an opportunity to re-engage towards more productive math learning. In future work, we plan to conduct a similar study during the school year to examine how results differ in a more controlled and less self-regulated learning environment.

# 7. ACKNOWLEDGEMENTS

We gratefully acknowledge funding from the NSF (ACI-1440753, DRL-1252297, DRL-1109483, DRL-1316736, DGE-1535428 & DRL-1031398), the US Dept. of Ed (IES R305A120125 & R305C100024 and GAANN), and the ONR.

# 8. REFERENCES

[1] Aleven, V., McLaren, B., Roll, I., & Koedinger, K. 2006. Toward Meta-Cognitive Tutoring: A Model of Help-Seeking with a Cognitive Tutor. *Int J Artif Int in Ed*, *16*, 101-130.

[2] Aleven, V., Stahl, E., Schworm, S., Fischer, F., Wallace, R. 2003. Help seeking and help design in interactive learning environments. *Rev Educ Res*, *73*(3), 277-320.

[3] Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. 2004. Off-Task Behavior in the Cognitive Tutor Classroom: When Students Game The System. *Proc ACM CHI*, 383-390.

[4] Baker, R.S.J.d., de Carvalho, A.M.J.A., Raspat, J., Aleven, V., Corbett, A.T., Koedinger, K.R. 2009. Educational Software Features that Encourage and Discourage "Gaming the System". *Proc 14th Int Conf Artif Int in Ed*, 475-482.

[5] Baker, R.S.J.d., Moore, G., Wagner, A., Kalka, J., Karabinos, M., Ashe, C., Yaron, D. 2011. The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software. *Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction*.

[6] Beck, J., & Gong, Y. 2013. Wheel-spinning: Students who fail to master a skill. *Arti Int in Ed*. Berlin: Springer.

[7] Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, 289-300.

[8] Broderick, Z., O'Connor, C., Mulcahy, C., Heffernan, N. & Heffernan, C. 2011. Increasing Parent Engagement in Student Learning Using an Intelligent Tutoring System. *J Int Learn Res*, 22(4):523-550.

[9] Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. 2009. Effectiveness of reading and mathematics software products: Findings from two student cohorts. Washington, DC: U.S. Dept Ed, Inst Ed Sci.

[10] Corbett, A. T., & Anderson, J. R. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Model User-Adap Intrc*, *4*(4), 253-278.

[11] Heffernan, N., & Heffernan, C. 2014. The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning & Teaching. *Int J AIED* 24(4),470-97.

[12] Heiner, C., Beck, J., & Mostow, J. 2004. Improving the help selection policy in a Reading Tutor that listens*. InSTILL/ ICALL Symposium.*

[13] Inventado, P.S. & Scupelli, P. 2015. Data-Driven Design Pattern Production: A Case Study on the ASSISTments Online Learning System. *Proc 20th Euro Conf Pattern Languages of Programs*.

[14] Inventado, P.S., Scupelli, P., Van Inwegen, E.G., Ostrow, K.S., Heffernan, N., Baker, R.S., Slater, S., & Ocumpaugh, J. 2015. Materials for *Hint Availability Slows Completion Times in Summer Work*. Retrieved from https://goo.gl/xyli5h

[15] Koedinger, K.R., & Aleven, V. 2007. Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educ Pscyhol Rev* 19.3: 239-264.

[16] Li, S., Xiong, X., & Beck, J. 2013. Modeling student retention in an environment with delayed testing. *Int Educ Data Mining Society*, 328-329

[17] Natl Gov Assoc Ctr Best Practices, Council of Chief State School Officers. 2010. Common Core State Stds. Washington D.C.

[18] Nelson-Le Gall, S. 1987. Necessary and unnecessary help-seeking in children. *J Genetic Pscyhol*, 148, 53–62.

[19] Ocumpaugh, J., Baker, R.S, Rodrigo, M.M.T. 2015. *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical & Training Manual.* NY, NY: Teachers College, Columbia U. Manila, Philippines: Ateneo Laboratory for the Learn Sciences.

[20] Pane, J.F., McCaffrey, D.F., Slaughter, M.E., Steele, J.L., & Ikemoto, G.S. 2010. An experiment to evaluate the efficacy of Cognitive Tutor geometry. *J Res Educ Eff*, 3(3), 254–281.

[21] Pardos, Z., Baker, R.S., San Pedro, M.O.C.Z., Gowda, S., Gowda, S. 2014. Affective states & state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. *J Learn Anlytc*, 1(1), 107-28.

[22] Razzaq, L.M., & Heffernan, N.T. 2009. To Tutor or Not to Tutor: That is the Question. *AIED,* 457-464.

[23] San Pedro, M.O., Baker, R., Heffernan, N., Ocumpaugh, J. 2015. Exploring College Major Choice and Middle School Student Behavior, Affect and Learning: What Happens to Students Who Game the System*? Proc 5th Int Learn Anlytc Know*, 36-40.

[24] Schofield, J. W. 1995. Computers and Classroom Culture. *Cambridge University Press*.

[25] VanLehn, K. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educ Psychol,* 46(4), 197-221.

[26] Whorton, S. 2013. Can a computer adaptive assessment system determine, better than traditional methods, whether students know mathematics skills? MA thesis, Computer Science Department, Worcester Polytechnic Institute.

[27] Wijekumar, K., Meyer, B., & Lei, P. 2012. Large-scale RCT with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educ Tech Res Dev*, 60(6), 987-1013.