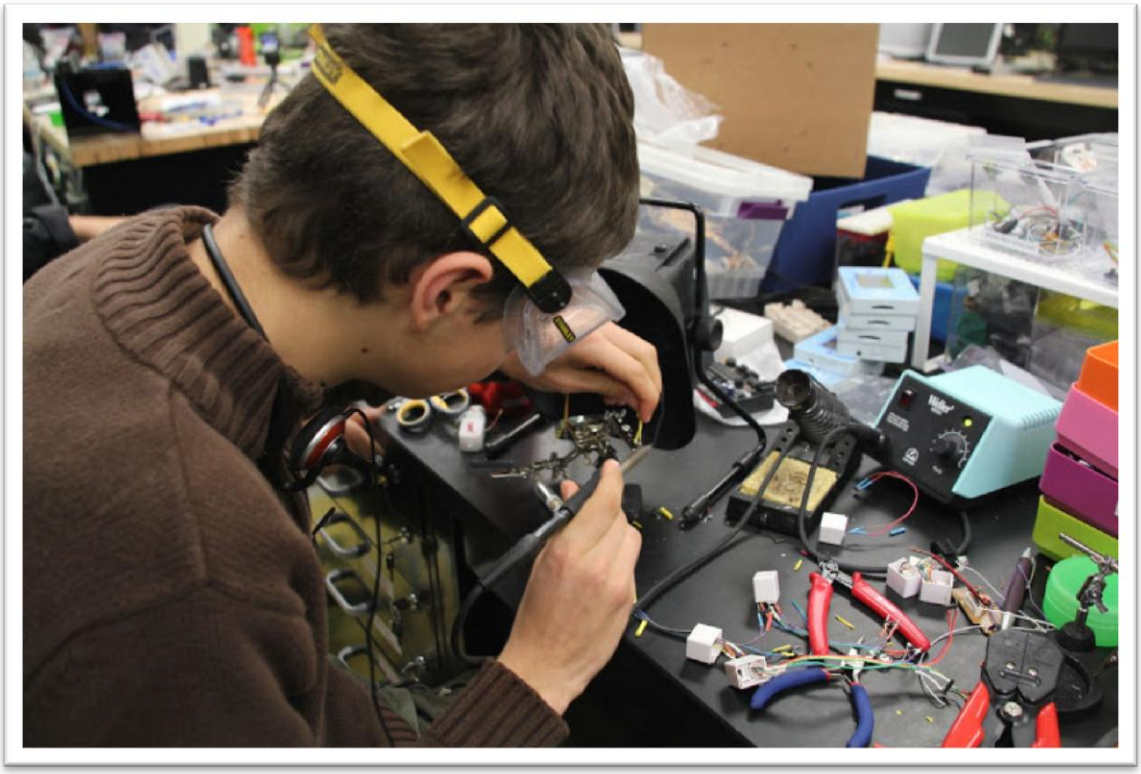
Amir Lopatin Final Report

Understanding Learning Processes in Student-Designed Project-Based Learning

Marcelo Worsley

Contents



Executive Summary

This report describes work completed through the generous funding of the Amir Lopatin Fellowship. Throughout this report I will identify a series of project-based learning programs that I undertook in order to study how student learning in constructionist learning environments. The analysis includes examining both semi-structured clinical interview data, as well as data that was captured throughout the students" learning experiences.

Methodologically, the work combines techniques from the learning sciences with techniques from computer science to explore ways to study learning processes in a more automated, scalable and systematic fashion. This combined methodology is used to study two main research questions:

1. Do high school students demonstrate positive changes in identity, depth of knowledge and precision through involvement in a 2 to 3-week long student centered, project based learning experience?

2. Does using a state-based analysis of student actions and gestures, while they are participating in a building activity, produce a meaningful representation of student learning that can be used to better understand that student's STEM intuitions and how their intuitions change over time?

Because of the difficulties in collecting and automatically analyzing the persistently captured data, the more detailed portion of this report will focus on data from the semi-structured clinical interviews. For the persistently data captured, I will focus on more high level observations, but will look to go into more depth in analyzing the richness of this data in future months. Portions of this work also appear in the proceedings for the 2013 Learning Analytics and Knowledge Conference.

Abstract

Project-based learning has found its way into a range of formal and informal learning environments. However, systematically assessing these environments remains a significant challenge. Traditional assessments, which focus on learning outcomes, seem incongruent with the process-oriented goals of project-based learning. In order to develop appropriate scalable measures for assessing project-based learning I will study 8 high school students as they engage in a 2 to 3-week engineering design workshop. During the workshop I will collect extensive multi-modal data which will help construct complete pictures of their learning gains, while also facilitating the use of automated analyses using data mining techniques.

Introduction

Since the work of Dewey (1897, 1913), Vygotsky (1978) and Papert (1980) there has been an increased integration of student-designed project-based learning (SPBL) – constructionism, inquiry based learning and constructivism, for example - in classrooms and in informal contexts. However, systematically assessing learning in these environments remains a significant challenge and an area of national importance (NETP 2010). Traditional forms of assessment, which are easy to distribute and analyze, may necessarily be incongruent with the objectives of SPBL because most assessments tend to focus on outcomes, where SPBL is largely focused on the process. Furthermore, the forms of assessment that may be well suited to the spirit of project based learning environments: ethnographies, micro-genetic analysis and portfolio based assessments; are unable to scale to the level that is needed for educators to consistently and reliably turn to SPBL on a regular basis. This creates a conflict that can often times be difficult to resolve, and leave individuals dissatisfied with the lack of measurable changes in learning outcomes in project based learning studies.

In order to develop a more appropriate measure and means for assessing SPBL I want to systematically construct a complete picture of how learning takes place in SPBL environments. To this end, the overarching research question that I have is, 'How do students learn in SPBL environments?' with sub-questions: 'Can the learning processes in SPBL environments be characterized using a system of continuous, multi-modal monitoring based on a composition of artificial intelligence technologies?'; 'What are the long term impacts of SPBL activities on students learning trajectories?'; and 'What factors influence effectively learning in SPBL environments?' Finally, in designing a method for describing learning in SPBL implementations, I am going to try to keep in mind the following design principles for my assessments: 1) basing assessment on the learning process and not simply

learning outcomes; 2) enabling scalable assessments; 3) allowing students to use multiple forms of expression; and 4) reducing non-integrated forms of formal assessment (i.e. traditional tests and exams).

Realizing that the above questions and principles are still relatively broad, I have identified two specific primary research questions that I would like to answer in the current instantiation of this project:

1. Do high school students demonstrate positive changes in identity, depth of knowledge and precision through involvement in a 2 to 3-week long SPBL learning experience?

2. Does using a state-based analysis of student actions and gestures, while they are participating in a building activity, produce a meaningful representation of student learning that can be used to better understand that student's STEM intuitions and how their intuitions change over time?

Previous Literature

This work builds on a wealth of literature from the learning sciences, cognitive science and learning analytics. More specifically, previous works describe the importance of student-designed project based learning experiences for promoting learning of STEM (Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman 2004; Roth, 1996; 1997; 1998 in Kelly and Copabianco, 2012); techniques for observing behavioral and developmental changes among students through rich ethnographic studies (Shulman 2006, Abrahamson 2009, Barron 2006; Barron 2004, diSessa 2002); and using automated techniques for identifying salient markers of learning in a range of modalities (Litman et al 2009, Forbes-Riley et al 2009, Forbes-Riley and Litman 2010). Furthermore, it builds on the work that I have done during the past two and a half years which has found that there are meaningful cues in student speech (Worsley & Blikstein 2011a, Worsley & Blikstein 2011b, Worsley 2011b), gaze (Worsley & Blikstein, 2012b), programming state (Piech et al 2012); epistemological beliefs and identity (Worsley 2012b, unpublished); and a combination of modalities (Worsley 2011a, Worsley & Blikstein 2011), for example. It also builds on a number published and of unpublished research tools that enable: object tracking (Worsley 2012a); user localization (Worsley, & Huang 2012); and multi-modal data capture of collaborative work (Worsley & Blikstein 2011, Worsley and Blikstein, 2012a). To this end, my previous research has looked at a variety of modalities in isolation, but what I am proposing for this study is a complete and all-inclusive study that will combine these different techniques and look at the

relationships that exist among the different modalities. Additionally, it will use a number of novel technologies that we've developed in order to collect data that will, hopefully, further substantiate the findings reached from my previous papers.

