Impact of a Sketch-Based Tutoring System at Multiple Universities

Dr. Vimal Kumar Viswanathan, San Jose State University

Dr. Vimal Viswanathan is an assistant professor in the Mechanical Engineering Department at San Jose State University. His research interests include design theory, design automation, design for X and engineering education. His engineering education work includes the application of brain-based learning protocols in engineering education, technology-assisted education, problem-based learning, and improving spatial visualization skills.

Josh Taylor Hurt

Josh Hurt is a first year graduate research student in the George W. Woodruff School of Mechanical Engineering at the Georgia Institute of Technological, where he received his Bachelors degree in the spring of 2020 . He has been working in the IDREEM lab under the guidance of Dr. Julie Linsey for most of his career at the Georgia Institute of Technology. Josh has been a part of the research into Maker Spaces and Engineering Education conducted within the IDREEM lab, and is currently focusing on Engineering Education.

Dr. Tracy Anne Hammond PhD, Texas A&M University

Director of the Sketch Recognition Lab and Professor in the Department of Computer Science and Engineering at Texas A&M University, Dr. Hammond is an international leader in sketch recognition, haptics, intelligent fabrics, SmartPhone development, and computer human interaction research. Dr. Hammond's publications on the subjects are widely cited and have well over a thousand citations, with Dr. Hammond having an h-index of 23, an h10-index of 65, and multiple papers with over 200 citations each. Her research has been funded by NSF, DARPA, Google, and many others, totaling over 9 million dollars in peer reviewed funding. She holds a PhD in Computer Science and FTO (Finance Technology Option) from MIT, and four degrees from Columbia University: an M.S in Anthropology, an M.S. in Computer Science, a B.A. in Mathematics, and a B.S. in Applied Mathematics. Prior to joining the TAMU CSE faculty Dr. Hammond taught for five years at Columbia University and was a telecom analyst for four years at Goldman Sachs. Dr Hammond is the 2011-2012 recipient of the Charles H. Barclay, Jr. '45 Faculty Fellow Award. The Barclay Award is given to professors and associate professors who have been nominated for their overall contributions to the Engineering Program through classroom instruction, scholarly activities, and professional service.

Dr. Benjamin W Caldwell, LeTourneau University

Benjamin W. Caldwell, Associate Provost for Academic Administration and Associate Professor of Mechanical Engineering at LeTourneau University, earned his B.S. (2007), M.S. (2009), and Ph.D. (2011) degrees from Clemson University, each in Mechanical Engineering. Caldwell served for five years in the Department of Mechanical Engineering at LeTourneau University where he taught courses in statics, dynamics, mechanics of materials, engineering design, and mechanisms and kinematics, among others, and conducted research in the areas of design and education. His research interests include validation of design methods, design creativity, idea generation, and engineering education. Caldwell collaborates with a variety of disciplines, including psychology, education, and theology, in both teaching and research, and has received both internal and external funding for this work. Caldwell also served as president of the teaching faculty at LeTourneau before moving to an administrative role. He is a member of ASME and ASEE.

Dr. Kimberly Grau Talley P.E., Texas State University

Dr. Kimberly G. Talley is an associate professor in the Department of Engineering Technology, Bobcat Made Makerspace Director at Texas State University, and a licensed Professional Engineer. She received her Ph.D. and M.S.E. from the University of Texas at Austin in Structural Engineering. Her undergraduate degrees in History and in Construction Engineering and Management are from North Carolina State University. Dr. Talley teaches courses in the Construction Science and Management and Civil Engineering Technology Programs, and her research focus is in student engagement and retention in engineering and engineering technology education. Contact: talley@txstate.edu

Dr. Julie S. Linsey, Georgia Institute of Technology

Dr. Julie S. Linsey is an Associate Professor in the George W. Woodruff School of Mechanical Engineering at the Georgia Institute of Technological. Dr. Linsey received her Ph.D. in Mechanical Engineering at The University of Texas. Her research area is design cognition including systematic methods and tools for innovative design with a particular focus on concept generation and design-by-analogy. Her research seeks to understand designers' cognitive processes with the goal of creating better tools and approaches to enhance engineering design. She has authored over 150 technical publications including over forty journal papers, and ten book chapters.

A Study on the Impact of a Sketch-based Tutoring System in Statics Instruction

Abstract

Large class sizes in engineering programs often prevent instructors from providing detailed and meaningful feedback to students on their homework problems. While the literature shows that frequent and immediate formative feedback has several benefits in terms of knowledge gain and academic motivation, several instructors struggle to provide any feedback. Motivated by this inability, a sketch-based virtual tutoring system, named Mechanix, has been developed and implemented. Mechanix lets the students to sketch their freebody diagram on a virtual interface and the process involved is very close to using a pencil and paper. The system provides real-time feedback on the accuracy of their Freebody diagrams and the solution to the problem. This paper reports the implementation of Mechanix at two large public universities in the United States – Georgia Institute of Technology and Texas State University. Mechanix is used to solve specific assignments from each school that involve the use of freebody diagrams. Pre- and post- concept inventories are used to measure the improvements in the conceptual understanding of the students. The results show that students who solve their homework using Mechanix outperform their peers who do not in one school, whereas the results are similar across the two groups in the second school. The evaluation of the concept inventories shows that the students who used Mechanix has the same level of improvement in their conceptual knowledge compared to the control group.

