Pilot Study Results from Using TrussVR© to Learn About Basic Trusses

Ryan Banow, University of Saskatchewan

Ryan Banow is an Educational Development Specialist at the Gwenna Moss Centre for Teaching and Learning at the University of Saskatchewan. He is also a PhD student in the within the College of Education at the University of Saskatchewan. He has worked as an Educational Developer since 2012 and has taught as a Sessional Lecturer since 2014. He is currently the chair of the University of Saskatchewan's Instructional Design Group. His educational background includes a BSc (Math), a BEd (Secondary Math and Physics), and an MEd (Educational Technology and Design). He is passionate about teaching and has a variety of research interests around the central theme of enhancing teaching and learning. He has worked, presented, and published on research in the fields of STEM education, educational technology, virtual reality, the Scholarship of Teaching and Learning, and others.

Dr. Sean Maw P.Eng., University of Saskatchewan

Dr. Maw currently holds the Huff Chair in Innovative Teaching in the College of Engineering at the University of Saskatchewan. In this capacity, his work focuses on learning facilitation methods especially as they pertain to engineering design. He earned his BASc and MASc degrees in Systems Design Engineering from Waterloo, and a PhD in Neuroscience from the University of Alberta. His current research interests lie in sports engineering and engineering pedagogy.

Pilot Study Results from Using TrussVR© to Learn About Basic Trusses

Abstract

TrussVR©, a custom-designed VR application, was developed to help engineering students learn about basic trusses in a virtual lab environment. Trusses are a mainstay of many first-year engineering Statics courses. They are relatively simple to analyze. However, hand calculations are typically time-consuming. As a result, most textbook problems involve evaluating one loading scenario and end with the calculated values of forces running through the truss's two-force members (2FMs). This scenario does not lend itself to a holistic understanding of how trusses behave under loads of various magnitudes and locations. It does not facilitate a comparison of the relative strengths and weaknesses of different truss designs, nor a constructivist learning style driven by curiosity.

TrussVR© carries out the computations of solving a truss almost instantaneously. What this affords is a new way to learn about trusses, and a way to learn features of trusses that have been previously impractical to learn through conventional lab techniques. Build a truss, apply an external force, and see the distribution of forces within the truss. This cycle can be repeated quickly in VR, allowing learners to gain an enhanced visual appreciation of how trusses behave.

In the 2018/19 academic year, 166 first-year engineering students at the University of Saskatchewan volunteered for a pilot study that examined the efficacy of TrussVR©. The study was approved by the U of S Behavioural Research Ethics Board. Volunteers completed a "pre-survey" that examined knowledge, skills, and attitudes regarding trusses, before trusses had been taught in the classroom. Participants were then randomly assigned to one of three groups: no lab, regular lab, and TrussVR© lab. All participants then attended 2 weeks of in-class lectures on trusses, frames, and machines, as they normally would. During this time, the "no lab" group did not complete an experimental lab on trusses. The "regular lab" group completed the same 3-hour lab as the rest of the course's non-participants. This involved the evaluation and construction of a physical truss using a Pasco© construction set. The "TrussVR©" group went through a 7-10 minute tutorial and then a 20 minute VR lab experience consisting of i) examining and playing with a variety of truss types, ii) solving a number of skill testing problems involving basic trusses, and iii) building a bridge and testing it. At the end of the 2 week unit on trusses, frames, and machines, participants in all groups completed a mixed-methods "post-survey" that re-examined their knowledge and skills, as well as their impressions of the TrussVR© and Pasco© systems, if they used one. A "distant post-survey" was conducted two months later followed by a "very distant post-survey" another seven months after that, to further re-examine knowledge, skills, and attitudes.

Statistically significant differences were found between TrussVR© and the other groups ($p < .01$) for recognition and recall of truss types two months and nine months afterwards. Likewise, the virtual lab experience was highly rated in most respects.

1.0 Introduction

About two and a half years ago, the authors of this study had the opportunity to engage in some virtual reality (VR) demonstrations. Emerging from that experience with a belief that part of the future of engineering education lay in the application of VR for teaching and learning, a software development project was undertaken with a local start-up (www.sprockety.com). The result of that project was TrussVR©, a piece of VR software that teaches students about trusses. Starting in the Fall of 2018, a study was carried out with volunteer first-year students to evaluate the potential efficacy of TrussVR©. This paper describes most of the results of that study.

The decision to focus on trusses in a first attempt at developing VR software for education was based on two main factors. One of the authors teaches first-year Statics to engineering students. Using a conventional teaching approach, truss units are generally regarded as relatively fun and easy. However, anecdotally, it appears that students do not have a "deep" grasp of the nuances of trusses. This is because students take approximately 30 minutes to solve a simple truss only to determine the values of the members in tension or compression. Then they move on. There is little appreciation for the fact that a truss will behave quite differently for different loads. Indeed, most solutions are simply numerical and do not involve a holistic perspective on the overall patterns of internal forces within the truss. Any sort of iteration for evaluating multiple loads on a truss is very laborious and time-consuming, which discourages exploration.

The other main factor was a consideration of what VR is good for, in education. VR's value lies in its abilities to reduce risk, reduce costs, and accelerate experimental iteration. As such, the truss application was a good choice because VR released the burden of doing the math, allowing students to focus on the structural behaviour of the trusses. Much like playing with a children's building set, it also offered a virtual building environment that most people could relate to.

The constructivist approach to learning (e.g. [8] and [11]) posits that we build understanding, starting from our current understanding of the world and then adding on new scaffolding. The traditional approach to learning trusses is a very slow approach which may obscure more holistic understanding due to a focus on numerical answers. If a student was able to build a truss, apply a force, and then instantly see how the truss' members reacted to that applied load, there would be enormous potential for iterative exploration and experimentation. Understanding could almost literally be built with the trusses.

This notion guided the development of TrussVR©. Ultimately, TrussVR© contained three separate activities or modules. The first was a guided exposure to different types of trusses. TrussVR© users are shown a shelf full of model trusses. They can select one, and then "play" with it (where "play" means they can apply external loads, see the resulting internal loads, and then modify the loads to see how the truss behaviour changes). If they spend at least 30 seconds on a truss, they get a "point". They can get up to 6 points in the exercise by looking at and playing with at least 6 different truss types. Figure 1 shows a Polynesian truss in TrussVR©.

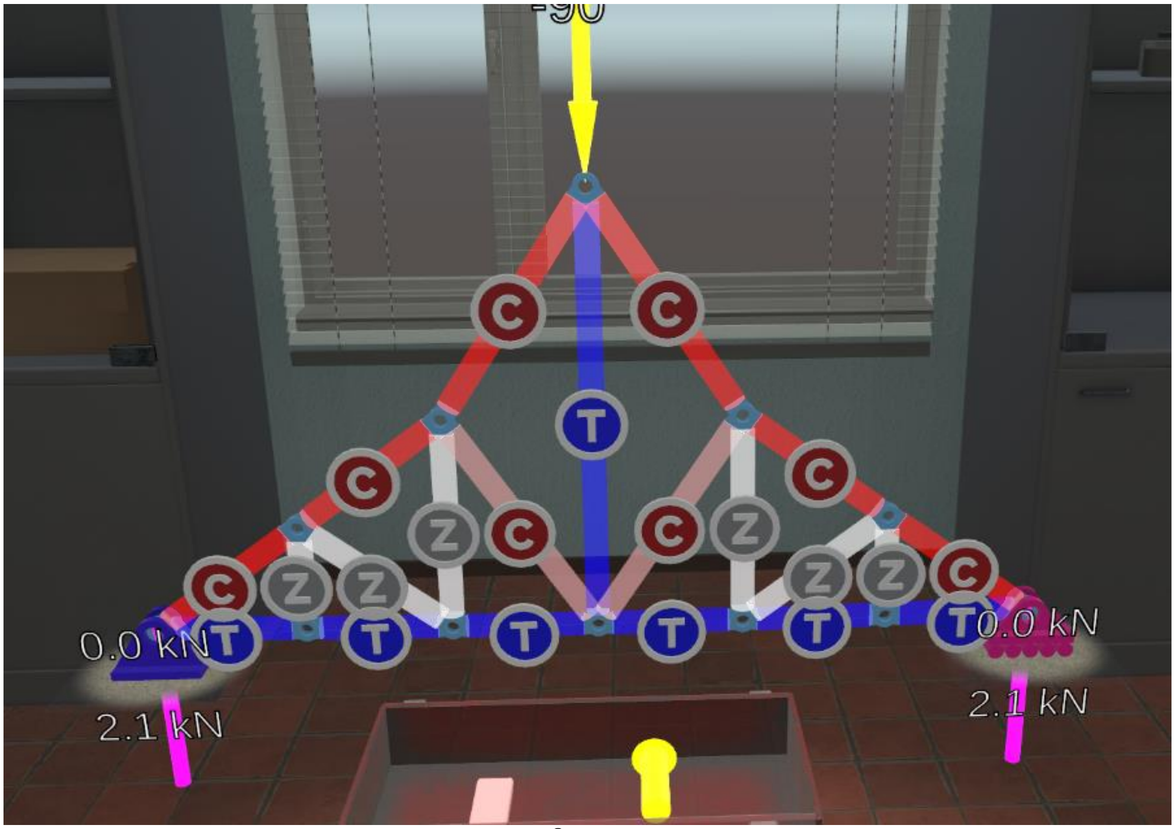

Figure 1: Polynesian truss in TrussVR©.