Student Participation

In order to realize the above considerations and answer my research questions, I conducted a study with 30 high school students. These students participated in an engineering design program that meets for 4 hours per day and that challenge them to explore: bifocal modeling, computational modeling, digital fabrication, robotics and introductory electronics, computer programming, wood working, polymer casting and more. During the first week, students will spend one hour per day learning about different tools and techniques that can be used for doing digital fabrication. During the second and third weeks, students will participate in both individual and group work. Week 2 will consist of a team competition where students must create a device to complete a certain set of tasks. In the third week, students will be challenged to complete an individual project focused around an invention of their choosing, that will likely solve a problem that is meaningful to them, and may also be geared towards creating an artifact for a friend or family member.

Data Collection and Analysis

Pre-Post Data

Interviews - Using a semi-structured interview protocol, I interviewed students for approximately 30 minutes before and after the summer program. During these interviews, students were asked to describe their recent experiences designing and building technology and their view of how the workshop transpired. Additionally students will be asked to do participate in a think-aloud as they complete design-based challenges. These design-based challenges will include creating devices that will typically combine different sensors and that can be completed using a multitude of techniques. These protocols will be based on previous work that we've done and reported on. Additionally, these protocols will have students make graphic designs of their ideas and will also have them build some physical objects.

Persistent Data Capture

In addition to pre-post data, I will also be collecting a wealth of data on a more continuous basis. This data will contain student behaviors, actions, speech and sentiment. These data sources will

leverage technological tools that I have developed and intentionally designed to be used for understanding student learning in this unstructured, open-ended environment.

Student Wifi-based Localization – Using a custom Android application, I captured student's relative locations at 1 second increments as they move around the lab. This information is useful for studying relative student collaboration.

Student Motivation and Sentiment – I have a mobile platform for asking students multiple choice, likert scale and free response questions.

Student Dialogue Capture – A head-mounted microphone was used to capture student dialogue throughout the lab.

Data Analysis

As one can imagine, this data can conceivably lend itself to a host of analyses. To be clear though, the majority of the analyses that I will be undertaking, at least initially, is grounded in previous work that we've done and will be facilitated through computer-based analysis techniques (Worsley and Blikstein 2011a, Worsley and Blikstein 2011b, Worsley 2012a, Worsley 2012b, Worsley and Blikstein 2012a, Blikstein 2011). I will also do some more traditional analysis of the data. Doing this will involve hiring an organization (or group of individuals) to do the following: transcribe audio at the utterance level, with time stamps and a set of labeling conventions typically used in linguistic and discourse analysis; annotating student transcriptions.

Again, the collection of the raw data and the annotated data will be used in computer-based techniques. These techniques make it more reasonable to work with this volume of data.

I will pay particular attention to a subset of the data for the purposes of having a strong deliverable for this project. This subset will consist of the transcribed audio from student interviews; classroom dialogue; as well as the gesture and object tracking data collected through the Kinect sensor and web cameras, from one-on-one interview. These have been selected because they are most closely related to my research questions around 1) student language and conceptual development, as well as 2) the development of process oriented learning assessments. Furthermore, they represent two different classes of analysis, with the former being well-grounded in traditional qualitative analysis while the later represents a challenging yet fruitful exploration that may be able to generalize to a host of learning processes.

In the analysis of gesture and object data, I constructed a state-based representation of each configuration of objects and upper-body gestures that a student uses while completing their building task. From these states, l examine the extent to which students pass through similar configurations of objects and gestures at their final solution, as well as the cognitive intuitions that aid students in deciding which design configuration to use next. The hypothesis here is that there is rich information embedded in the process that students use to build things. Furthermore, I hypothesize that this type of representation can be fruitful in understanding a student's current mental models, and perhaps identify different shared intuitions that they may be using.

Summary of Persistent Capture Data Results

Student speech was captured throughout their participation in the program. However, due to complications with consistently capturing their data, as a result of student absences and students forgetting to turn their audio capture devices after lunch, doing a systematic analysis of the amount of student speech throughout the program was quite challenging. However, I do still see from the following plot that there are noticeable fluctuations in student speech patterns. While this in itself, does not tell us what is happening, this is an important piece of information to have in order to help me better identify which segments of audio merit additional human coding.

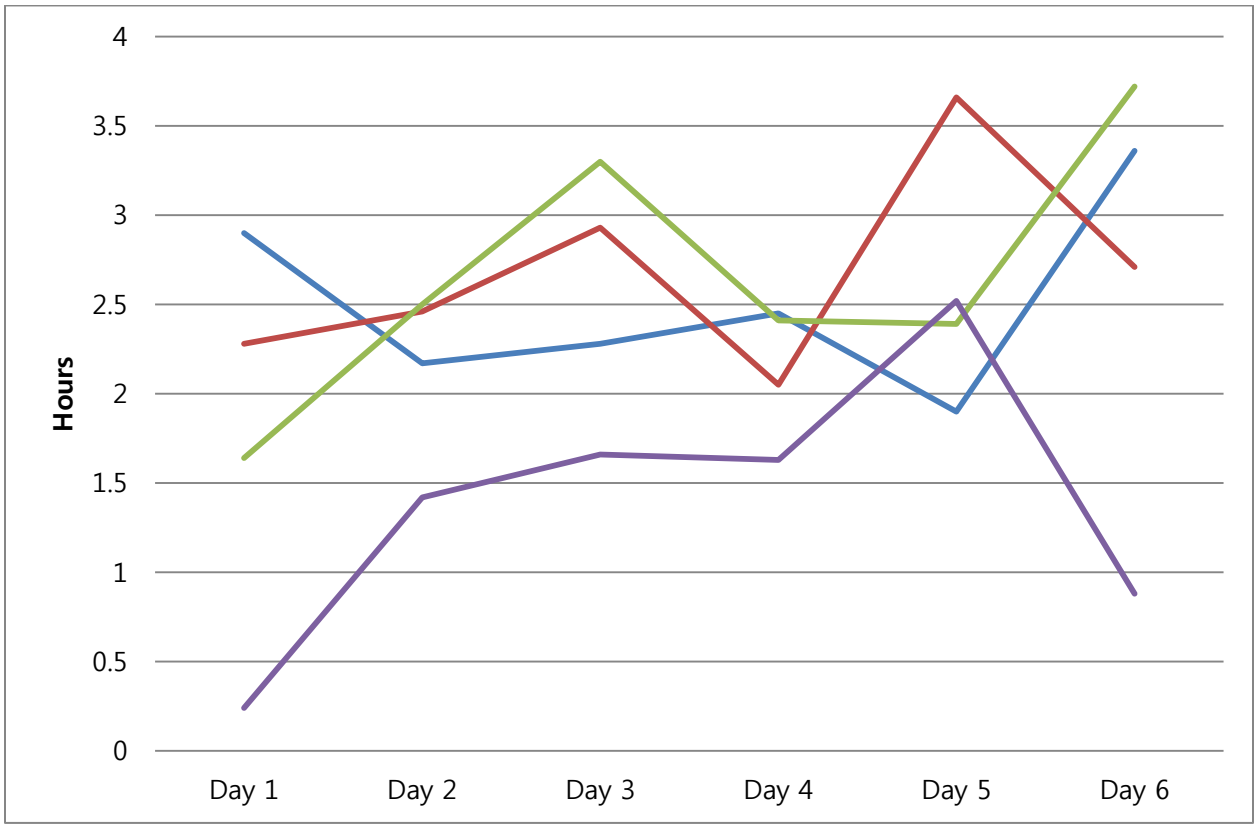


Figure 1 - Sample Student Speech across Six Days

In addition to the above, I also had some coders look at the topicality of each audio segment. When doing this, I found that most of the students spent the majority of their time engaging in discussions that pertained to their projects.

In addition to looking at speech, I also captured information about student motivation. Here I again see fluctuations in the level of motivation that each student experiences over the course of three days

in the program. This data is again useful to better pinpointing important segments of audio that I should study moving forward. Since the data is time-stamped, it will allow me to see what is happening before and after students' motivation, or frustration, changes.