Keywords: FBD, sketch, statics, dynamics, engineering education.

Introduction & Background

In engineering education, feedback on a student's work plays a vital role as shown by existing literature [1-5]. It can promote knowledge acquisition and motivate new learning [6, 7]. Formative feedback can defined as the information communicated to a student with the intention of modifying the student's thinking or behavior with respect to the learning [8]. A plethora of research shows the importance of providing formative feedback in education [8-11]. While homework can serve as a perfect opportunity for formative feedback, several instructors find it very difficult to provide meaningful feedback to their students due to the time commitment involved. This is especially true while teaching large classes. In such cases, the feedback is often binary (correct or incorrect) along with a summative feedback from the exams.

While the research on timing of feedback is complex and often contradictory, a meta-analysis of studies comparing immediate and delayed feedback [12], concluded the overall superiority of the immediate feedback. When the feedback is delayed, students may have moved onto new content and the feedback may be irrelevant to them. The benefit of immediate feedback has been found across many populations, including undergraduate college students [13].

One of the fundamental concepts that mechanics instructors deal with is that of freebody diagrams (FBDs). FBDs allows to represent rigid bodies in a two-dimensional format along with the external forces acting on them. Typically introduced in the Statics course, the ability to create accurate FBDs is a key skill for their success in many subsequent classes. However, several students find it challenging to create accurate and complete FBDs which may eventually lead to their inability to learn more complex concepts in their future classes.

In large Statics classes, instructors often rely on web-based textbook platforms where homework is graded automatically by the system. Often, these systems grade the final numerical answer without providing any meaningful feedback on the method followed by the student. While this reduces the burden on the instructor, the lack of meaningful formative feedback affects the student's ability to learn the concepts. This is especially true when the student needs to draw an FBD to solve the problem. While some instructors encourage students to submit their FBD either in a hardcopy form or an uploaded scanned copy form, the feedback that the students receive is often delayed.

Mechanix is a sketch-recognition based tutoring system that allows students to hand-draw solutions just as they would with pencil and paper, while also providing iterative real-time personalized feedback (some references are removed to maintain anonymity). Sketch recognition algorithms use artificial intelligence to identify the shapes, their relationships, and other features of the sketched student drawing. Other AI algorithms then determine if and why a student's work is incorrect, enabling the tutoring system to return immediate and iterative personalized feedback facilitating student learning that is otherwise not possible in large classes, while saving instructor time. As it uses freehand sketching of FBDs, as they do with a pen and paper, the emphasis is put on learning the concepts, not on learning a new software tool, as several of the existing FBD tools do. In such cases, the students become experts in using the software tool while failing to learn the concepts they are expected to.

Currently, the effectiveness of the Mechanix platform is being tested in Statics courses at multiple universities across the country. This paper reports the results of the implementation of Mechanix at two large public universities. Specifically, the paper looks at the performance of the participating students in their homework and established concept inventories.

What is Mechanix?

Mechanix is a sketch interface that uses a sketch-recognition algorithm developed at Texas A&M University. In this platform, students can draw FBDs on a virtual interface as they would with a pencil and paper. Figure 1 shows an example truss problem in Mechanix with a corresponding FBD drawn by a student. When Mechanix recognizes the shape of the body, it creates nodes at the points where the forces are expected to be entered. The students are expected to draw arrows representing forces at the appropriate locations. In addition to recognizing the sketches and forces drawn, Mechanix compares the solutions entered by the students with pre-loaded solutions and provides immediate and detailed feedback.