The second module consists of two parts. In the first part, students are shown a simple truss with the members labeled for tension, compression or zero force. Students are then asked to apply an external load at a joint that generates these internal forces. Rapid feedback tells them how close they are to the correct answer. Indeed, users are shown the internal forces created by their guess and they receive a coaching comment ("try another joint" or "try another direction of the applied force"), depending on the error made. They work through three problems of this sort. The second part of this second module, is the exact opposite. Students are shown a simple truss with an applied external load, and they have to label all of the truss members as being in tension, compression, or zero-force members (ZFMs). Once they have made a guess, they are told how many of the members are correct, but not which ones. They have three guesses to get the fully correct answer. They have 6 minutes for this module as well.

The third module is a construction game. Users have to create a truss bridge, and then drive a load across the bridge without it collapsing. If they are successful, they can increase the load for further testing and modify their bridge design. They have 7 minutes for this module.

Students try TrussVR© in a teaching lab created specifically for VR education. Outfitted with a number of VR stations, students book times to visit the VR classroom and do their truss lab. With an initial operating capacity of four students/hour (each getting 30 minutes VR time), throughput was not especially high. However, the evaluation of the student's performance was happening in real-time, so no further markers were required to evaluate the lab. As such, it was actually more economical on an operating basis, versus a paper-based lab report assessment.

In having a significant number of students try TrussVR© within the context of an existing course, we were able to explore the potential efficacy of TrussVR© from a variety of perspectives. Knowledge enhancement and retention was measured, as were truss analysis skills, as well as attitudes towards trusses. These were compared with other students that were taking part in the regular truss lab for the course, as well as a group of students that did not take part in any truss lab experience. By looking at knowledge, skills, and attitudes towards trusses before, immediately after, two-months after, and long after the truss unit (and lab experience) in the Statics course was over, a set of interesting results were gathered. This paper builds on the initial results previously reported in [2].

2.0 Methods

2.1 Recruitment

The participants in this study were recruited from the students enrolled in the Fall 2018 offering of first-year GE 124 Engineering Mechanics 1 (Statics). There were 440 students enrolled. A member of our team visited the class within the first month of the term and explained the study to the students and asked those interested in taking part to complete an online sign-up survey on SurveyMonkey. We received responses from 228 students.

2.2 Pre Survey

About one month later, and three weeks prior to trusses being taught in the class, we emailed a SurveyMonkey link to the Pre Survey to all of the signed-up participants. The Pre Survey contained a consent form and 45 questions in total. There was a mix of closed- and open-ended questions. There were questions on demographics, prior experiences with VR, Likert scale questions on interest, confidence, and excitement about learning trusses, a question asking participants to name as many trusses as possible, and a series of skill-testing questions to assess their prior knowledge and skills regarding trusses. The skill-testing questions were of five main types: identify the type of truss from a list of names based on an image of the shape (7 questions); given a diagram of a simple structure, identify if it is a truss or not (3 questions); given a diagram of a truss and an internal pattern of tension and compression, identify the joint and direction of the external force (3 multi-part questions); and given a diagram of a truss with an external force acting on a joint, identify whether each member is in compression, in tension, or is a zero-force member (3 multi-part questions). Since the students had not yet covered trusses in the course, it was expected that most students would have little prior knowledge of them. A total of 166 students completed the entire Pre Survey.

2.3 Lab Experiences

After the Pre Survey closed, we randomly assigned the participants to one of three experimental groups: Pasco©, No Lab, and TrussVR©. The Pasco© group would complete the regular truss lab along with the students in the course that did not sign up to take part in the study. The Pasco© lab is 2-3 hours long and involves students using a physical construction set to build and test a truss design (see pasco.com), after having calculated its behaviour. The No Lab group did not complete a lab related to trusses. Their truss lab was omitted from calculations of the students' final grades in the course. The TrussVR© group engaged in a 25-min VR truss lab. This included a 5-7 min tutorial on how to use TrussVR©. During the same two-week time period that the truss labs were taking place, there were approximately four 80-minute lectures related to trusses, for all students.

The TrussVR© lab was conducted in software designed at the University of Saskatchewan in cooperation with Sprockety Ventures (www.sprockety.com). TrussVR© runs in Unity on HTC Vive Pro hardware. The goal of the software is to enhance the understanding of trusses and to allow users to quickly build and test truss designs. To recap, the 25-minute lab experience involved five main stages: a tutorial, examination and testing of common truss designs, identification of an external force that would create a given pattern of internal forces, identification of internal forces resulting from a given external force, and building a bridge. Students completed the lab in the Virtual Reality Teaching and Cloud Based Learning (ViRTCL) Lab facilities, as shown in Figure 2. Student grades on the lab were computed based on the latter four stages and were used as part of calculations of the final grades in the course. Biometric data was also collected on these participants as they used TrussVR©, primarily to assist in improving the software. Lab grades and biometric data are not explored in this paper.

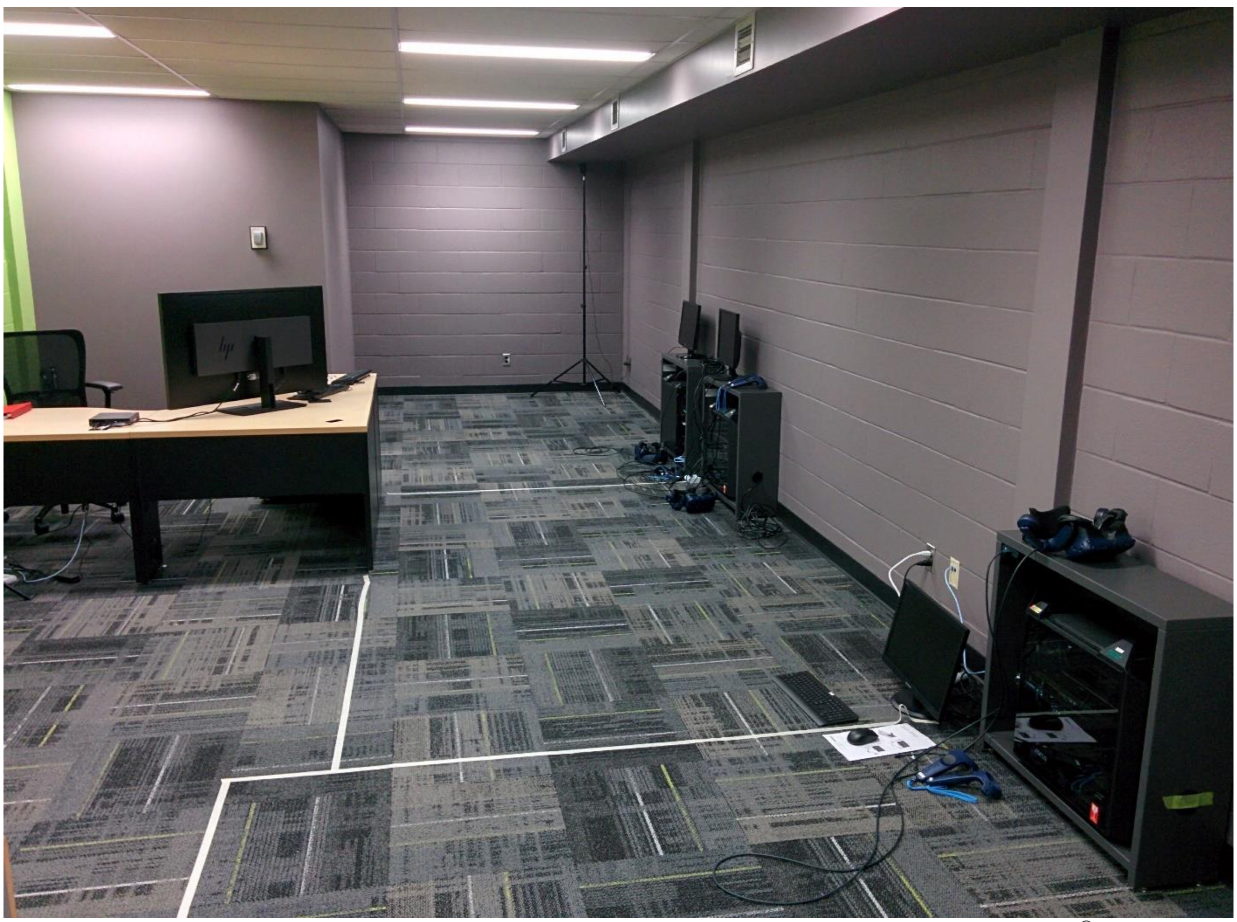

Figure 2: Virtual Reality Teaching and Cloud Based Learning (ViRTCL) Lab©.