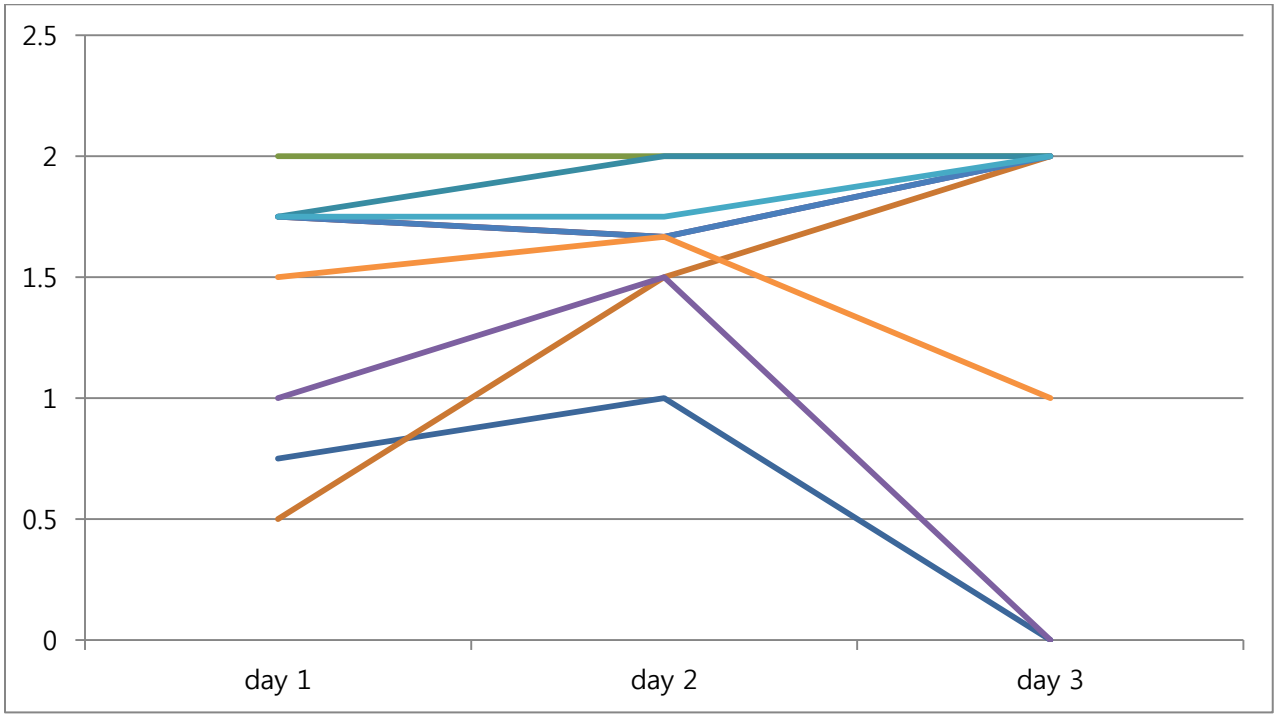


Figure 2 - Sample Student Motivation across Three Days

Finally, I also recorded estimated collaboration data. Like the motivation and speech data, the collaboration data will also be useful for pinpointing important segments of student speech that should receive greater attention. That said, the classifier that I trained was mostly able to predict non-collaboration at a high degree, so in future work I'll work to build a better way for actually predicting when students are collaborating. The other thing that I noticed was that students tended to work in the same groups during the entirety of the program, thus the collaboration data may not be as useful for studying network structure for this population of students.

Analyzing Student Gesture in Building Activities

Data is drawn from thirteen participants. Each participant is given everyday materials, and asked to build a tower that could hold a mass of approximately 3 lbs. Participants were also challenged to make the structure as tall as possible. Figures 3 and 4 depict structures created by two different participants.

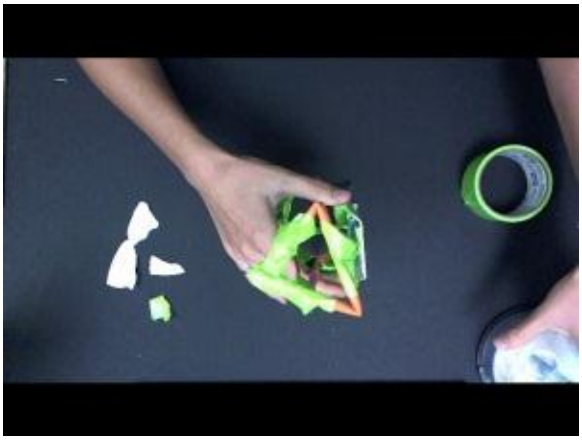


Figure 3 - Sample Expert Structure

Figure 4 - Sample Novice Structure

The task was designed to successfully students are able to take their intuitions about mechanical engineering and physics and translate them into a stable, well-engineered structure. As such, I expected students to use knowledge about forces, symmetry, and the affordances of different geometric objects, to enable them to complete the task. The additional challenge of making the structure as tall as possible was introduced to push all students, regardless of expertise, to the limits of their ability.

An additional design consideration for this task was the existence of explicit metrics for measuring the success of their work. These metrics include whether the structure could hold the mass, how tall the structure is and how long the structure is able to hold the 3 lbs. mass.

In terms of the actual building task, students were given four drinking straws, five wooden popsicle sticks, a roll of masking tape and a paper plate; and were told They were told that they would receive approximately ten minutes to complete the activity. However, they were permitted to work for as long as they wanted, with participation time ranging from eight minutes to fifty-two minutes.

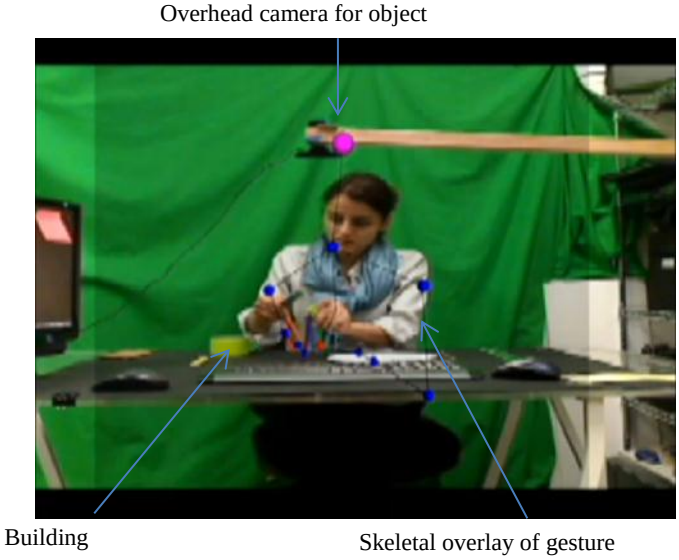


Figure 5 – The Data Capture Environment

Figure 5 depicts the capture environment used to record the audio, video and gesture data streams. Audio was used to capture meaningful utterances made by the participants, though students were not required to engage in a think-aloud. Audio was also captured of each student's metacognitive analysis of their building approach. Video captured the movement of objects as students progressed through the task, while gesture data, which consisted of twelve upper-body parts, recorded the students' physical actions.

Defining Expertise

Prior to the study students were classified based on their perceived level of expertise in the domain of engineering design. Expertise was primarily based on participants' previous experience with engineering design. Such experiences could be in either a formal or informal context. More specifically classification was made along two main dimensions. The first dimension pertains to the amount of formal instruction students had received in engineering. Individuals who had completed bachelors or graduate degrees in engineering were labeled as experts. The second dimension for

determining expertise in engineering was based on observations that the researchers made while watching the students over the course of more than two-hundred hours in an engineering and digital fabrication class. As a part of these two-hundred hours of observation, the researchers also had the chance to learn about the ways that participants engaged in engineering activities in extra-curricular activities and at home.

This definition of expertise resulted in a population of three experts (graduate students in mechanical engineering), two high expertise students, five medium expertise students, and three low expertise students.

Coding

In order to establish a basis for comparing across the students, I created a coding scheme. This coding scheme consists of eleven object manipulation codes. This set of codes was identified through open coding of a sample of the videos, and agreed upon by a team of research assistants. The codes are entirely based on participant object manipulation, or lack thereof, and are not an attempt to explicitly interpret a student's intentions. Nonetheless, I would argue that in most cases, the codes are necessarily tied to user intent, since they are strictly action oriented.

Table 1- Fine-Grain Object Manipulation Codes

Code	Description
Building	Joining things together by tape or other means that is relatively permanent.
Prototyping Mechanism	Seeing if putting two (or more) things together will work well. This could also include acting out a mechanism with the materials.
Testing Mechanism	Involves testing of a subsection of the overall system.
Undoing	Taking things apart as to make a change to a previous build.
Single Object Examination	Pressing or bending on an object to explore its properties
Thinking without an object in hand	Simply surveying the pieces, but not touching anything, or actively doing anything.
Thinking with an object in hand	Not building, or testing the objects properties explicitly, but still holding the object.
System Testing	Putting force on a collection of relatively permanently affixed pieces to see if they will hold the mass
Organizing	Repositioning the raw materials but not actually building,

	examining or prototyping.
Breaking	Breaking apart sticks, bending straws, or ripping tape (in an usual way)
Adjusting	Often times involves moving something to slightly reposition it, or applying more tape to make something stay better.