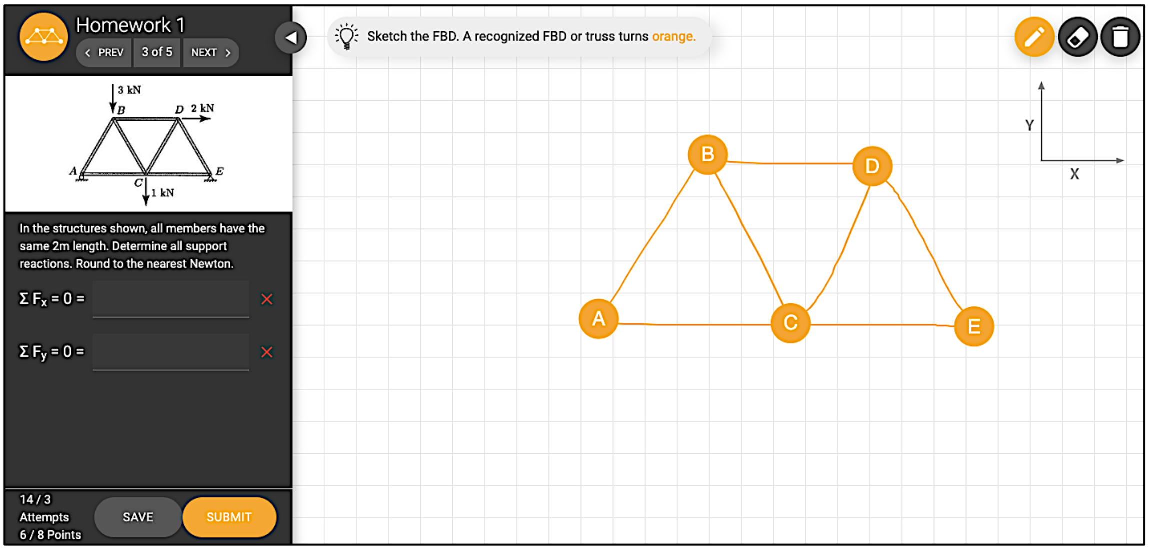

Figure 1: Example problem with hand-drawn FBD in Mechanix

A primary goal of Mechanix is increased and engaged student learning. Since the system is online, continuous learning opportunities are presented to the students. They are allowed to correct any errors in their work and resubmit until the entire content is correct, thus all of the objectives are learned, with the instructor knowing exactly where students had the most difficulty. Instructors are also able to assign open-ended truss design questions in which there are many possible solutions. Such questions encourage student creativity, and such free-form questions increase teacher's knowledge of student comprehension. Additionally, Mechanix can be made to be useful for a test scenario, as the iterative correction process allows students to learn during a test. An advantage of having a freehand sketch-based interface for learning truss analysis, and truss FBDs is that it takes the focus off of learning the simulation tool and puts the focus back on learning the concepts behind truss systems.

Currently, Mechanix can correct three different types of static homework problems: (1) Free-form free body diagrams such as that in Figure 2, (2) Planar truss problems requiring calculations of method of joints such as in Figure 1, and (3) Creative design problems such as in Figure 3. In Creative Design Mode, the student must think creatively to create a viable truss that abides by the constraints. Creative Design Mode offers a large number of possible solutions for the student, and thus presents an interesting recognition problem for the software.

