Towards the Development of Multimodal Action Based Assessment

Marcelo Worsley
Stanford University
520 Galvez Mall, CERAS 102
Stanford, CA 94305, USA
mworsley@stanford.edu

Paulo Blikstein
Stanford University
520 Galvez Mall, CERAS 232
Stanford, CA 94305, USA
paulob@stanford.edu

ABSTRACT
In this paper, we describe multimodal learning analytics techniques for understanding and identifying expertise as students engage in a hands-on building activity. Our techniques leverage process-oriented data, and demonstrate how this temporal data can be used to study student learning. The proposed techniques introduce useful insights in how to segment and analyze gesture- and action-based generally, and may also be useful for other sources of process rich data. Using this approach we uncover new ideas about how experts engage in building activities. Finally, a primary objective of this work is to motivate additional research and development in the area of authentic, automated, process-oriented assessments.

Categories and Subject Descriptors

General Terms
Algorithms, Human Factors.

Keywords
Keywords are your own designated keywords.

1. INTRODUCTION
As much as we might like to think otherwise, assessment remains a critical component of the educational system. Whether students are engaged in a formal classroom lesson, or participating in play-based learning, there is the expectation that one can identify a measurable outcome concerning how the student thinks, acts or feels. Systematically demonstrating such learning outcomes in project-based learning environment has long been a challenge faced by education researchers [1, 2]. Early education researchers [3, 4] recognized the merits of project-based learning, but widespread adoption of the practice has largely been hampered by this need to demonstrate its effectiveness at scale. The observed challenge manifests itself in researchers having to choose between traditional assessments that scale, but may be fundamentally inconsistent with the process-oriented goals of project based learning; and finding creative ways to use student portfolios, micro-genetic analysis and ethnographies, all of which are unable to scale to larger populations. Fortunately, we are arriving at a time when the technological tools that are available through machine learning and artificial intelligence can help make process-oriented analyses for project-based learning, more scalable.

Beyond the goal of moving away from traditional assessments, that tend to neglect learning processes, we are also concerned about how some traditional assessments are divorced from the actual practices of the discipline in which they are administered. This is particularly the case within many engineering disciplines. In computer science, for example, it is not uncommon to have students write pseudo code on an exam as a method for assessing their programming proficiency, despite the fact that when writing pseudo code, the student is restricted from utilizing the various tools that may be used when actually programming. Similarly, mechanical engineering students may be asked to derive an equation or prove a theory on an exam, but seldom engage in activities that are directly akin to the practices of the field. In order to fill this gap, our current work looks to advance our ability to understand and utilize forms of assessment that are more closely tied to the practices of their respective disciplines. More specifically, we study patterns in how students of different levels of expertise go about completing the design and construction of simple machines and structures.

At a high-level, this paper intends to:

- Present techniques for doing automated multimodal analysis of student expertise while they engage in building tasks,
- Justify the pedagogical merit of our techniques
- Discuss the implications that these techniques have on the future of assessments, and on our understanding of how expertise is manifested through building.
- Motivate more widespread development and adoption of process-oriented assessments through the use of multimodal learning analytics.

2. THEORETICAL FRAMEWORK

2.1 Constructivism/Constructionism
This work fundamentally builds on Piaget's notion that knowledge is actively and dynamically constructed by the learner based on resources that she already has, and Papert's constructionism [5]. The study takes place within a constructionist learning environment and involves students participating in the physical construction of artifacts. These construction activities give students the opportunity to develop their ideas by completing several cycles of building and debugging. Furthermore, students have an opportunity to explore engineering design in an authentic way that is challenging and engaging.

LAK '13 April 08 - 12 2013, Leuven, Belgium

2.2 Knowledge in Pieces

This work also builds the knowledge in pieces (KiP) framework [6] which considers how students make sense of fundamental concepts in science by dynamically articulating and reorganizing atomistic intuitions about the physical world, rather than making use of robust theoretical systems. Moreover, KiP speaks to the transition from being a novice to being an expert in a given field by reorganizing rather than replacing ideas. According to diSessa [6] experts share many of the same intuitions as novices, but have the additional ability to know when those intuitions apply, how they are connected, and when one must employ other concepts in order to fully understand a scientific phenomenon. While diSessa is primarily concerned with spoken descriptions of phenomena in physics, we hypothesize that similar types of dynamic re-articulation of intuitions could be at play when students engage in simple engineering building activities.

We also borrow from KiP theorists their focus on microgenetic analysis [7] in that we look at student actions over small timescales in order to interpret the mental constructs governing their thinking and actions.