Using the above codes I was able to condense the students' actions into comparable sequences of time-stamped codes. These codes will serve as a primary data source for the analysis described in the following section.

Object Manipulation Data Analysis

Sequence Construction

I begin the automated portion of the analysis with the time-stamped action logs for each student. I first compress similar action codes. More specifically, I compress the codes to the following five classes:

Table 2 - General Object Manipulation Action Classes

Class	Codes
BUILD	Building and Breaking
PLAN	Prototyping mechanism, Thinking with or without an object, Single object examination, Organizing and Selecting materials
TEST	Testing a mechanism and System testing
ADJUST	Adjusting
UNDO	Undoing

With these more general classes of behaviors, I construct a sequence of user actions that are based on half-second increments. Thus, for each user I have an ordered list of actions, as observed every half a second.

Sequence Segmentation

Each sequence of actions is then segmented any time a TEST action occurs. My assumption is that I need to have a logical way for grouping sequences of user actions and each time a user completes a TEST action, they are essentially signaling that they expect for their previous set of actions to produce a particular outcome. Each segments is recorded, based on the proportion of each of the five action classes (BUILD, PLAN, TEST, ADJUST, UNDO) that took place during that segment. Put differently, I

now have a five dimensional feature vector for each segment, where each dimension corresponds to one of the action classes. As an example, consider the following set of codes:

PLAN, PLAN, BUILD, TEST, ADJUST, UNDO, BUILD, TEST

This sequence of 8 codes would be partitioned into four segments. The first segment would be PLAN, PLAN, BUILD; the second would be TEST; the third would be ADJUST, UNDO, BUILD; and the fourth would be TEST. These four segments would then be used to construct four feature vectors based on the proportion of each of the action classes. Accordingly, I would have the following:

Table 3 - Sample Segmented Feature Set

Segment	ADJUST	BUILD	PLAN	TEST	UNDO
1	0.00	0.33	0.67	0.00	0.00
2	0.00	0.00	0.00	1.00	0.00
3	0.33	0.33	0.00	0.00	0.33
4	0.00	0.00	0.00	1.00	0.00

Segment Standardization

Each column of the feature set is then standardized to have unit variance and zero mean. This step is taken in order to ensure that there are no biases when I perform clustering in the next step.

Segment Clustering

Following standardization the segments are clustered into ten clusters using k-means. Each segment is now associated with one of ten clusters. Each participant's action sequence is reconstructed to reflect one of the ten clusters for each segment, recalling that the action sequence is segmented based on TEST.

Dynamic Time Warping

Finally, dynamic time warping [15] is used to compute the minimum distance between each pair of participants. The distance between two clusters is determined by the cluster centroids from k-means, and is based on Cosine distance. This computation yields an n-by-n matrix of minimum distances, where each distance is normalized by the length of the vectors being compared.

Distance Clustering

The n-by-n matrix from the dynamic time warping calculation is standardized along each column, before being used to construct the final clustering, again with k-means. In order to compare the

clusters to expertise classifications, I find the cluster to expertise alignment that minimizes the total error.

In summary, this algorithm converts an action sequence into segments based on when a subject tests their structure or tests a mechanism. The proportions of actions in the different segments are used to find representative clusters, which are used to re-label each users sequence of segments. Finally, I compare sequences across participants and perform clustering on the pair-wise distances in order to find a natural grouping of the participants.

Gesture Data Analysis

The gesture data analysis, while similar in spirit to the object manipulation analysis, involves markedly less complexity. This is partially due to the particularly fine-grained nature of the data, which was captured every millisecond. Capturing millions and sometimes billions of data points for each user and attempting to use these for doing sequence alignment is a computationally expensive task, which I may endeavor to explore further in later work. Instead, for this analysis I take a simpler approach. This approach is motivated by an observed difference between the amounts of two-handed, coordinated, movement among individuals of differing expertise. Here I consider two-handed coordinated movement to be when a participant is using both of their hands within a given action. Figures 5 and 6, which graph the cumulative displacements for the right and left hand, depict this difference. The expert's hands typically move in sync with one another, whereas the novice's hand movements are markedly asynchronous. I look to exploit this difference in constructing my algorithm.

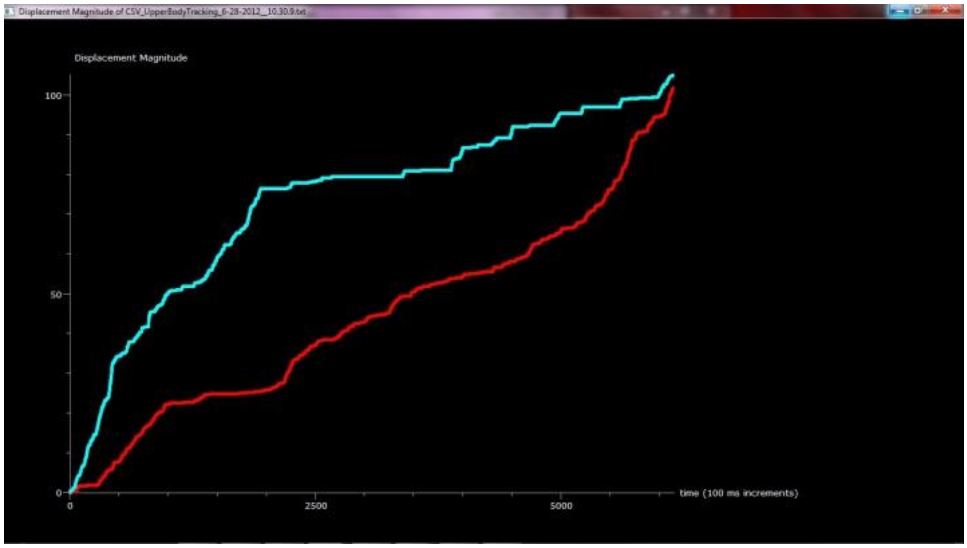


Figure 5 - Novice Cumulative Hand Movements

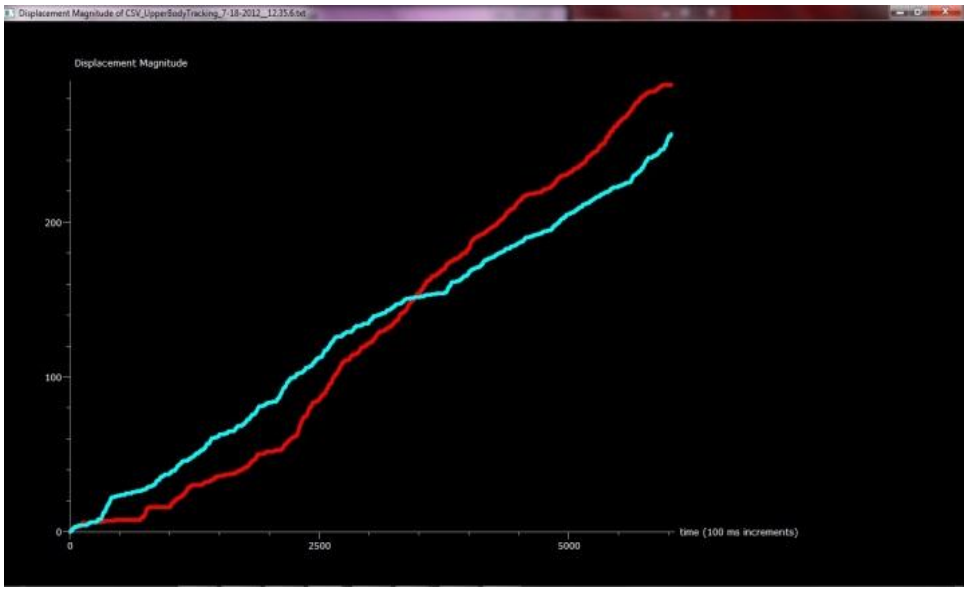


Figure 6 - Expert Cumulative Hand Movements

Given the gesture data from each individual's hands, I begin by constructing a vector based on the absolute difference in the cumulative displacement of their two hands. I then sample each of those distributions at five percent increments, such that all participants will have feature vectors of equal length. These feature vectors are then used to compute the pairwise Euclidean distance between every set of two participants. Those distances are standardized by column, and used as the input for Hierarchical Agglomerative Clustering, with four clusters. I tried using K-means clustering also, but found that most students were being assigned to the same clusters. In future work I will more closely examine why Hierarchical Agglomerative clustering was most successful for this analysis. Finally, the clusters are aligned to the levels of expertise as to minimize the total error.