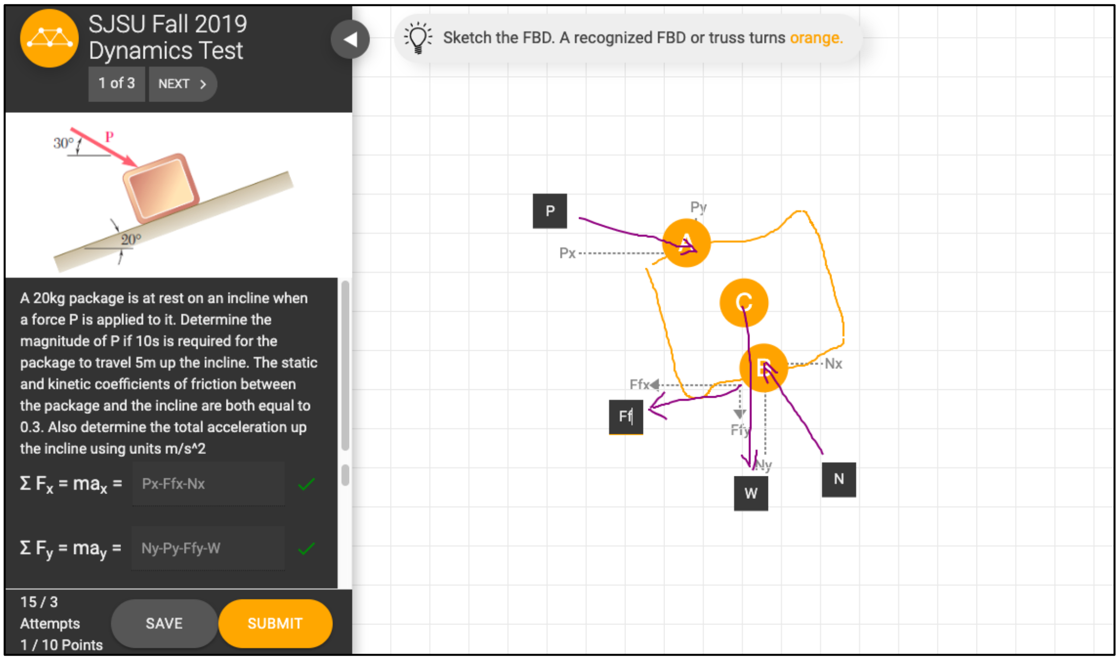

Figure 2: Example of a freeform FBD problem in Mechanix

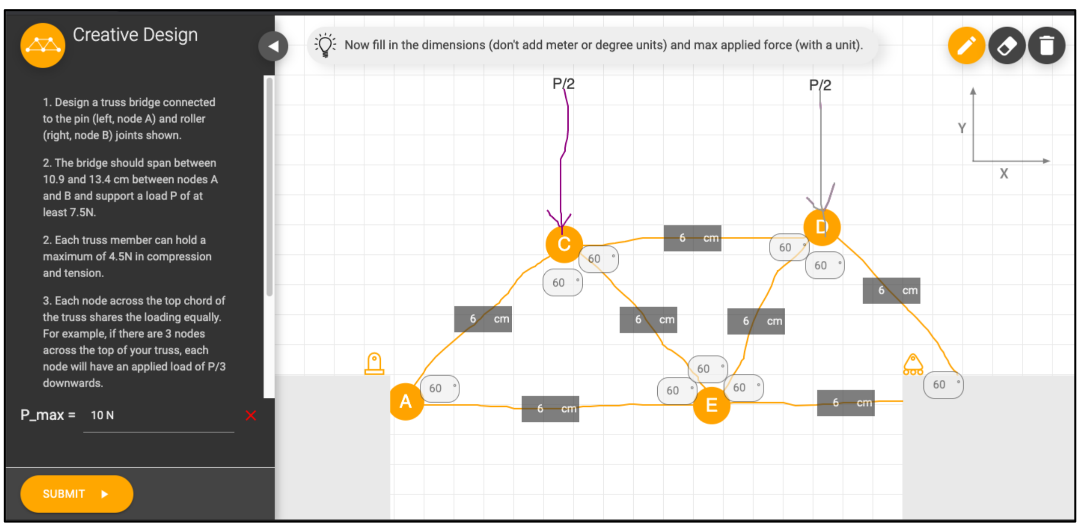

Figure 3: Example of a creative design problem in Mechanix

Additionally, Mechanix contains three different interfaces: (1) the student interface, where the student answers the problem, (2) the instructor question creation interface, and (3) the instructor review mode, where the instructor reviews the existing solutions. To add questions, the instructor simply types the question, uploads an image, sketches the answer, and types in the numerical answers. The drawn answer is then compared to the student's answer for correction. Since the student needs to know where he or she is wrong, Mechanix performs sophisticated analysis on the student's solution in an attempt to determine where the student has gone wrong. In the case of Creative Design Mode, no solution is drawn, instead only constraints are specified, and Mechanix then uses these constraints to grade the student's solution. Two types of feedback are given, that of a simple dropdown box, and that of a complete checklist for them to follow. Mechanix provides two types of feedback so that the instructor can allow the software to provide more feedback on initial problems and less later, scaffolding the feedback.

Research Questions

This study explores the potential benefits of using Mechanix platform for solving FBD problems in Statics courses. Specifically, the study investigates the following two questions:

RQ1: Can students using Mechanix as a platform to solve their FBD homework outperform their peers who do not use Mechanix?
RQ2: Does the use of Mechanix affect the conceptual knowledge gains in a Statics course, as measured using established concept inventories?

Method

To evaluate the learning gains while using Mechanix, empirical studies are conducted at multiple universities. This paper reports the results from two universities – Georgia Institute of Technology (School 1) and Texas State University (School 2). Both universities are large public universities with large class sizes. In School 1, "Statics" course is used as the testbed, while in School 2, a "Structural Analysis" course is used. The study is conducted across two semesters in School 1– the "control" semester where the students completed their homework using the tools they typically employ and a "experimental" semester where the students solved some specific homeworks using the Mechanix interface. The homework problems used remain the same across the two semesters. In School 2, the experimental setup was very similar, but the two experimental groups ran concurrently. Two sections of the Structural Analysis course offered in the same semester and taught by the same instructor were used for the data collection.

The primary data collection was performed from selected homework assignments that the participants completed on Mechanix (in the experimental group). These homework assignments required the students to draw FBDs and use the information to further solve the problem. Traditionally, both schools used a learning management system to administer these homework and collect the student responses. The control group completed the selected homework using the same learning management system. The experimental group received the same homework problems, but they were also given access to Mechanix. They were expected to draw their FBDs in Mechanix and submit the complete solutions to the problems using Mechanix. Mechanix also graded their work.

As part of the study participants completed a Statics Concept Inventory (SCI) in full and an abridged Force Concepts Inventory (FCI) at the beginning and at the end of their involvement in the study. The intent of these concept inventories is to measure a participants understanding of the material at the beginning of the class and then re-measure their understanding at the end of the semester long class. The material found within the Force Concept Inventory is concepts that students are expected to know leading into the class. The content on the SCI is expected to have lower initial scores, while an improvement in the score is expected for participants that complete the concept inventory at the end of the semester.