2.4 Post Survey

Once all the Pasco© and TrussVR© labs were complete, participants were emailed an invitation and link to the Post Survey. Each group received a slightly different survey, as there were unique feedback questions related to experiences with each intervention. There were 31 common survey questions. The survey contained repeats of the Likert scale questions on interest, confidence, and excitement about learning trusses, a question asking participants to name as many trusses as possible, and a series of skill-testing questions. The skill-testing questions were all based on the same trusses as the Pre Survey, but the force locations and directions had changed i.e. these were isomorphic questions. For the students that were in the Pasco© and TrussVR© groups, there was a series of 11 Likert scale questions focused on the lab experiences e.g. "It was easy to see the forces acting in the truss" and "I would recommend the use of Pasco©/TrussVR© in teaching trusses". The Post Survey closed with a series of 8 open-ended questions asking the students' opinions on the approaches to the teaching of trusses e.g. "What did you enjoy most in learning trusses?" and "What is the hardest skill to learn when analyzing trusses?"

2.5 Distant Post Survey

Two months later, participants were emailed an invitation and link to the Distant Post Survey. This survey was 31 questions long and repeated the common questions from the Post Survey. Again, the skill-testing questions were based on the same trusses, but the forces had been changed. Between the Post and Distant Post Surveys, students also completed the final exam in the course. The final exam occurred six weeks prior to the Distant Post Survey. The final exam contained one truss problem and one frame problem.

Once the Distant Post Survey was closed, we offered participants in all three of the experimental groups the option to take part in any interventions that they missed. For example, No Lab group members were offered the opportunity to try TrussVR© and the Pasco© lab. In addition, those that completed all three of the Pre, Post, and Distant Post surveys were entered into draws for gift cards.

2.6 Very Distant Post Survey

Another seven months later i.e. nine months after the lab experiences, participants were emailed an invitation and link to the Very Distant Post Survey. Since this survey was so far removed from when the participants originally provided consent, the survey began with participants providing consent again. This survey was the same for all three groups and was 37 questions long. It repeated the common questions from the Post Survey. Again, the skill-testing questions were based on the same trusses, but the forces had been changed. On this survey, we doubled the number of skill-testing questions i.e. given a diagram of a truss and an internal pattern of tension and compression, identify the joint and direction of the externally applied force, and given a diagram of a truss with an external force acting on a joint, identify whether each member is in compression, in tension, or is a zero-force member. We increased the number of questions in an attempt to increase the reliability of the scores. When reported (as shown below), the scores on these questions on the Very Distant Post Survey have been halved to match the sums of the scores at the other time points. Participants that completed the Very Distant Post Survey were entered into another draw for gift cards.

The University of Saskatchewan Behavioural Research Ethics Board provided ethics clearance for the study. Since the Very Distant Post Survey was not in the original application, approval was sought and received to amend the original application and include the final survey.

2.7 Analysis of Multiple Choice Responses

The research questions were all difference questions that compared two or three experimental groups. To analyze the differences on the skill-testing questions, questions were grouped based on the sets of different types of questions and each question was weighted equally. These questions were of five main types: identify the type of truss from a list of names based on an image of the shape (7 questions); given a diagram of a simple structure, identify if it is a truss or not (3 questions); given a diagram of a truss and an internal pattern of tension and compression, identify the joint and direction of the external force (3 multi-part questions); and given a diagram of a truss with an external force acting on a joint, identify whether each member is in compression, in tension, or is a zero-force member (3 multi-part questions).

For these skill-testing questions, Post Gain (Post - Pre), Distant Post Gain (Distant Post - Pre), 2-Month Retention (Distant Post - Post), Very Distant Post Gain (Very Distant Post - Pre), and 9-Month Retention (Very Distant Post - Post) scores were calculated for each set of questions. This method allowed for analyses of differences between each group at each time point. This analysis method maintained the scores of each participant that responded at two compared time points e.g. Post Gain evaluated those that completed both the Pre Survey and the Post Survey. Another option would have been to analyze using repeated measures, but since the response rate was relatively low on the Distant Post and Very Distant Post, it was determined that using gain and retention scores would most effectively utilize the breadth of responses collected on the earlier surveys.

After exploring the data, it was found that in every analysis that would be performed, at least one set of values were non-normal. This was based on looking at the skewness and kurtosis values and the Shapiro-Wilk Test of Normality. This led to the use of non-parametric tests to analyze the differences. These questions were all analyzed using IBM SPSS Statistics 26.

2.8 Analysis of Open-Ended Responses

A series of open-ended questions were asked on the Post Survey. A grounded theory approach was taken to analyze the text responses and group them into themes. Each question was analyzed separately. For each question, the steps that were taken to arrive at clear conclusions included: 1) delete null/nonsense responses; 2) where applicable, break a response into core concepts/ideas; 3) sort the entire set of core concepts/ideas for all experimental groups into themes based on what was said, and not on any pre-conceived themes; and 4) count the number of responses corresponding to each theme and calculate what percentage of responses each theme accounts for within each experimental group.

3.0 Results

Results are presented below, beginning with the multiple choice questions followed by the open-ended questions. The multiple choice questions are presented in the following order: 1) questions on interest, confidence, and excitement; 2) skill-testing questions; and 3) impressions of the lab experience. The open-ended questions are presented in the following order: 1) what do you remember most about trusses, 2) compare two trusses, 3) what is the hardest skill to learn when analyzing trusses, 4) what was the most noteworthy thing that you gained from your lab experience, 5) what was the most surprising thing you learned about trusses, 6) what did you enjoy most about trusses, 7) what did you enjoy least about trusses, and 8) do you have any comments about your bridge-building experience. All effect size interpretations are based on [3].

3.1 Questions on Interest, Confidence, and Excitement

There were five questions on each survey regarding students' interest, confidence, and excitement about learning trusses. Responses were similar across all three experimental groups at the Pre, Post, and Distant Post time points. A series of Kruskal-Wallis non-parametric tests were conducted to test for statistically significant differences between the three experimental groups. This non-parametric test was chosen because the *n*s across the groups differed and every set of ratings failed the Shapiro-Wilk Test of Normality. The Kruskal-Wallis non-parametric tests indicated that none of the differences were statistically significant between the experimental groups at any of the examined time points.

3.2 Skill-Testing Questions

As described earlier, there were five types of skill-testing questions. Table 1 shows the Post Gain, Distant Post Gain, Two-Month Retention, Very Distant Post Gain, and Nine-Month Retention scores for the three study groups based on each question type, plus a straight sum of all of the scores together (Total Score). At the various time points, different groups scored best for the different types of questions. The TrussVR© group scored significantly higher on six of the measures while the Pasco© group scored significantly higher on one.

3.2.1 Identifying Truss Name by Shape

In general, as can be seen in Table 1, when asked to identify the name of a truss given an image of its shape, the TrussVR© group had better gain and retention scores than both other groups. The only exception was the 9-month retention score, which was slightly better for the other two groups. Figure 3 also shows the raw scores (not gain) for each group at each time point in this question. The TrussVR© group scored higher on each of the Post, Distant Post, and Very Distant Post surveys. A series of Kruskal-Wallis non-parametric tests were conducted to test for statistically significant differences between the three experimental groups. This non-parametric test was chosen because the *n*s across the groups differed and, in each case, at least one of the group's scores being analyzed failed the Shapiro-Wilk Test of Normality.

For Post Gain, the Kruskal-Wallis non-parametric test indicated that the three groups differed on being able to identify truss names based on seeing an image of the shape, χ^2 (2, N = 120) = 6.66, *p* = .04. Post hoc Mann-Whitney tests compared the three pairs of groups using a Bonferroni corrected *p* value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the TrussVR© group (48.27, n = 51) was significantly higher than that of participants in the Pasco© group (35.09, n = 34), z = -2.45, p = .014, r = -.27 i.e. a small-to-medium effect size. There was no statistically significant difference between the TrussVR© and No Lab groups or the No Lab and Pasco© groups.