Results

This study focuses on the nature and frequency of building patterns that I observed among the students, through process-oriented data analysis techniques. In order to motivate the utility of my approach, I begin by taking a static, non-process-oriented, view of the students' actions. Here I take non-process-oriented to mean that instead of looking at the entire sequence as an ordered set of data points, I will only look at the data in aggregate.

Non-Process Oriented Analysis

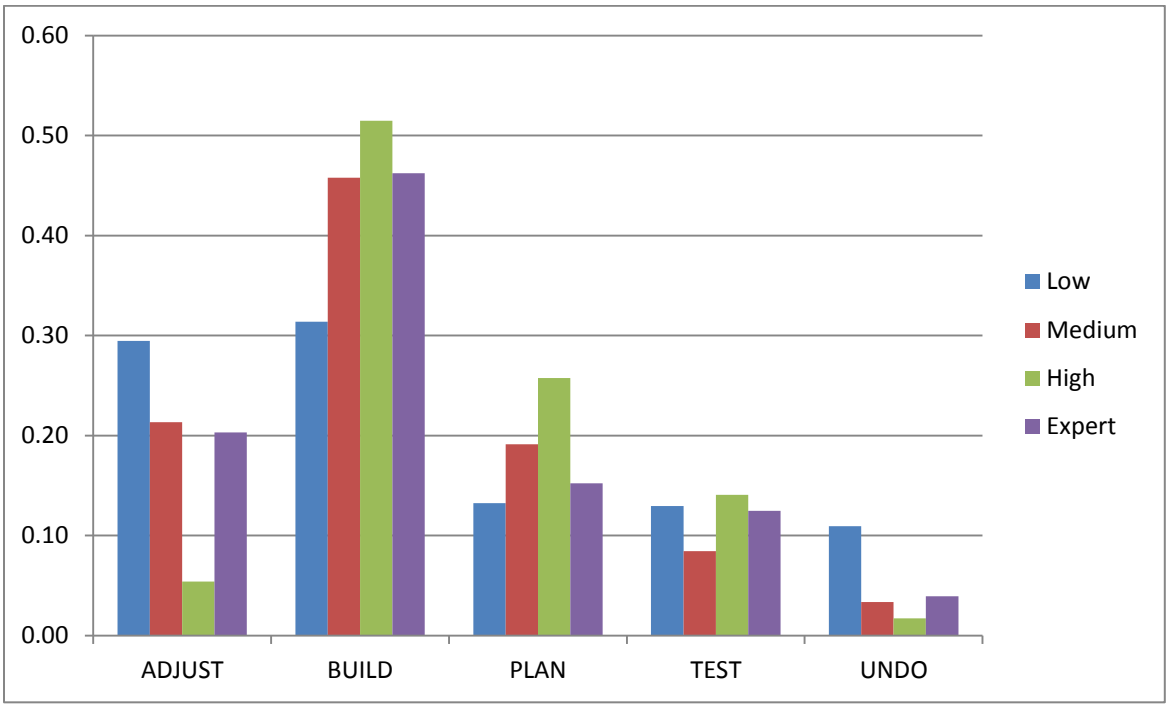


Figure 7 - Proportion of Object Manipulation Classes by Expertise

Figure 7 presents the proportion of time that each student spent on the five general action classes. From the graph it is quite unclear as to how one would go about accurately predicting expertise based solely on these overall proportions. More specifically, there does not seem to be a linear relationship between any of the five general classes and expertise. Instead I see that in some cases, as in the case of time spent in PLAN, experts are most similar to novices. However, in other cases, as in the case of ADJUST (Figure 7), experts and people of medium expertise are the most similar. This is merely one example of a non-linear progression. Nonetheless, I can take these values and learn models that are aligned with expertise. Figure 8 presents the results from a logistic regression model, with 10-fold cross validation, as well as k-means clustering. As a point of comparison, two baseline measures are also reflected in Figure 8. Similar analyses were also completed using other machine learning algorithms: Decision Trees, Neural Networks and Bayesian Networks, but all with similar results. Furthermore, I am cautious about using supervised learning with such a small dataset, because the algorithms are likely to over fit to the data.

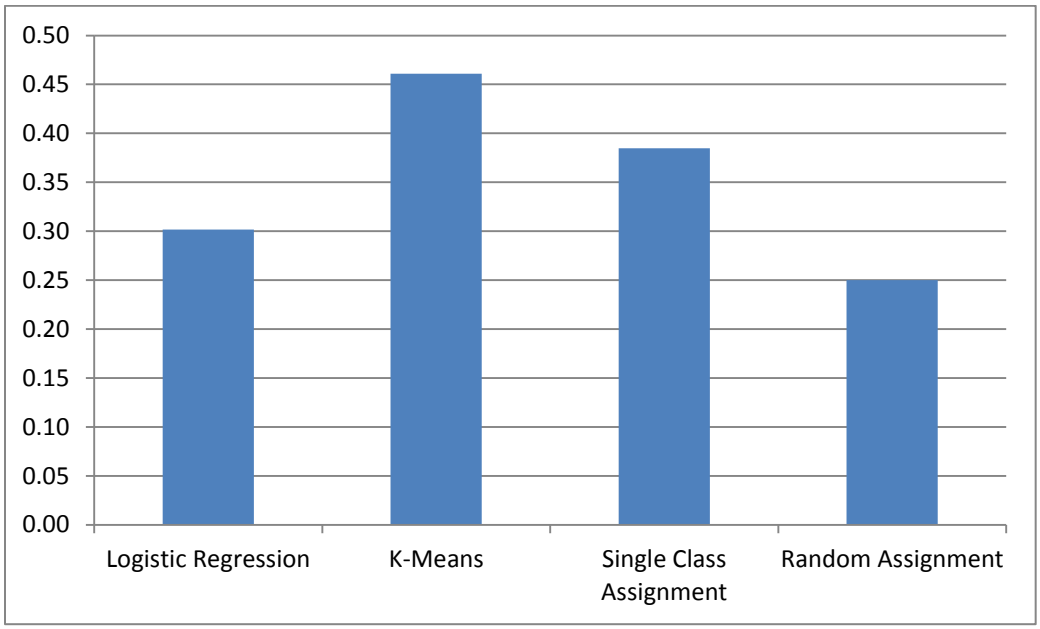


Figure 8 – Classifier Accuracy Based on Proportion of Object Manipulation Classes by Expertise

Another non-process-oriented metric for comparison could be the time spent to complete the task and the overall success of a given build. Table 4 shows the amount of time each student took to complete the task, as well as a binary scoring concerning the success of their structure.

Table 4 - Elapsed time and success for each participant

Subject	Expertise	Time(s)	Success
1	Medium	1387	Yes
2	High	909	Yes
3	Medium	491	Yes
4	Low	1550	No
5	Low	3077	No
6	Medium	1265	Yes
7	Medium	1366	Yes
8	Medium	1373	Yes
9	Low	1730	No
9	Medium	2363	No
10	High	713	Yes
11	Expert	834	Yes
12	Expert	1100	Yes
13	Expert	1122	No

While previous literature would suggest that experts take less time to complete tasks [16] this is only partially true for my population and task. Using these values to differentiate between different levels of expertise worked better than the action code proportions, (see Figure 9). Nonetheless elapsed time and success represent very unsatisfying features. They are unsatisfying because the nature of the problem is not one that would easily align with this paradigm. For example, because of the challenge to make the structure as tall as possible, experts may find themselves spending more time than novices in an effort to perfect their design. This would distort the expected time trend. At the same time, it could also distort my expectations around success, since an expert may take a functioning structure and render it unsuccessful in an effort to make it taller.