Results & Discussion

Performance in Homework

The study considers two homework assignments given to students at School 1, with both the Mechanix and Control groups completing the same problems on two separate systems. The experimental group completes the assignments on Mechanix, while the control group completes the assignments on Wiley Plus. The first homework deals with a statics problem that needs an FBD to solve and the second problem uses the method of joints to solve for forces on a truss. The students that use the experimental condition are graded on the correctness of their answers, as well as the completion of the FBD. This provides them with more opportunities to lose points within the assignment, but also provides more feedback to the students in the form of missing points directly related to the assumed correctness of their FBDS.

For a data set to be considered for the study, a student had to show effort within their group, which was assumed to be participation in their assignment group. This effort was considered if they had both a score on the assignment that was not an outlier from the other assignments, as well as if they continued to complete assignments in the assigned group. It should be noted that due to how the experiment was setup, only the experimental group had the opportunity to attempt problems using both homework systems and therefore could be invalidated due to completing the assignment using the incorrect system (the system that they were not assigned to).

Within the data gathered, a total of 12 experimental data sets and 6 control sets were invalidated due to the lack of correct participation in the study. Between the two data sets at school 1, these numbers varied, as some participants had correct participation for one set, while failing to correctly participate on the later assignment causing part of their data to be invalidated. Each data set was considered on a case by case basis, meaning within a participant's case, failure to comply with the study for one assignment did not invalidate all of that student's work.

At School 1, the first assignment has 19 data sets that qualified within the experimental group, while 29 students qualified for the experimental group. Within the groups, the total standard variation within the experimental group was less than the control group, which is directly reflected in the total standard error of each groups. However, both groups performed comparably well, with the experimental group scoring an average of 80.10%, and the control group scoring an average of 82.87% (as shown in Figure 4). These two scores had no significant difference as shown through a two-tailed t-test. The results showed that there was no significant impact on a student's ability to complete assignments based on basic free body diagrams when using either type of homework system.

On the second assignment, a total of 14 data sets qualified for the study within the experimental group, while 23 students qualified within the control group. In this case, there was a clear distinction between the experimental and control groups, with the variations being more significant. For the experimental group, the average score was 94.66%, while for the control group, it was 78.26% (see Figure 4). These scores did have statistically significant differences, as shown through a two-tailed t-test assuming unequal variances. This showed that the experimental group outperformed the control group on more complex problems, especially among students that did not show as much ability to complete problems sets on more difficult concepts.

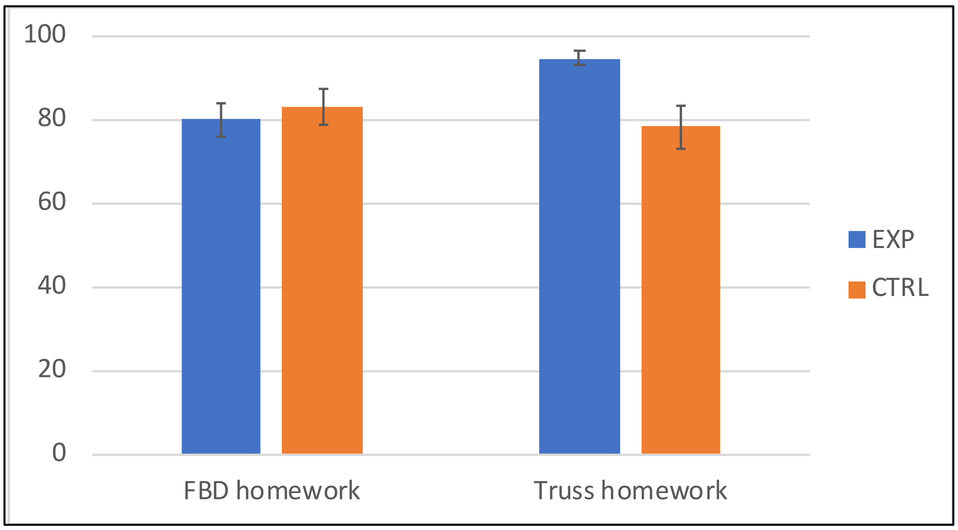

Figure 4: Homework average scores, percentage of correct responses in School 1. Homeworks were completed 2 weeks apart, with the control group using a separate online system and the experimental group using Mechanix. All error bars show (+ or -)1 S.E.

In School 2, data collection from homework was performed in two separate sections of the same class during the same semester. Each of these classes was assigned a homework condition, either experimental or control. Only one homework was considered for the study. Students that participated in the control condition were given several chances on assignments, allowing many of them to achieve higher grades on the assignment as each submission showed their overall grade prior to the final submission. Due to this, most of the control participants scored well on this assignment, while students that participated in the experimental condition did not score as much as the control. These students were given several submissions as well, but they were not shown their full credit on each submission - once they could view their grade, they had no opportunity to change it. Because of this difference, in the control condition most of the participants scored high on the assignment, while in the experimental group the scores were distributed evenly from high to low.