2.3 Multimodal Learning Analytics

Finally, the approach described, follows in a growing body of literature concerned with developing learning analytics [10, 11]). Here we specifically look at approaches that are multimodal in nature, as we posit that identifying student learning likely requires the ability to analyze and synthesize a variety of data streams. Much of the previous work in this area of research has looked at speech, gaze, sentiment and drawings as primary elements of analysis [12, 13, 14]. Here we depart from those and study the interaction of actions and gestures.

3. METHODS

3.1 Data

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 1 and 2 depict structures created by two different participants.

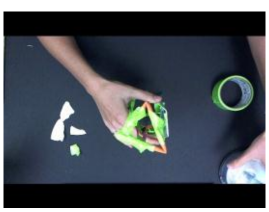

Figure 1 - Sample Expert Structure

Figure 2 - 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, we 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.

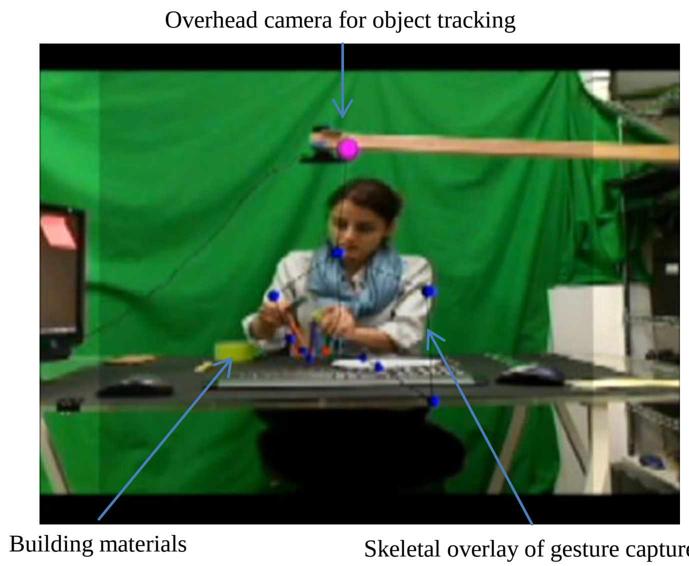

Figure 3 – The Data Capture Environment

Figure 3 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.

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

3.2 Coding

In order to establish a basis for comparing across the thirteen students, we 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, we 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 we were 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.

3.3 Object Manipulation Data Analysis

3.3.1 Sequence Construction

We begin the automated portion of the analysis with the time-stamped action logs for each student. We first compress similar action codes. More specifically, we 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, we construct a sequence of user actions that are based on half-second increments. Thus, for each user we will have an ordered list of actions, as observed every half a second.

3.3.2 Sequence Segmentation

Each sequence of actions is then segmented any time a TEST action occurs. Our assumption is that we 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, we 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, we 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

3.3.3 *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 we perform clustering in the next step.

3.3.4 *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.

3.3.5 *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.

3.3.6 *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, we 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, we compare sequences across participants and perform clustering on the pair-wise distances in order to find a natural grouping of the participants.

3.4 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 we may endeavor to explore further in later work. Instead, for this analysis we 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 we consider two-handed coordinated movement to be when a participant is using both of their hands within a given action. Figures 4 and 5, 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. We look to exploit this difference in constructing our algorithm.

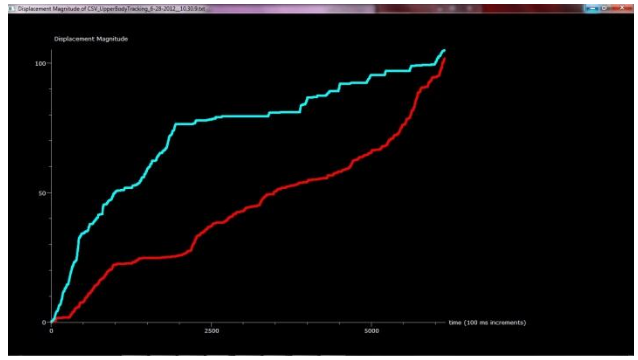

Figure 4 - Novice Cumulative Hand Movements

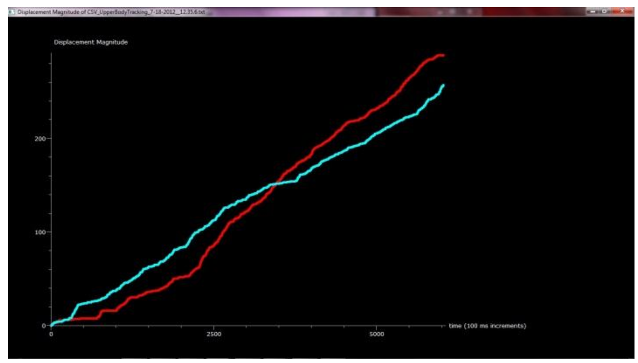

Figure 5 – Expert Cumulative Hand Movements

Given the gesture data from each individual's hands, we begin by constructing a vector based on the absolute difference in the cumulative displacement of their two hands. We 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. We tried using K-means clustering also, but found that most students were being assigned to the same clusters. In future work we 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.