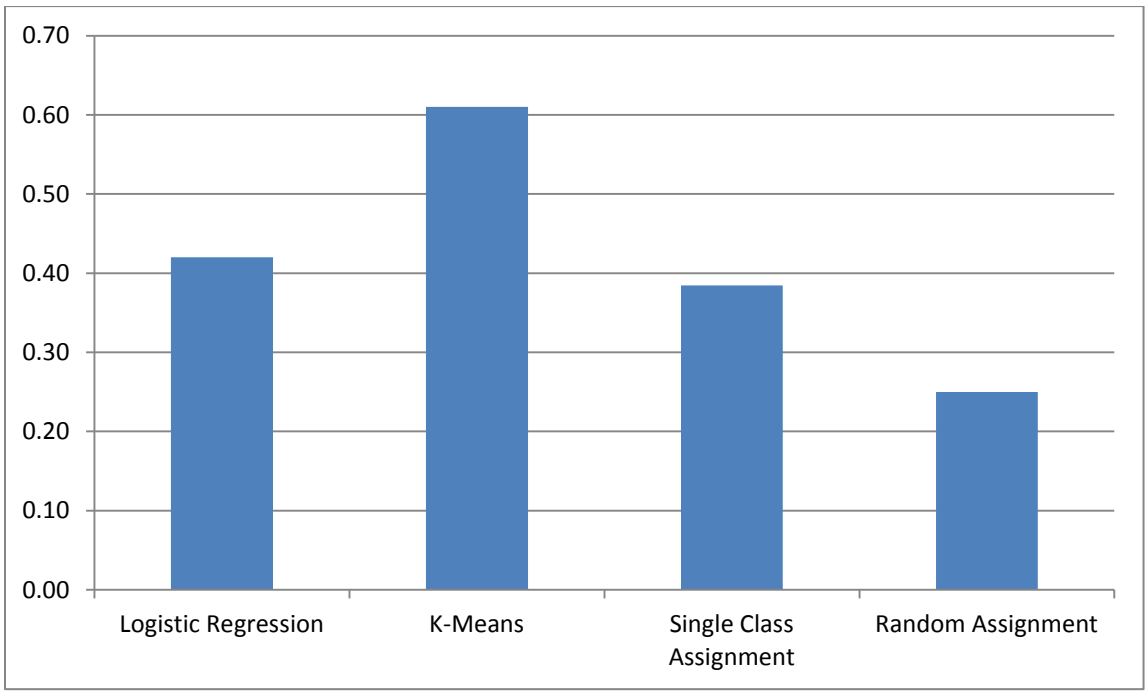


Figure 9 - Classifier Accuracy Based on Elapsed Time and Success

Taken as a whole, these non-process oriented analyses fail to account for the temporality of the data, and the important ways that the temporality of actions is associated with user expertise. At the same time, simply using time and success takes a very naïve view of expertise and begs for an algorithm that can more closely capture the nuances of expertise.

Object Manipulation Results

In contrast to the non-process-oriented approach, my object manipulation analysis algorithm is able to significantly outperform both random assignment and majority class assignment, all while

preserving the process-oriented nature of the task. Figure 10 highlights the accuracy attained through my object manipulation analysis, and the other techniques, keeping in mind that my approach has been completely unsupervised.

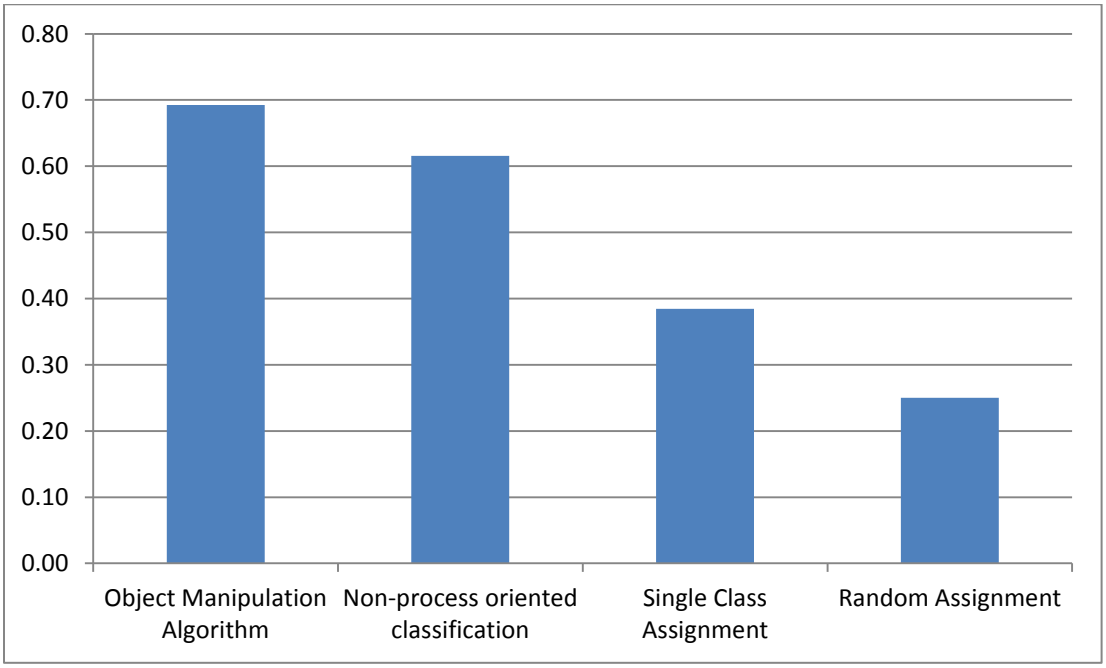


Figure 10 – Classifier Accuracy Based on Object Manipulation Algorithm as Compared to Other Techniques

Similarly, the confusion matrix derived from my work is seen in Table 5.

Table 5 - Confusion Matrix of Expertise

	Low	Medium	High	Expert
Low	3	0	0	0
Medium	3	1	1	0
High	0	0	2	0
Expert	0	0	0	3

From the confusion matrix I see that the algorithm worked best at uniquely clustering expert behavior which it did at an accuracy of 1. It also attained recall of 1 for individuals of low expertise. However, for those individuals of intermediate levels of expertise, the algorithm was less accurate, but was still able to do a reasonable job, considering that my metric of expertise may be noisy for participants of medium expertise.

Of additional interest is the cluster centroids for the segments, as these elucidate what each cluster segment represents. Figure 11 highlights these differences along the dimensions of the five general object manipulation action classes (the cluster centroids that I discuss here do not correspond to clustered students, but the clusters of different segments.) Showing the clusters centroids for the students would only show how different each cluster is from the other clusters based on average dynamic time warp distance.

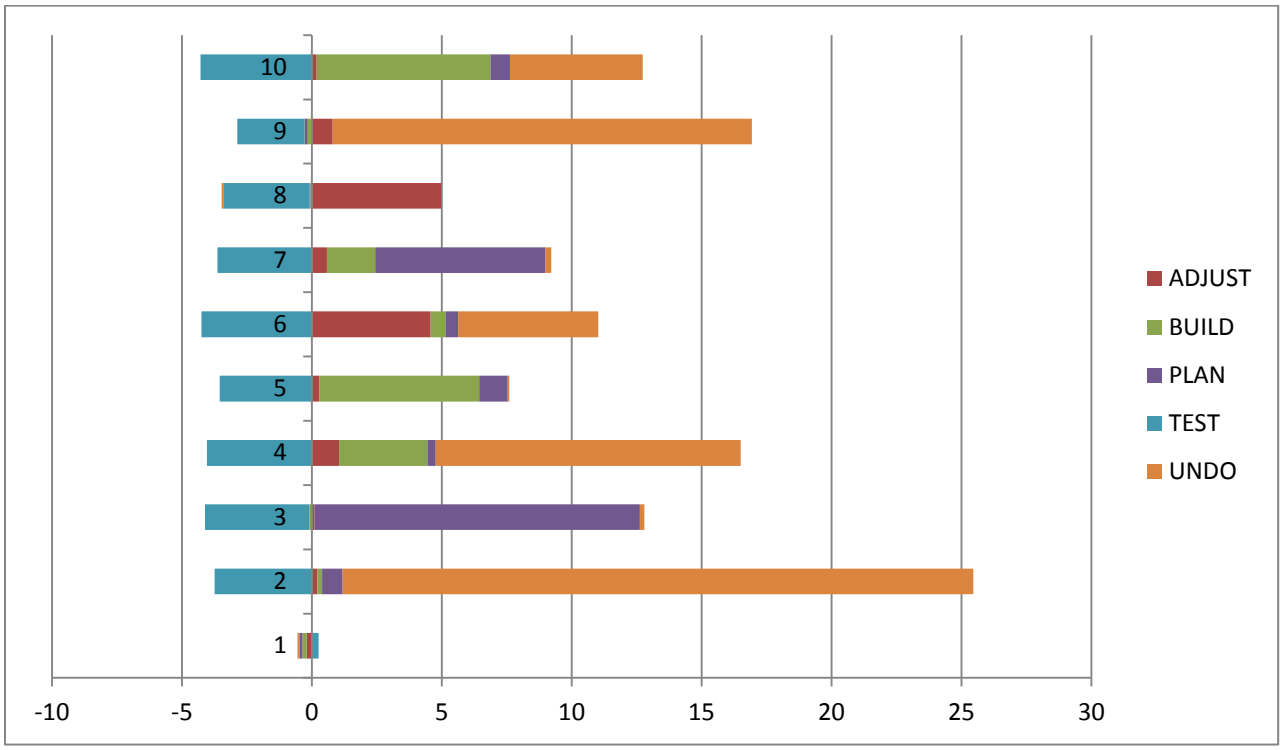


Figure 11 – Cluster Centroids from K-means Clustering

TEST Cluster

Cluster 1 represents the TEST action, and was used for segmenting the sequence of actions. Accordingly, I expect for this to be small in magnitude, and for all of the other clusters to include below average TEST action proportions.