In both experimental conditions, there were a few students who chose to ignore the assignment. After counting these students out, the final sample sizes were 32 for the control group, and 31 for the experimental group. It was observed that the average scores of the two groups were comparable. As shown in Figure 5, the control group outperformed the experimental group when comparing the final scores on the assignment. However, considering that the control group re-attempted the assignment in a significantly higher number of cases, this difference was not surprising. To make a meaningful comparison, the scores of all the participants in their first submission of the assignment was considered. As shown in Figure 5, this comparison showed that the two experimental groups performed at the same level. Based on a two-tailed t-test, no statistically significant differences were observed across the two conditions.

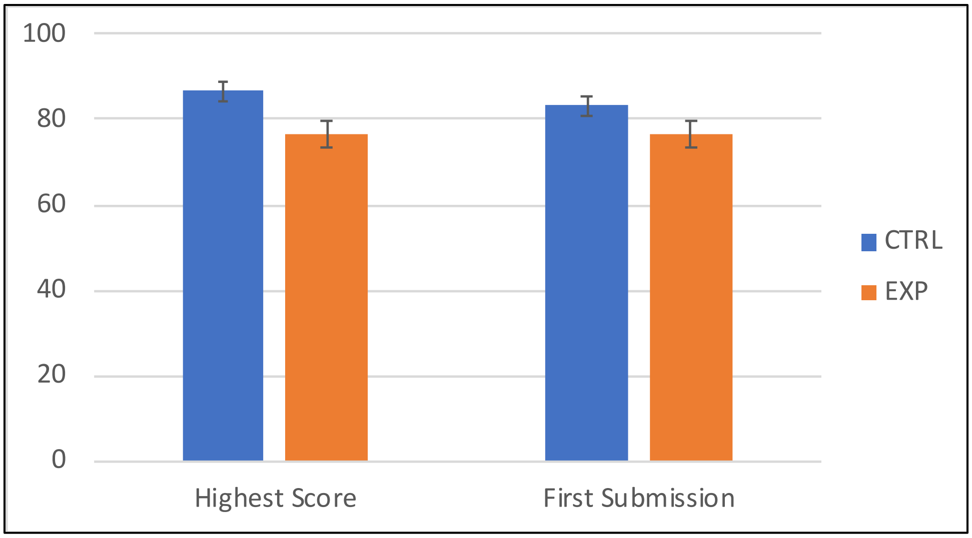

Figure 5: School 2 uses the Mechanix software for one homework assignment, where the control setting continues to use the online homework system that is used throughout the semester. All error bars show (+ or -)1 S.E.

It needs to be noted that the population of undergraduate students participating in the study at School 2 consists of many students that have more responsibilities than the students in School 1. The homework presented to the participants in the experimental groups of both schools can be considered a longer assignment, with less opportunities to get a higher score. It may be argued that the students in School 2 cannot spend as much time on Mechanix as their counterparts in School 1. This may contribute to the difference in results on the homework performance across the two schools.

Performance in Concept Inventories:

Students that participated in the study completed two concept inventories, Statics Concept Inventory (SCI) [14] and Force Concept Inventory (FCI) [15], at the beginning and end of the study. They were rewarded with a homework credit as a compensation for their participation. The SCI is a peer reviewed analysis tool, which is intentionally more difficult than most of the concepts given throughout the semester. In contrast, FCI is considered to be easier for a majority of the study participants, as the material is presented in an easier fashion, and tests them on their pre-requisite knowledge.

The results from the comparison of concept inventories are shown in Figures 6 & 7. Within each school, the students in both experimental and control groups at both schools performed almost identically in the concept inventories at the beginning of the semester. However, comparing across the two schools, there was a significant difference in the scores. School 1 outperformed School 2 on both the pre- and post- implementation of both concept inventories. In both schools, an improvement was observed in the scores from pre- to post- implementation of the concept inventories, but this improvement was dependent on the school. It was observed that the students in School 1 gained significantly more in their post-test scores compared to their counterparts in School 2.

This lack of differences in the pre-test SCI scores within a school's experimental and control groups showed that both groups had similar understanding of the material coming into the study, and saw a similar amount of growth through the completion of the study (Figure 6). The significant differences in scores across the two schools highlighted the demographic differences that were expected at the different universities. At the same time, it could be observed that the students who used Mechanix for their homework gained the same amount of conceptual knowledge as the other techniques they were used to.