Table 1: Gain and retention scores for skill-testing questions

Question Set (# of possible marks)	Group	Post Gain M ± SD (n)	Distant Post Gain M ± SD (n)	2-month Retention M ± SD (n)	Very Distant Post Gain M ± SD (n)	9-month Retention M ± SD (n)
Identify Name of Truss (7)	No Lab	1.4 ± 1.6 (n = 35)	0.9 ± 1.8 (n = 35)	-0.5 ± 2.0 (n = 35)	0.2 ± 1.6 (n = 29)	0.9 ± 1.6 (n = 25)
	Pasco©	0.7 ± 2.2 (n = 34)	0.3 ± 1.4 (n = 20)	-0.5 ± 1.5 (n = 16)	0.1 ± 1.3 (n = 14)	0.8 ± 1.6 (n = 13)
	TrussVR©	*1.9 ± 2.4 (n = 51)	*1.8 ± 2.9 (n = 35)	-0.4 ± 2.5 (n = 35)	*2.0 ± 2.0 (n = 28)	0.3 ± 2.1 (n = 28)
Identify if Truss or Not (3)	No Lab	1.2 ± 1.1 (n = 35)	0.9 ± 1.3 (n = 35)	-0.3 ± 0.5 (n = 35)	0.8 ± 1.1 (n = 29)	0.6 ± 1.0 (n = 25)
	Pasco©	1.5 ± 1.2 (n = 33)	1.2 ± 1.4 (n = 20)	-0.2 ± 0.8 (n = 16)	*1.6 ± 1.2 (n = 14)	0.2 ± 1.4 (n = 13)
	TrussVR©	1.1 ± 1.2 (n = 51)	0.9 ± 1.1 (n = 35)	-0.3 ± 0.6 (n = 35)	0.5 ± 1.2 (n = 28)	0.5 ± 1.0 (n = 28)
Identify Force Joint and Direction (6)	No Lab	1.7 ± 1.6 (n = 35)	0.0 ± 2.0 (n = 35)	-1.7 ± 1.7 (n = 35)	0.2 ± 1.6 (n = 29)	1.7 ± 1.5 (n = 25)
	Pasco©	1.8 ± 1.6 (n = 33)	-0.4 ± 2.1 (n = 18)	-2.1 ± 2.0 (n = 16)	0.5 ± 2.0 (n = 14)	-2.0 ± 4.9 (n = 13)
	TrussVR©	1.7 ± 1.8 (n = 50)	0.6 ± 1.9 (n = 34)	-1.0 ± 2.2 (n = 34)	1.0 ± 2.0 (n = 28)	1.2 ± 1.7 (n = 28)
Identify members in C, T, or ZFM (15)	No Lab	0.2 ± 3.7 (n = 35)	0.3 ± 3.7 (n = 35)	0.2 ± 3.3 (n = 35)	-0.2 ± 3.8 (n = 29)	0.7 ± 4.1 (n = 25)
	Pasco©	0.6 ± 3.9 (n = 32)	1.8 ± 3.5 (n = 15)	0.3 ± 3.4 (n = 14)	1.1 ± 3.7 (n = 14)	1.0 ± 3.1 (n = 13)
	TrussVR©	0.4 ± 3.4 (n = 50)	1.0 ± 3.3 (n = 33)	0.7 ± 3.4 (n = 33)	-0.2 ± 3.2 (n = 28)	0.4 ± 3.2 (n = 28)
Total Score (31)	No Lab	4.5 ± 5.3 (n = 35)	2.2 ± 5.2 (n = 35)	-2.3 ± 4.1 (n = 35)	0.9 ± 5.1 (n = 29)	-3.9 ± 5.0 (n = 25)
	Pasco©	4.6 ± 5.0 (n = 32)	3.3 ± 5.4 (n = 15)	-2.0 ± 4.9 (n = 14)	3.2 ± 4.0 (n = 14)	-3.6 ± 5.4 (n = 13)
	TrussVR©	5.1 ± 5.6 (n = 50)	4.3 ± 5.8 (n = 33)	-0.9 ± 4.6 (n = 33)	3.4 ± 6.0 (n = 28)	-2.4 ± 4.0 (n = 28)

* $p < 0.05$, shading indicates the best scores for each question set at each time point

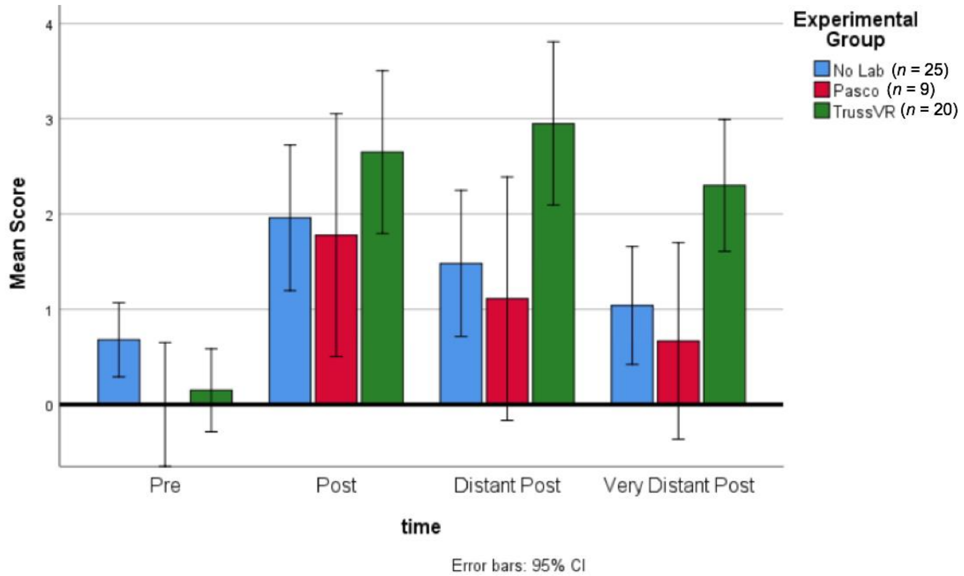

Figure 3: Identify the names of trusses based on provided images. Note that the figure only includes scores for students that responded to all four surveys.

For Distant Post Gain, once again the Kruskal-Wallis non-parametric test indicated that the three groups differed on being able to identify truss names based on seeing an image of the shape, χ^2 (2, N = 90) = 8.71, p = .01. Post hoc Mann-Whitney tests compared the three pairs of groups, using a Bonferroni corrected p value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the TrussVR© group (32.29, n = 35) was significantly higher than that of participants in the No Lab group (20.50, n = 20), z = -2.67, p = .008, r = -.36 i.e. a medium effect size. There was no statistically significant difference between the TrussVR© and Pasco© groups or the No Lab and Pasco© groups.

Finally, for Very Distant Post Gain, the Kruskal-Wallis non-parametric test indicated that the three groups differed on being able to identify truss names based on seeing an image of the shape, χ^2 (2, N = 71) = 15.009, p = .001. Post hoc Mann-Whitney tests compared the three pairs of groups, using a Bonferroni corrected p value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the TrussVR© group (36.43, n = 28) was significantly higher than that of participants in the No Lab group (21.83, n = 29), z = -3.45, p = .001, r = -.46 i.e. a medium-to-large effect size. Also, the mean rank for participants in the TrussVR© group (36.43, n = 28) was significantly higher than that of participants in the Pasco© group (13.89, n = 14), z = -2.94, p = .003, r = -.45 i.e. a medium-to-large effect size. There was no statistically significant difference between the No Lab and Pasco© groups.

3.2.2 Is This a Truss?

The scores on the questions which asked participants to identify whether a given image of a structure was a truss or not were generally similar for all three groups across the time points (as shown in Figure 4), but one result did stand out. For Very Distant Post Gain, the Kruskal-Wallis nonparametric test indicated that the three groups differed on being able to determine if a structure is a truss or not based on seeing an image of the structure, χ^2 (2, N = 71) = 6.961, p =

.03. Post hoc Mann-Whitney tests compared the three pairs of groups, using a Bonferroni corrected p value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the Pasco© group (28.14, $n = 14$) was significantly higher than that of participants in the TrussVR© group (18.18, $n = 28$), $z = -2.56$, $p = .01$, $r = -.40$, a medium effect size. There was not a statistically significant difference between the No Lab and Pasco© groups or the No Lab and TrussVR© groups.

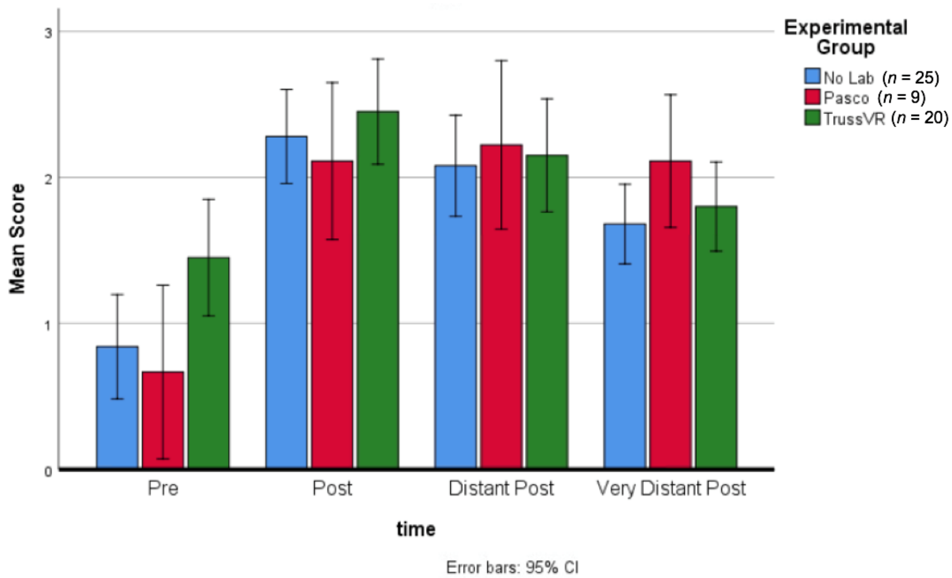

Figure 4: Given an image of a structure, decide if it is a truss or not. Note that the figure only includes scores for students that responded to all four surveys.