UNDO Clusters

Beyond this, one immediate observation is the amount of UNDO actions. For clusters 2, 4, 6, 9 and 10, undoing represents the primary component of that segment. This, on the whole, suggests that undoing is an important behavior to pay attention to when studying expertise. However, simply looking at UNDO by itself is not sufficient. Instead, one needs to observe what other actions are taking place in the context of the UNDO action. In the case of cluster 2, the user is performing

significant UNDO actions in the absence of any other action. This is in contrast to cluster 4, for example, where the user is completing a large number of UNDO actions, but is also doing several BUILD actions. From this perspective, cluster 2 seems to correspond to doing a sustained UNDO, without any building. An example of this would be a student completely deconstructing their structure. Cluster 4, on the other hand, is more akin to undoing a few elements of one's structure with the intent of immediately modifying the structure. These may be more microscopic UNDO actions, whereas cluster 2 consists of more macroscopic UNDO actions. Clusters 6 and 9 appear to be characterized by a combination of UNDO actions and ADJUST actions. So in this case, the user is undoing, not to make large structural changes to their design, but to make small adjustments. Cluster 6 differs from cluster 9, however, in that cluster 6 also contains both BUILD and ADJUST elements, as well as more PLAN actions.

PLAN, BUILD, ADJUST Clusters

The remaining clusters, 3, 5 and 7, involve few UNDO actions, but can be characterized as different combinations of PLAN, BUILD and ADJUST. Cluster 3 almost exclusively consists of PLAN actions, whereas clusters 5 and 7 primarily include BUILD and PLAN actions.

In summary I see that six of the cluster centroids play a large emphasis on UNDO actions, and the context that they appear in while the remaining four are aligned with different proportions of TEST, PLAN, BUILD and ADJUST actions.

Finally, to make sense of the different levels of expertise, I present the transition probability tables for each of the levels of expertise. These tables show how likely a student is to transition from one cluster to another based on level of expertise.

Table 6 - Transition Probabilities for Students of Low Expertise

	0	1	2	3	4	5	6	7	8	9
0	0.955	0.001	0.004	0.001	0.012	0.001	0.003	0.018	0.001	0.004
1	0.667	0	0	0	0	0.333	0	0	0	0
2	0.571	0	0.143	0	0.143	0	0.143	0	0	0
3	0.667	0	0	0	0	0	0	0	0	0.333
4	0.875	0.031	0	0	0.094	0	0	0	0	0
5	0.25	0	0	0	0.5	0	0	0.25	0	0
6	0.5	0	0	0.1	0.1	0	0.1	0.2	0	0
7	0.66	0	0	0.019	0	0.019	0	0.302	0	0
8	0	0	0	0	0	0	0	0	0	1
9	0.9	0	0	0	0.1	0	0	0	0	0

Table 7 - Transition Probabilities for Students of Medium Expertise

	0	1	2	3	4	5	6	7	8	9
0	0.947	0.000	0.003	0.001	0.010	0.000	0.004	0.032	0.003	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.750	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000
3	0.500	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000
4	0.950	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.750	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000
7	0.619	0.000	0.000	0.000	0.095	0.000	0.000	0.286	0.000	0.000
8	0.667	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8 - Transition Probabilities for Students of High Expertise

	0	1	2	3	4	5	6	7	8	9
0	0.960	0.000	0.002	0.002	0.002	0.000	0.000	0.031	0.000	0.004
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.667	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000
5	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.722	0.000	0.000	0.000	0.000	0.056	0.000	0.222	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 9 - Transition Probabilities for Experts

	0	1	2	3	4	5	6	7	8	9
0	0.925	0.000	0.004	0.000	0.028	0.000	0.006	0.034	0.000	0.002
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.750	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.692	0.000	0.000	0.000	0.308	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000

6	0.600	0.000	0.200	0.000	0.200	0.000	0.000	0.000	0.000	0.000
7	0.591	0.000	0.000	0.000	0.091	0.045	0.000	0.273	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The four tables (6-9) show how individuals of different expertise transition between states differently. For example, the experts transition into 5, 6 and 9 with statistically different probabilities than novices.

Gesture Analysis Results

The gesture analysis also yielded promising results. Recall that here I used the difference between the cumulative displacement of the right hand and the cumulative displacement of the left hand.

Table 10 - Confusion Matrix from Gesture Analysis

	Low	Medium	High	Expert
Low	1	2	0	0
Medium	1	2	1	1
High	0	1	0	1
Expert	0	0	1	2

From the confusion matrix in Table 10 I see that the gesture channel appears to be less conclusive than the action code modality. And, in fact, this is expected given the fact that I was unable to take as fine-grained of an approach to this analysis. The results are also reflective of only looking at a single set of gesture data points, namely the hands. That said, when I relax my levels of expertise to simply be binary, I see that the algorithm performs significantly better (see Table 11)

Table 11 - Confusion Matrix from Binary Expertise Gesture Analysis

Expertise	Low-Medium	High-Expert
Low-Medium	6	2
High-Expert	1	4

Again, this resulted in an accuracy of .77, surpasses accuracy from single class assignment, .62. Thus, while it is apparent that this model does not perfectly segment the data, is does correlate with previous findings concerning two-handed inter-hemispheric interaction [17]. More specifically, previous work on the brain has identified that two-handed interaction is crucial for successful problem solving. By using two hands, individuals can simultaneously engage the right and left hemispheres of the brain. Doing so permits them to create new ideas, which are mediated by the

right hemisphere, and logically choose which of those ideas to utilize, which is mediated by the left hemisphere. These results can therefore be interpreted to suggest that more expert individuals are able to engage both of the processes needed to successfully solve the problem: idea generation and logical selection of the appropriate idea. Furthermore, this ability to select the most applicable idea is analogous to the reprioritization and appropriate use of intuitions that diSessa [6] observed in his expert-novice comparisons. Thus it may not be that the novices are unable to develop the same ideas, it may instead be that they are less capable of identifying which of their structural building ideas to use, and *when* each one should be used. As I will describe later, future research will help us explore this theory in more detail.

Discussion

Pedagogical Considerations

From a pedagogical perspective, I would like to begin this discussion by first taking a moment to acknowledge the non-traditional, yet well-received nature of this form of assessment on the part of the students. Many of the students that I work with have difficulty fully engaging with STEM content. The students often times require frequent encouragement from their instructors in order to successfully complete their assignments, and, if left alone, will quickly deviate from their assigned task. However, for a number of these students, the construction of the simple tower as a form of assessment, not only increased their engagement, but caused some to ask for additional opportunities to demonstrate their knowledge through building. This is largely because the activity didn't feel like a test, but, instead, was a fun engineering challenge. In particular, one student, who typically was shy and apprehensive about attempting to tackle STEM assignments, experienced a significant boost in confidence from participating in the building task. This is merely to suggest that at least for the population of students that tend to struggle within traditional STEM classrooms, making available to them novel forms of assessment that allow them to demonstrate their knowledge through other means represents a promising opportunity.

Object Manipulation Analysis Discussion

Moving now to the results of the object manipulation analysis, I see three primary contributions. On the whole, I have presented an algorithm that can effectively be used to group students based on the actions that they take while participating in the building of simple machines and structures. A key component of this algorithm is the identification of the appropriate unit of analysis. I showed that looking at the proportion of different actions across the entire building task fails to generate meaningful comparisons. Instead one should use an approach that captures the temporality of the data. I also explored the use to constant time based segmentation - segmenting every 10 seconds, for example - and normalized time based segmentation - segmenting every five percent of

someone's codes - however, neither of these approaches were met with success. Instead, segmentation should take place based on mechanism testing and system testing, as it's these actions that appear to accurately represent a unit of work.