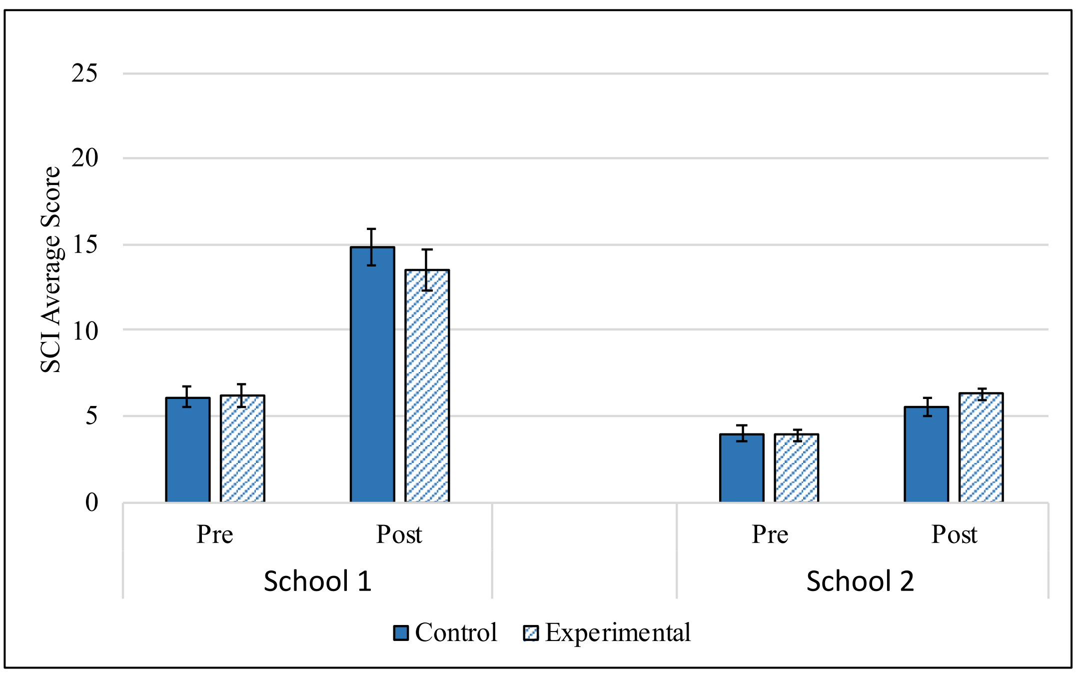

Figure 6: Overall Results of the Statics Concept Inventory Surveys at Georgia Tech and Texas State, resulting in statistically insignificant results between the experimental and control groups of that university. All error bars show (+ or -) 1 S.E.

The FCI scores can be compared pictorially in Figure 7. These results are as expected, where both the experimental and control groups found at a specific university performed almost identically on all of the surveys given. The difference in the scores across the two schools can be again attributed to the geographical differences.

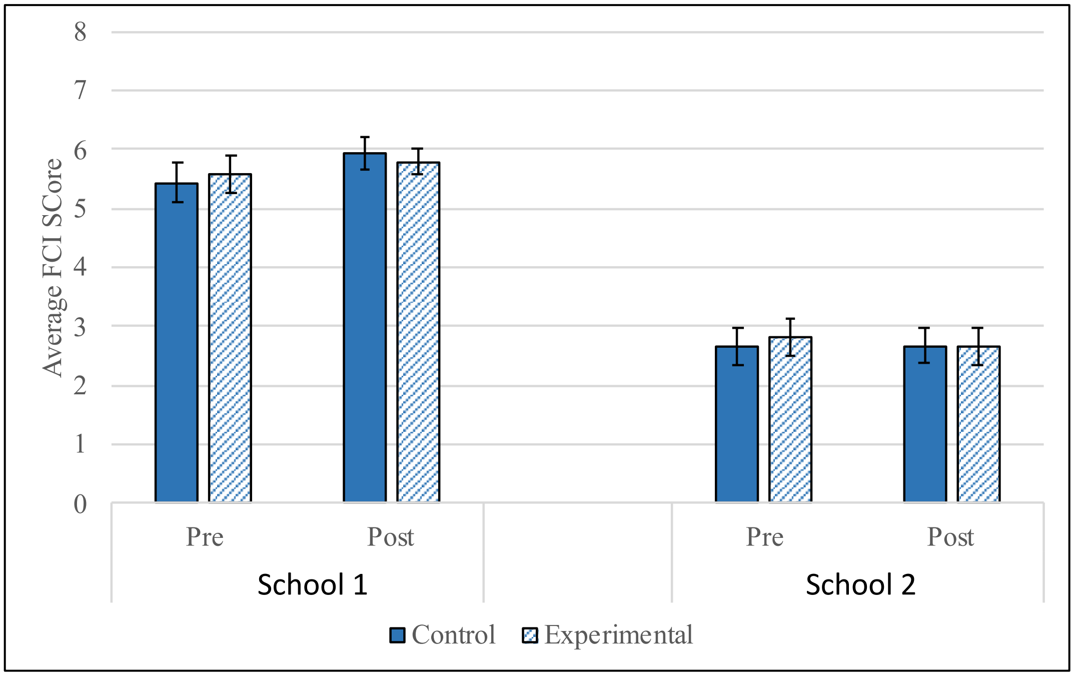

Figure 7: Pre and Post scores of both the experimental and control groups at both schools, showing the expected result of having similar understanding of the FCI at the beginning and end of the study. All error bars show (+ or -)1 S.E.