3.2.3 Identifying Internal and External Forces

The skill-testing questions that asked participants to identify the location and direction of an externally applied force and those that asked participants to identify the internal forces given a certain loading situation, were similar for the three groups across all time points (as shown in Table 1). The TrussVR© group scored best on the majority of the external force questions and the Pasco© group scored best on the majority of the internal force questions. None of the differences were statistically significant.

3.2.4 Name as Many Trusses as You Can

In each of the four surveys, study participants were asked to name as many types of trusses as they could. For example, "Howe", "Pratt", "simple", "space", and "roof", would all be acceptable answers. Figure 5 and Table 2 show the results. All groups exhibited higher Post scores compared to their Pre scores, but in the No Lab case, the Distant Post scores went down. In the Very Distant Post, the Pasco© group decreased, but the No Lab group showed a slight increase. The TrussVR© group, on the other hand, actually showed a large increase from Pre to Post, a moderate increase from Post to Distant Post, and a stabilization on the Very Distant Post.

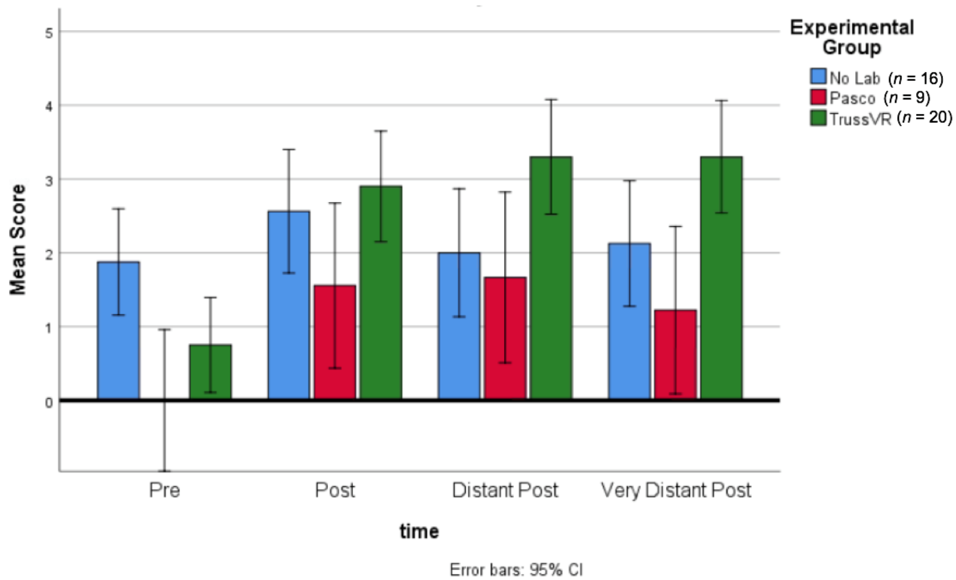

Figure 5: Name as many trusses as you can. Note that the figure only includes scores for students that responded to all four surveys.

Table 2: Gain and retention scores for name as many trusses as you can

Group	**Post Gain** *M* ± *SD* (*n*)	**Distant Post Gain** *M* ± *SD* (*n*)	**2-month Retention** *M* ± *SD* (*n*)	**Very Distant Post Gain** *M* ± *SD* (*n*)	**9-month Retention** *M* ± *SD* (*n*)
No Lab	1.8 ± 2.7 (n = 35)	0.6 ± 2.3 (n = 35)	-1.2 ± 2.9 (n = 35)	0.3 ± 1.9 (n = 16)	-.4 ± 2.0 (n = 16)
Pasco©	1.6 ± 2.1 (n = 16)	1.3 ± 1.7 (n = 16)	-0.3 ± 1.4 (n = 16)	1.2 ± 1.3 (n = 9)	-.3 ± 1.7 (n = 9)
TrussVR©	1.9 ± 2.0 (n = 35)	*2.2 ± 1.9 (n = 35)	*0.3 ± 1.6 (n = 35)	*2.6 ± 1.9 (n = 20)	0.4 ± 1.4 (n = 20)

* $p < 0.05$, shading indicates the best scores for each question set at each time point

There was a statistically significant difference between groups for the Distant Post Gain as determined by the Kruskal-Wallis non-parametric test, χ^2 (2, *N* = 86) = 6.52, *p* = .04. Post hoc Mann-Whitney tests compared the three pairs of groups using a Bonferroni corrected *p* value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the TrussVR© group (41.39, *n* = 35) was significantly higher than that of participants in the No Lab group (29.61, *n* = 35), *z* = -2.45, *p* = .01, *r* = -.29 i.e. a medium effect size. There was no statistically significant difference between the No Lab and Pasco© groups or the Pasco© and TrussVR© groups.

There was also a statistically significant difference between groups for the Very Distant Post Gain using the Kruskal-Wallis non-parametric test, χ^2 (2, *N* = 45) = 12.54, *p* = .002. Post hoc Mann-Whitney tests compared the three pairs of groups using a Bonferroni corrected *p* value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the TrussVR© group (23.53, *n* = 20) was significantly higher than that of participants in the No Lab group (12.22, n = 16), z = -3.24, p = .001, r = -.54 i.e. a large effect size. There was no statistically significant difference between the No Lab and Pasco© groups or the Pasco© and TrussVR© groups.

In addition, Two-Month Retention was also significantly different as determined the Kruskal-Wallis non-parametric test, χ^2 (2, N = 86) = 7.80, p = .02. Post hoc Mann-Whitney tests compared the three pairs of groups using a Bonferroni corrected p value of 0.017 (0.05 ÷ 3) to indicate statistical significance. The mean rank for participants in the TrussVR© group (41.90, n = 35) was significantly higher than that of participants in the No Lab group (29.10, n = 35), z = -2.66, p = .008, r = -.32 i.e. a medium effect size. There was no statistically significant difference between the No Lab and Pasco© groups or the Pasco© and TrussVR© groups.

3.3 Impressions of the Lab Experience

A set of lab experience questions were asked of the Pasco© and TrussVR© groups on the Post and Distant Post Surveys. As shown in Table 3, more questions were asked on the Post Survey than on the Distant Post Survey. Each of these questions were asked using a 5-point Likert Scale (1 = strongly disagree, 5 = strongly agree). A summary of the questions and mean responses are shown in Table 3. Since at least one of the sets of ratings was non-normal in every case, based on skewness values and Shapiro-Wilk's Test for Normality, Mann-Whitney U tests were performed to find differences between the lab groups. As can be seen in Table 3, in almost every case the participants in the TrussVR© group rated their experience higher than the Pasco© group.

3.3.1 Recommend Use in Teaching Trusses

On the Post Survey, based on a Mann-Whitney U test, the 50 participants in the TrussVR© group had higher mean ranks (48.30) than the 30 participants in the Pasco© group (27.50) on the statement "I would recommend the use of this lab in teaching trusses", U = 360, p < .001, r = -.48, which was a statistically significant difference. Two months later, on the Distant Post Survey, based on a Mann-Whitney U test, the 33 participants in the TrussVR© group had higher mean ranks (28.18) than the 15 participants in the Pasco© group (16.40) on the same statement, U = 126, p = .002, r = -.44, which was a statistically significant difference. In both of these cases, these are medium-to-large effect sizes.

3.3.2 Helped in Learning Trusses

On the Post Survey, based on a Mann-Whitney U test, the 50 participants in the TrussVR© group had higher mean ranks (47.29) than the 30 participants in the Pasco© group (29.18) on the statement "This lab helped my learning of trusses", U = 410.5, p < .001, r = -.40, which was a statistically significant difference and is a medium-to-large effect size. Once again, two months later on the Distant Post Survey, based on a Mann-Whitney U test, the 33 participants in the TrussVR© group had higher mean ranks (27.41) than the 15 participants in the Pasco© group (18.10) on the same statement, U = 151, p = .03, r = -.32, which was a statistically significant difference and a medium effect size.