Another key insight has to do with the nature of collapsing the original eleven codes. Collapsing codes has important cognitive and computational implications. Given that I would like to enable automatic labeling of the different actions taken by a participant, code collapsing makes this increasingly feasible. Instead of having to identify very fine grained, hard to detect differences between building and breaking, for example, the action classification algorithm will only need to be trained on five classes of actions. From a cognitive perspective, these findings may suggest that while an observer may see the activities in each state, prototyping a mechanism or examining an object, for example, as distinct activities, sets of activities may actually serve the same cognitive role within the participant. This is to say that prototyping a mechanism may be cognitively the same as examining an object – and I can say they are the same because it appears as though individuals of the same level of expertise use them in similar ways, as they plan their design. Nonetheless, further analysis is required to gain additional insight into these potential cognitive similarities.

Finally, the algorithm provides a very fine-grained representation of the action "states" that are salient for the data set. Following the first instantiation of k-means, I was left with a set of representative "states" that were shared across several participants. Recall that each state consisted of the proportion of time spent doing each of the five general action classes, within a given segment. This representation of the action states is several levels of granularity beyond what could reasonably be inferred by a human observer. Instead, humans tend to be limited to seeing "states" that are largely characterized by a single action code. For example, a human may be inclined to group all UNDO actions into the same "state," when, in fact, the context in which UNDO actions are happening is very important. My analysis is able to get "states" that are characterized by relative proportions of all of the action codes. This provides a much more precise representation of the different "states" and helps in articulating a clearer difference among participants of differing expertise.

Gesture based analysis

The gesture based analysis also produced a number of key findings. First, there are clearly correlations between the gestures individuals make and the object manipulation action that they undertake. This finding is inferred from the fact that both techniques were able to yield relatively accurate results. This, again, may be useful for improving automatic detection of object manipulation actions. Additionally, the analysis was able to make use of a theory concerning two-handed coordination and the implications that this has on problem solving. In my case I found that two-handed coordinated

actions were correlated with expertise. It is my conjecture that there are additional theories related to embodied cognition that can be discovered or leveraged in research concerning building-based assessments.

Finally, the gesture-based analysis highlights a potential area of easy intervention for trying to effect behavioral changes among students. Though I have yet to explore these interventions, one can imagine showing a student a plot of their own hand movements while they are participating in a building task, and see how this additional awareness of their body movements either helps, or hurts their ability to successfully complete the task. Such an intervention could be enhanced by sharing with the student knowledge about two-handed inter-hemispheric interactions, to see how this helps the student perform more like an expert.

Looking at the analysis as a whole, I am looking to motivate the development of authentic, process-oriented assessments that can be enacted in minimally instrumented environments. My interest in doing this is to create additional ways for validating student learning in project-oriented environments. This goal is also grounded in a desire to develop techniques that can eventually be utilized within both formal and informal learning environments.

In future work, I plan to combine my data capture technique with a think-aloud protocol, as so I can begin to align user actions and user cognition more explicitly. I will also endeavor to study how collaboration influences the emergence of expert-like behaviors. Finally, I will continue to work towards developing techniques for automatically labeling user object manipulation actions during the task explored in this analysis, as well as with other tasks.

Summary

In this report, I have focused on two ways for studying student learning in project based learning environments. First, by using persistent data capture of speech, motivation and collaboration, I motivated the ways that this data can help me pinpoint the specific segments of speech that merit human analysis and transcription. This is important, given that I began with thousands of hours of audio. In the latter part of this report I presented a pair of techniques for analyzing and detecting expertise as recognized through object manipulation and gestures. In so doing, I identified key elements in how to segment and compress object manipulation codes, while also showing how dynamic time warping combined with clustering can be used to accurately classify student expertise. In addition to classification, I have generally motivated the use of multimodal learning analytics for supporting authentic, process-oriented assessments, as this technique has permitted us to realize a more fine-grained level of expertise delineation than could have been reasonably perceived by a

human. Finally, the approach has made it evident that meaningful analysis can be gleaned from simply watching and measuring student actions as they participate in building tasks, a realization that I hope will encourage other researchers to embark upon this promising, yet challenging, area of study.

As a whole, this report will hopefully contribute to on-going work that studies learning processing in project based learning environments using multimodal learning data. I believe that this type of work will be important for the future of education, and for the development of new learning technologies.

References

Abrahamson, D. (2009). Orchestrating Semiotic Leaps from Tacit to Cultural Quantitative Reasoning—The Case of Anticipating Experimental Outcomes of a Quasi-Binomial Random Generator. *Cognition and Instruction*, *27*(3), 175-224.

Barron. B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecologies perspective. Human Development, 49, 193-224.

Barron, B. (2004). Learning ecologies for technological fluency: Gender and experience differences. Journal of Educational Computing Research, 31(1), 1-36.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. Proceedings of the I Learning Analytics Knowledge Conference (LAK 2011), Banff, Canada.

Dewey, J. (1913). *Interest and effort in education*. Cambridge, MA: The Riverside Press.

Dewey, J. (1897). My Pedagogic Creed. *School Journal 54*, 77-80.

diSessa, A.A. 2002. Why "conceptual ecology" is a good idea. In M.Limón & L.Mason (Eds.), Reconsidering conceptual change: Issues in theory and practice (pp. 29–60). Dordrecht: Kluwer.

Forbes-Riley, K. and Litman, D. 2010. Metacognition and Learning in Spoken Dialogue Computer Tutoring. Proceedings 10th International Conference on Intelligent Tutoring Systems (ITS), Pittsburgh, PA.

Forbes-Riley, K., Rotaru, M. and Litman, J. (2009). The Relative Impact of Student Affect on Performance Models in a Spoken Dialogue Tutoring System. User Modeling and User-Adapted Interaction (Special Issue on Affective Modeling and Adaptation), 18(1-2), February, 11-43.

Kelly, T and Copabianco, B. (2012) Think-aloud Protocol Analysis as a Measure of Students' Science Learning through Design Assessment. Paper Presented at National Association for Research in Science Teaching Annual Meeting, March 25-28, 2012, Indianapolis, IN.

Li, M. Ruiz-Primo, M.A. & Shaveloson, R.J. (2006). Towards a science achievement framework: The case of TIMSS 1999. In S. Howie & T. Plomp (Eds.), Contexts of learning mathematics and science: Lessons learned from TIMSS (pp. 291–311). London: Routledge

Litman, D., Moore, J., Dzikovksa, M. and Farrow, E. (2009). Using Natural Language Processing to

Analyze Tutorial Dialogue Corpora Across Domains and Modalities. Proceedings 14th International Conference on Artificial Intelligence in Education (AIED), Brighton, UK, July.

Litman, D., and Forbes-Riley, K. 2009. Spoken Tutorial Dialogue and the Feeling of Another's Knowing. Proceedings 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), London, UK, September.

Papert, S. (1980). *Mindstorms : children, computers, and powerful ideas*. New York: Basic Books.

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012). *Modeling how students learn to program.* Paper presented at the 43rd ACM technical symposium on Computer Science Education (SIGCSE '12).

Shulman, L. S. (2006). From hermeneutic to homelitic: The professional formation of clergy. *Change*. March/April, 28-31.

Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press

Worsley, M. (2011). Qualifying Paper – What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis.

Worsley, M. (2011). Deconstructing Constructionism: An Analysis of Identity and Epistemological Changes from a High School Constructionist Learning Experience. (unpublished)

Worsley, M. (2012). Semi-Supervised Object Tracking. (unpublished)

Worsley, M. and Blikstein P. (2010) Towards the Development of Learning Analytics: Student Speech as an Automatic and Natural Form of Assessment. Paper Presented at the Annual Meeting of the American Education Research Association (AERA).

Worsley, M. and Blikstein P. (2011). What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In Proceedings for the 4th Annual Conference on Educational Data Mining.

Worsley, M. and Blikstein, P. (2012). An Eye For Detail: Techniques For Using Eye Tracker Data to Explore Learning in Computer-Mediated Environments. In the Proceedings of the 2012 International Conference of the Learning Sciences.

Worsley, M. and Blikstein P. (2012). OpenGesture: A Low-Cost, Easy-to-Author Application Framework for Collaborative, Gesture-, and Speech-Based Learning Applications. Paper Presented at the Annual Meeting of the American Education Research Association (AERA)

Worsley, M., Johnston, M. and Blikstein P. (2011) OpenGesture: a low cost authoring framework for gesture and speech based application development and learning analytics. In Proceedings for the 10th Annual Conference on Interaction Design and Children.

Worsley, M. and Huang, Z (2012). Mining through Mobile: Using Smart Phones to Monitor Student Action, Affect and Interaction In Open-Ended Learning Environments. (unpublished)

U.S. Department of Education, Office of Educational Technology (2010). Transforming American Education: Learning Powered by Technology. *National Education Technology Plan 2010*. Washington, DC.