It should be noted that at both schools, participants had difficulties understanding some concepts on the SCI, specifically concepts that were not taught in the class. This understanding led to a minority of students choosing not to fully complete SCI. These attempts were not included in the overall results, as they did not reflect the overall understanding of students. Some students found the SCI to be more difficult than anything that had been taught during the semester and showed less effort overall on this study. Because of the vocal lack of support in the concept Inventories, a possible fix for these major problems has been implemented in the study for other participating universities. This being the reordering of the SCI questions, with the easiest ones at the beginning and a gradual increase in the difficulty as the students move on. The order of questions was directly determined by the average score of each question set at all participating universities. The results using the reordered SCI will be included in the future work.

RQ1: Can students using Mechanix as a platform to solve their FBD homework outperform their peers who do not use Mechanix?

Based on the results from the limited data in this study, it may be concluded that when students spend sufficient time on Mechanix, they may outperform their peers who do not using Mechanix. Additional data are needed to derive more concrete conclusions on this. The learning gains are more significant for complex problems, compared to simple FBD problems. In School 2, a good fraction of the students have other commitments and they are unable to spend extra time on their homework. This lack of time can explain why their learning gains are not as significant as in School 1.

RQ2: Does the use of Mechanix affect the conceptual knowledge gains in a Statics course, as measured using established concept inventories?
The comparison of the concept inventories shows some interesting results. It can be concluded that in each school, students who use Mechanix have the same level of conceptual knowledge gain as their peers who do not. When comparing across the two schools, there is a significant difference in the pre- and post- test scores on both concept inventories, which can be attributed to geographical differences in the schools.

Conclusions

The study reported in this paper aims to evaluate the effectiveness of Mechanix, a sketch-based tutoring system for freebody diagram problems. Mechanix provides real-time detailed feedback to students on the accuracy of their FBDs and the complete solution to the problem. The current study explored the effectiveness of this platform using homework and concept inventory data collected from two large public universities in the United States. The scores of students who completed selected homework assignments on Mechanix are compared against that of their peers who did not use Mechanix. The results show that Mechanix is an effective tool for solving FBD-based problems, but students do need to spend a sufficient amount of time using it. Further, the use of Mechanix helps students to gain the same extent of conceptual knowledge as the learning management systems in the schools, as measured by the concept inventories.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant numbers DUE- 1726047, 1725659, 1725423, 1726306, and 1725785. 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 National Science Foundation.

References

[1] R. Azevedo and R. M. Bernard, "A meta-analysis of the effects of feedback in computer-based instruction," *Journal of Educational Computing Research,* vol. 13, no. 2, pp. 111-127, 1995.

[2] R. L. Bangert-Drowns, C.-L. C. Kulik, J. A. Kulik, and M. Morgan, "The instructional effect of feedback in test-like events," *Review of educational research,* vol. 61, no. 2, pp. 213-238, 1991.

[3] A. Corbett and J. R. Anderson, "Feedback timing and student control in the LISP Intelligent Tutoring System," in *Proceedings of the Fourth International Conference on AI and Education*, 1989, pp. 64-72.

[4] M. L. Epstein *et al.*, "Immediate feedback assessment technique promotes learning and corrects inaccurate first responses," *The Psychological Record,* vol. 52, no. 2, p. 5, 2010.

[5] D. R. Pridemore and J. D. Klein, "Control of practice and level of feedback in computer-based instruction," *Contemporary Educational Psychology,* vol. 20, no. 4, pp. 444-450, 1995.

[6] M. R. Lepper and R. W. Chabay, "Intrinsic motivation and instruction: Conflicting views on the role of motivational processes in computer-based education," *Educational Psychologist,* vol. 20, no. 4, pp. 217-230, 1985.

[7] S. Narciss and K. Huth, "How to design informative tutoring feedback for multimedia learning," *Instructional design for multimedia learning,* vol. 181195, 2004.

[8] V. J. Shute, "Focus on formative feedback," *Review of educational research,* vol. 78, no. 1, pp. 153-189, 2008.

[9] T. R. Guskey, "Reporting on student learning: Lessons from the past-Prescriptions for the future," *ASSOCIATION FOR SUPERVISION AND CURRICULUM DEVELOPMENT-YEARBOOK-,* pp. 13-24, 1996.

[10] J. B. Nyquist, "The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis," Vanderbilt University, 2003.

[11] G. Smith, "How does student performance on formative assessments relate to learning assessed by exams?," *Journal of College Science Teaching,* vol. 36, no. 7, p. 28, 2007.

[12] J. A. Kulik and C.-L. C. Kulik, "Timing of feedback and verbal learning," *Review of educational research,* vol. 58, no. 1, pp. 79-97, 1988.

[13] G. Gibbs and C. Simpson, "Conditions under which assessment supports students' learning," *Learning and teaching in higher education,* no. 1, pp. 3-31, 2005.

[14] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," The physics teacher, vol. 30, no. 3, pp. 141–158, 1992.

[15] P. S. Steif and J. A. Dantzler, "A statics concept inventory: Development and psychometric analysis," Journal of Engineering Education, vol. 94, no. 4, p. 363, 2005.