Table 3: Impressions of the lab experience

Question	Group	Post Survey $M \pm SD$ (*n*)	Distant Post Survey $M \pm SD$ (*n*)
I would recommend the use of this lab in teaching trusses	TrussVR©	*4.70 ± .59 (*n* = 50)	*4.64 ± .60 (*n* = 33)
	Pasco©	3.64 ± 0.93 (*n* = 30)	3.93 ± .92 (*n* = 15)
This lab helped my learning of trusses	TrussVR©	*4.15 ± .71 (*n* = 50)	*4.15 ± .87 (*n* = 33)
	Pasco©	3.14 ± 1.10 (*n* = 30)	3.57 ± .85 (*n* = 15)
I enjoyed using TrussVR© / Pasco©	TrussVR©	*4.82 ± .39 (*n* = 50)	*4.73 ± .45 (*n* = 33)
	Pasco©	3.64 ± .75 (*n* = 30)	3.86 ± .77 (*n* = 15)
This lab has enhanced my interest in trusses/ civil engineering	TrussVR©	*4.00 ± .94 (*n* = 50)	*4.09 ± .84 (*n* = 33)
	Pasco©	3.29 ± .99 (*n* = 30)	3.21 ± .98 (*n* = 15)
The instructions in the lab were clear	TrussVR©	4.26 ± .71 (*n* = 53)	
	Pasco©	4.19 ± .48 (*n* = 31)	
It was easy to use the tools	TrussVR©	4.36 ± .74 (*n* = 53)	
	Pasco©	4.35 ± .49 (*n* = 31)	
It was easy to see the forces acting in the truss	TrussVR©	4.11 ± .89 (*n* = 53)	
	Pasco©	3.74 ± .93 (*n* = 31)	
It clearly showed what members were in tension, compression, or were zero force members	TrussVR©	*4.47 ± .61 (*n* = 53)	
	Pasco©	3.90 ± .87 (*n* = 31)	
It was easy to see the reaction forces at supports of the truss	TrussVR©	*4.08 ± .87 (*n* = 53)	
	Pasco©	3.68 ± .87 (*n* = 31)	
It was easy to build and modify trusses	TrussVR©	3.94 ± .93 (*n* = 53)	
	Pasco©	*4.55 ± .51 (*n* = 31)	
It was easy to test trusses	TrussVR©	4.02 ± .93 (*n* = 53)	
	Pasco©	4.06 ± .85 (*n* = 31)	

* $p < 0.05$, shading indicates the best scores for each question set at each time point

3.3.3 Enjoyed Using

On the Post Survey, based on a Mann-Whitney *U* test, the 50 participants in the TrussVR© group had higher mean ranks (50.52) than the 30 participants in the Pasco© group (23.80) on the statement "I enjoyed using TrussVR©/Pasco©", $U = 249$, $p < .001$, $r = -.64$, which was a statistically significant difference. On the Distant Post Survey, based again on a Mann-Whitney *U* test, the 33 participants in the TrussVR© group had higher mean ranks (29.50) than the 15 participants in the Pasco© group (13.50) on this statement, $U = 82.5$, $p < .001$, $r = -.60$, which was a statistically significant difference. Both of these are large-to-much larger than typical effect sizes.

3.3.4 Enhanced Interest in Trusses and/or Civil Engineering

On the Post Survey, using a Mann-Whitney *U* test, the 50 participants in the TrussVR© group had higher mean ranks (48.28) than the 30 participants in the Pasco© group (27.53) on the statement "This lab has enhanced my interest in trusses/civil engineering", $U = 361$, $p < .001$, $r = -.45$, which was a statistically significant difference. When asked the same statement on the Distant Post Survey, based on a Mann-Whitney *U* test, the 33 participants in the TrussVR© group had higher mean ranks (28.32) than the 15 participants in the Pasco© group (16.10), $U = 121.5$, $p = .003$, $r = -.43$, which was a statistically significant difference. Both of these results are a medium-to-large effect size.

3.3.5 Showed What Members were in Tension, Compression, or were ZFMs

On the Post Survey, based on a Mann-Whitney *U* test, the 53 participants in the TrussVR© group had higher mean ranks (48.25) than the 31 participants in the Pasco© group (32.68) on the statement "It clearly showed what members were in tension, compression, or were zero force members", $U = 517$, $p = .002$, $r = -.21$, which was a statistically significant difference and is a small-to-medium effect size.

3.3.6 Easy to See the Reaction Forces

On the Post Survey, based on a Mann-Whitney *U* test, the 53 participants in the TrussVR© group had higher mean ranks (46.60) than the 31 participants in the Pasco© group (35.48) on the statement "It was easy to see the reaction forces at supports of the truss", $U = 604$, $p = .03$, $r = -.24$, which was a statistically significant difference and a small-to-medium effect size.

3.3.7 Easy to Build and Modify Trusses

On the Post Survey, based on a Mann-Whitney *U* test, the 31 participants in the Pasco© group had higher mean ranks (52.24) than the 53 participants in the TrussVR© group (36.80) on the statement "It was easy to build and modify trusses", $U = 519.5$, $p = .002$, $r = -.33$, which was a statistically significant difference and a medium or typical effect size. Within the lab experience questions, this was the only statement where the Pasco© group rated higher.

3.4 Open-Ended Responses

There were a series of open-ended questions on the Post Survey that asked participants about their experiences and learnings. These questions were posed to participants in all three groups, except for two questions: 1) a question that asked for feedback on "bridge-building aspects" but the No Lab group did not build any bridges, and 2) a question that asked for the "most noteworthy thing that you gained from using TrussVR©/Pasco©?"

3.4.1 What Do You Remember Most About Trusses?

Respondents could list up to 3 ideas/concepts in their answer to this question. Indeed, some respondents provided more than 3 ideas/concepts within their answer(s). The average number of responses per respondent (2.23 ± .03) and the standard deviation in the number of responses per respondent (.83 ± .03) were not significantly different between the three groups. The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 4 summarizes the thematic analyses. Results of interest are highlighted in light gray. The lab experiences elicited more emotionally laden responses, and more responses about experiences with trusses. This was counterbalanced by a reduced focus on truss solving methods in the Pasco© and TrussVR© groups. Truss qualities were the most popular type of response for every group.

Table 4: What do you remember most about trusses?

No Lab (n = 75)	Pasco© (n = 63)	TrussVR© (n = 113)	Response Theme
5.3%	0.0%	2.6%	difficulty
2.7%	1.6%	0.0%	efficiency
0.0%	4.8%	7.0%	expression of emotion/affinity
0.0%	3.2%	8.8%	experiences
45.3%	47.6%	44.7%	qualities of trusses
37.3%	33.3%	28.9%	truss solving methods
4.0%	1.6%	4.4%	types
2.7%	7.9%	3.5%	uses of trusses

3.4.2 Hardest Skill to Learn When Analyzing Trusses

Respondents were expected to give one answer to this question. Some respondents provided more than one idea/concept within their answer. The average number of responses per respondent (1.17 ± .01) and the standard deviation in the number of responses per respondent (.38 ± .01) were not significantly different between the three groups. The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 5 summarizes the thematic analyses. Results of interest are highlighted in light gray. TrussVR© respondents clearly identified the analysis of forces as the hardest skill, while the Pasco© group found truss solving methodology hardest.

Table 5: Hardest skill to learn when analyzing trusses?

No Lab (n = 40)	Pasco© (n = 35)	TrussVR© (n = 57)	Response Theme
35.0%	25.7%	29.8%	ZFMs
22.5%	17.1%	49.1%	analyzing/determining forces
25.0%	37.1%	10.5%	truss solving methodology
2.5%	8.6%	0.0%	calculations
2.5%	2.9%	3.5%	visualization
12.5%	8.6%	7.0%	other

3.4.3 Most Surprising Thing You Learned About Trusses

Respondents were expected to give one answer to this question. Some respondents provided more than one idea/concept within their answer. The average number of responses per respondent (1.09 ± .06) and the standard deviation in the number of responses per respondent (.31 ± .13) were not significantly different between the three groups. The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 6 summarizes the thematic analyses. Results of interest are highlighted in light gray. No Lab respondents focused on solving trusses and features of trusses, while the Pasco© group keyed in on the value of trusses as well as features. The TrussVR© respondents skewed on the difficulty of solving trusses versus the other groups, and noted the variety of types of trusses more than the other groups.

Table 6: What was the most surprising thing you learned about trusses?

No Lab (n = 39)	Pasco (n = 30)	TrussVR (n = 53)	Response Theme
30.8%	13.3%	13.2%	solving/understanding trusses
15.4%	6.7%	13.2%	what types/methods work
25.6%	30.0%	24.5%	features of trusses
0.0%	16.7%	9.4%	the value of trusses
5.1%	0.0%	13.2%	difficult challenge
12.8%	13.3%	3.8%	easy challenge
5.1%	10.0%	15.1%	types
5.1%	6.7%	5.7%	nothing

3.4.4 What Do You Enjoy Most About Trusses?

Respondents were again expected to provide one response to this question, but some participants mentioned more than one theme in their response. The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 7 summarizes most of the thematic analyses. Results of interest are highlighted in light gray. The most popular theme for the No Lab and Pasco© groups was the solving process. On the other hand, the TrussVR© group's most popular theme was focused on how trusses are practical and/or relevant to the real-world. Other common themes were that the No Lab group enjoyed the puzzle-like features of trusses and the

Pasco© group enjoyed that trusses were easy. The Pasco© group was the only group to describe the structure of trusses.

Table 7: What do you enjoy most about trusses?

No Lab (n = 39)	Pasco© (n = 35)	TrussVR© (n = 51)	Response Theme
18.4%	5.7%	27.5%	Practical/real-world
18.4%	22.9%	17.6%	Easy
0.0%	5.7%	15.7%	Lab
21.1%	5.7%	9.8%	Puzzle
2.6%	5.7%	9.8%	How Trusses work
23.7%	28.6%	9.8%	Solving process
0.0%	11.4%	0.0%	Structure

3.4.5 What Do You Enjoy Least About Trusses?

Respondents were again expected to provide one response to this question, but some participants mentioned more than one theme in their response. The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 8 summarizes most of the thematic analyses. Results of interest are highlighted in light gray. There was not one common theme that stood out for the TrussVR© group; rather, there were three themes that each accounted for 12.8% of the total responses for the group: time-consuming, the analysis process, and internal forces. Time-consuming and the analysis process were also the most common themes for the Pasco© group. The No Lab group also did not enjoy the analysis process and specifically mentioned zero force members in many of their responses to this question. The TrussVR© group was the only group to mention terminology.

Table 8: What do you enjoy least about trusses?

No Lab (n = 34)	Pasco© (n = 26)	TrussVR© (n = 47)	Response Theme
8.8%	19.2%	12.8%	Time consuming
23.5%	19.2%	12.8%	Analysis Process
5.9%	0.0%	12.8%	Internal Forces
8.8%	15.4%	10.6%	Difficulty
14.7%	11.5%	10.6%	Tedious
17.6%	11.5%	8.5%	Zero Force Members
2.9%	0.0%	6.4%	Specific Types of Problems
0.0%	0.0%	6.4%	Terminology

3.4.6 What was the Most Noteworthy Gain from Using TrussVR©/Pasco©?

Respondents were expected to give one answer to this question. Some respondents provided more than one idea/concept within their answer. The average number of responses per respondent (1.24 ± .04) and the standard deviation in the number of responses per respondent (.48 ± .04) were not significantly different between the two groups (Pasco© and TrussVR©). The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 9 summarizes the thematic analyses. Results of interest are highlighted in light gray. In general, it appears that TrussVR© users felt that they better understood trusses perhaps based on the iterative nature of experimentation and/or improved visualization capabilities. These students clearly felt the potential of VR for learning trusses. On the other hand, Pasco© users took away one major lesson that TrussVR© did not provide: real life is different than theory (as a result of the challenges associated with getting useful measurements from load cells and "non-ideal" truss models).

Table 9: What was the most noteworthy thing gained from using TrussVR©/Pasco©?

Pasco© (n = 37)	**TrussVR© (n = 60)**	**Response Theme**
18.9%	26.7%	better understanding of trusses
10.8%	13.3%	improved visualization of truss behaviours
0.0%	18.3%	interactivity/feedback
32.4%	0.0%	realizing that real life is different
2.7%	5.0%	application insights
5.4%	3.3%	insights on personal shortcomings
5.4%	23.3%	seeing/experiencing lab's potential
10.8%	5.0%	seeing real world relevance
13.5%	5.0%	no gains noted/complaints

3.4.7 Feedback on Bridge-Building

Respondents were asked to "Provide any feedback you wish regarding the bridge building aspects of TrussVR©/Pasco©, including how fun it was, how much you learned, and how much it stimulated your interest in trusses?". Students could give any number of responses. Indeed, some respondents provided several ideas/concepts within their answer, addressing different aspects of the posed question. The average number of responses per respondent (2.6 for Pasco©, 3.1 for TrussVR©) and the standard deviation in the number of responses per respondent (1.42 ± .01) were not significantly different between the Pasco© and TrussVR© groups. The *n* values indicate the total number of ideas/concepts from respondents within that group. Table 10 summarizes the thematic analyses. Results of interest are highlighted in light gray. The Pasco© responses were marked by tangibility and practicality, as well as a neutral or negative bias, although nearly a third of responses were focused on fun/joy/excitement. The TrussVR© responses were similarly focused on fun/joy/excitement, with a slightly more frequent reference to enhanced learning. TrussVR© respondents had fewer negative comments except for time pressure issues, which were a major factor in their feedback. They were also more likely to mention visualization enhancement.

Table 10: Feedback regarding bridge-building using TrussVR©/Pasco©

Pasco© (n=77)	TrussVR© (n=165)	Response Theme
31.2%	35.4%	fun/excitement/joy/surprise
9.1%	14.6%	enhanced teaching, learning and understanding
13.0%	3.0%	tangibility/practicality/hands-on nature
3.9%	2.4%	stimulated/increased interest
1.3%	2.4%	responsive
7.8%	4.9%	interesting/engaging/inspiring/intriguing
11.7%	4.9%	negative comments
1.3%	14.6%	time pressure issues
0.0%	5.5%	s/w development suggestions
11.7%	1.8%	nothing special
1.3%	5.5%	visualization was enhanced
7.8%	4.9%	difficulty/challenge

3.5 Can You Compare Two Trusses?

One final open-ended question was posed to all three study groups in the Post survey. They were asked whether they were able to compare two types of trusses. If they answered "yes", they were asked to do so. Table 11 summarizes the thematic analyses. If respondents provided a comparison, it was subjectively evaluated as being either a great comparison, a good one, an OK one, or a poor one (on the basis of correctness and insight). The topical basis for comparison is given in the middle portion of Table 11. Note that one comparison could have addressed more than one topical basis. At the bottom of Table 11, descriptive statistics regarding the data set for each group are provided. Clearly, there was an increased inclination to say that one could compare two types of trusses from No Lab to TrussVR©. As well, there appeared to be an increasing ability from No Lab to TrussVR©, although the scarcity of responses from the No Lab and Pasco© groups preclude a definitive answer on that issue.

3.6 Health Effects

No TrussVR© users experienced any significant negative health effects from using the VR sets. A couple of students started to feel dizzy and sat down when this occurred, but they did not have any lasting symptoms. Our lab supervisors were trained to anticipate and assist in such circumstances.

4.0 Discussion

It is easy to find references in the literature that say that learning with VR is more effective than conventional means e.g. [1], [6], [13], [14], and [15]. Finding convincing cases that demonstrate this contention is more difficult e.g. [17]. Is it the VR itself that creates learning effects, or what one is doing in VR? This is the difficult question. As such, extracting lessons from the present study is challenging because, as with many other VR studies, the medium confounds the message. It may well be that the novelty of VR creates some of its apparent benefits [13].

However, it is clear that there are benefits and some of these benefits are not surprising, if one takes a constructivist approach to learning. TrussVR© shifts the intellectual work of the student from the underlying math to the higher level understanding of dynamic structural behaviour. This was reflected in some of the patterns of comments which differed from the No Lab and Pasco© groups. TrussVR© comments more often looked beyond superficial features and rote-learning from class. There was more willingness and ability to compare truss types. TrussVR© students self-reported more learning, understanding, and visualization abilities. Not surprisingly, TrussVR© participants did better at identifying and naming different types of trusses. After all, they had more training time on that specific task. Ultimately, TrussVR© made an impression and the memorable experience seems to have resulted in some long-lasting benefits as well as enhanced interest levels. Students who used TrussVR© were generally very keen to see its use broadened in the teaching of trusses.

Table 11: Comparing trusses: self-assessed ability and responses

	No Lab		**Pasco©**		**TrussVR©**	
# of Respondents	35		35		53	
# of "No" Reponses	34		32		40	
# of "Yes" Responses	1		3		13	
% of "Yes" Responses	2.9%		8.6%		24.5%	
Quality of Comparison	Great	1	Great	1	Great	6
	Good	0	Good	1	Good	3
	OK	0	OK	0	OK	3
	Poor	0	Poor	1	Poor	1
Topic of Comparison	shape	1	shape	1	shape	10
	function	1	function	1	function	5
	misunderstanding	0	misunderstanding	1	misunderstanding	0
	practical use	0	practical use	0	practical use	3
	can't recall names	0	can't recall names	0	can't recall names	4

In general, VR's educational advantages are purported to primarily depend upon active learning, higher levels of engagement, and/or embodied cognition. Active learning can be defined as any activity which meaningfully engages the learner in thinking about what they are doing e.g. [16]. This commonly happens in VR. Prince [16] makes the point that active learning is more than just learning activities though. These activities need to be tied to important learning outcomes and they need to encourage deep engagement. Findings in the field of cognitive psychology suggest that additional information processing after initial learning is key for long-term retention [18]. In embodied cognition, the theory is that we make sense of the world through our body's interaction with it, which VR readily provides e.g. [12].

Then there is the question of what activity one should be doing in VR. Testing and gaming are both known to enhance learning, regardless of whether VR is used to create these opportunities e.g. [10], [11], and [13]. TrussVR© is aligned well with the philosophy of cognitive constructionists like Jean Piaget, wherein knowledge is actively constructed [4]. Similarly, the TrussVR© software embraces the principles of constructivist learning environments that feature authentic tasks and reflective practice as per [8] and [11]. Additionally, problem-based learning, or PBL, is clearly a feature of the bridge building module in TrussVR© and PBL been shown to have similar benefits e.g. [16]. Having learners generate explanations for phenomena can be used to produce better long-term retention and application of knowledge [10]. This happens within most of the modules in TrussVR©.

For all the focus on VR, much was learned from the Pasco© responses as well. These students were better able to identify what was a truss as compared to the other groups. Perhaps one of the greatest lessons from the current study was that the Pasco© users acquired a benefit unique to a "real" lab; an appreciation for the differences between the ideal textbook answers and real-life. While this could conceivably be achieved in VR, it may well be best achieved in physical reality.

In their systematic review of the VR education literature, Radianti et al [17] noted that there is not a lot of depth to the literature yet. While this study has some clear weaknesses, it does feature many of the attributes that Radianti's team recommends in a good study. TrussVR© was designed with constructivism in mind. We have noted how the TrussVR© lab was incorporated into the existing curriculum. We employed qualitative and quantitative measures that looked at knowledge gains and retention (learning outcomes) in the short and long term. We have provided some bases for comparison in the nature of the answers given for the types of open-ended questions that we posed. Indeed, very few studies of a similar nature have been previously reported e.g. [5] and [6]. In some respects, the most similar has been that of Madden et al [12] that involved three conditions (normal lab, desktop app, and VR app) with a pre/post assessment model employing isomorphic sets of questions. However, their study was focused on Astronomy.

If we were to repeat or expand the current study, it would be interesting to give TrussVR© users more time with TrussVR©, as that was the single biggest concern and criticism from the TrussVR© group. Indeed, it should not be forgotten that all of the noted benefits described in this paper arose from 6 minutes of studying different truss types, 6 minutes of doing some analytical problems, and 7 minutes of bridge building. If nothing else, this showed the learning efficiency of VR!

It also speaks to important future work. There are three main avenues of fruitful future work on TrussVR©. The first, reflecting the point made in the previous paragraph, is time on task. If the value of VR is strictly its novelty, then time on task should not have much of a learning effect. We should have students spend 30, 60 and 90 minutes on TrussVR© to see if and how learning scales with training time.

The next issue is learning task. Our three modules in TrussVR© were three very different tasks. The first was fairly open-ended and not particularly goal-driven. Just observe and play for a period of time. The second tasks were problem-solving exercises. These were "tests" and performance was scored/evaluated. The third was a goal-directed game/contest. The second and third tasks are known to be effective in stimulating learning, although it is unclear which is superior. We should have users of TrussVR© focus on the different activities within TrussVR© to see which activities lend themselves to specific learning outcomes.

Finally, there is the question of medium (of experience). In the current study, we compared the "no experience" (No Lab) with the "conventional experience" (Pasco©) with the "VR experience" (TrussVR©). These are three very different types of experiences, although they each have their relevance in a comparison study. Other experiences which need to be evaluated include a) a 2D laptop version of TrussVR©, b) a lower-resolution VR version of TrussVR© on a more cost-effective platform (e.g. on an Oculus Quest© platform), and c) a "paper" version of TrussVR© i.e. have learners do all of the module activities on paper in a similar time frame (look at different truss types, solve problems, design a good bridge). These variations would truly allow us to tease out the relative value (if any) of using immersive VR.

It should be noted that software tools with similar learning goals to TrussVR do exist. Hubing et al [7] developed a computer-based teaching tool for truss analyses. However, their software did not facilitate the creation and testing of truss structures. Karweit [9] created "Bridge Designer" which does much of what TrussVR© does, only in 2D. This software was suspected to be very effective in enhancing the performance of trusses in a competitive bridge design contest and may well be useful in the role of "2D laptop version of TrussVR".

5.0 Acknowledgements

The authors would like to thank Ron and Jane Graham for their moral and financial support of this project, without which it would not have occurred. Likewise, we want to thank and recognize Tod Baudais and Bruce Cory of Sprockety Ventures Inc, and Ian Stavness of University of Saskatchewan Computer Science, for programming TrussVR©. In developing TrussVR©, Bruce Sparling played an important role as a structural engineering consultant. Richard Heese was very helpful as tech support during the study, and in carrying out this study, Jim Bugg was vitally helpful as the Coordinator of GE 124 Mechanics I (Statics). We would also like to thank Stan Yu for his suggestions on how to complete the statistical analysis, and Yasi Delkash and Amir Ravanbod for their work as lab supervisors during the VR sessions.

References

[1] Abdul-Hadi G. Abulrub, Alex N. Attridge, and Mark A. Williams, "Virtual Reality in Engineering Education", *Proceedings of the 2011 IEEE Global Engineering Education Conference*, Amman, Jordan, pp. 751-757, April 2011.

[2] Ryan Banow and Sean Maw, "First Results from a Study on the Efficacy of Using Virtual Reality (VR) to Teach Truss Mechanics", *Canadian Engineering Education Association Annual Conference*, Ottawa, June 2019

[3] J. Cohen (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic.

[4] Grant Cooper, and Li Ping Thong, *Implementing Virtual Reality in the Classroom: Envisaging Possibilities in STEM Education*. In: STEM Education; An Emerging Field of Inquiry, Tasos Barkatsas, Nicky Carr and Grant Cooper (eds.), Brill Sense, Boston, pp. 61-73, 2019.

[5] Chelsea Ekstrand, Ali Jamal, Ron Nguyen, Annalise Kudryk, Jennifer Mann, and Ivar Mendez, "Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: a randomized controlled study", *The Canadian Medical Association Journal (OPEN)*, Vol. 6, No. 1, pp. E103-E109, 2018.

[6] Laura Freina, and Michela Ott, "A Literature Review on Immersive Virtual Reality in Education: State of the Art and Perspectives", *Proceedings of eLearning and Software for Education (eLSE)*, Bucharest, April 23-24, 2015.

[7] Nancy Hubing, David B. Oglesby, Timothy A. Philpot, Vikas Yellamraju, Richard H. Hall, and Ralph E. Flori, "Interactive Learning Tools: Animating Statics", *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition*, Session 2368, 2003.

[8] David H. Jonassen, and Lucia Rohrer-Murphy, "Activity theory as a framework for designing constructivist learning environments", *Educational Technology, Research and Development*, Vol. 47, No. 1, pp. 61-79, March 1999

[9] Michael Karweit, "Enhanced learning through a 'virtual laboratory'", *Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition*, Session 1520, 2002.

[10] Douglas P. Larsen, Andrew C. Butler, and Henry I. Roediger III, "Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial", *Medical Education*, Vol. 43, pp. 1174-1181, 2009.

[11] Douglas P. Larsen, Andrew C. Butler, and Henry I. Roediger III, "Comparative effects of test-enhanced learning and self-explanation on long-term retention", *Medical Education*, Vol. 47, pp. 674-682, 2013.

[12] J.H. Madden, A.S. Won, J.P. Schuldt, B. Kim, S. Pandita, Y. Sun, T.J. Stone, and N.G. Holmes, "Virtual Reality as a Teaching Tool for Moon Phases and Beyond", 2018 *Physics Education Research Conference*, 4 pages, 2018.

[13] Zahira Merchant, Ernest T. Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: a meta-analysis", *Computers & Education*, Vol. 70, pp. 29-40, 2014.

[14] John I. Messner, Sai C.M. Yerrapathruni, Anthony J. Baratta, and Vaughn E. Whisker, "Using Virtual Reality to Improve Construction Engineering Education", *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition*, Session 1121, 2003.

[15] Veronica S. Pantelidis, "Virtual Reality and Engineering Education", *Computer Applications in Engineering Education*, Vol. 5, No. 1, pp. 3-12, 1997.

[16] Michael Prince, "Does Active Learning Work? A Review of the Research", *Journal of Engineering Education*, pp. 223-231, July 2004.

[17] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda", *Computers & Education*, Vol. 147, April 2020

[18] Henry L. Roediger III, and Jeffrey D. Karpicke, "The Power of Testing Memory: Basic Research and Implications for Educational Practice", *Perspectives on Psychological Science,* Vol. 1, No. 3, pp. 181-210, 2006.