4. RESULTS

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

4.1 Non-Process Oriented Analysis

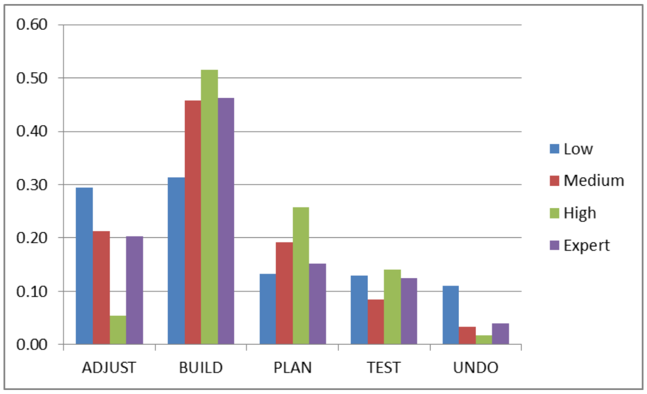

Figure 6 - Proportion of Object Manipulation Classes by Expertise

Figure 6 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 we 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 6), experts and people of medium expertise are the most similar. This is merely one example of a non-linear progression. Nonetheless, we can take these values and learn models that are aligned with expertise. Figure 7 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 7. Similar analyses were also completed using other machine learning algorithms: Decision Trees, Neural Networks and Bayesian Networks, but all with similar results. Furthermore, we are cautious about using supervised learning with such a small dataset, because the algorithms are likely to over fit to the data.

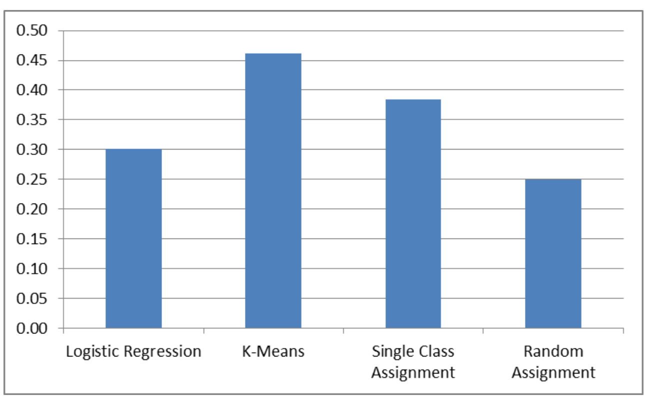

Figure 7 – Classifier Accuracy Based on Proportion of Object Manipulation Classes by Expertise

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 our population and task. Using these values to differentiate between different levels of expertise worked better than the action code proportions, (see Figure 8). 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 our expectations around success, since an expert may take a functioning structure and render it unsuccessful in an effort to make it taller.

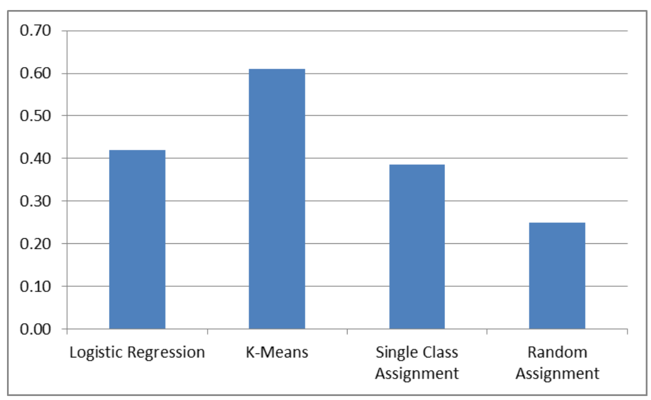

Figure 8 - Classifier Accuracy Based on Elapsed Time and Success

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.

4.2 Object Manipulation Results

In contrast to the non-process-oriented approach, our 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 9 highlights the accuracy attained through our object manipulation analysis, and the other techniques, keeping in mind that our approach has been completely unsupervised.

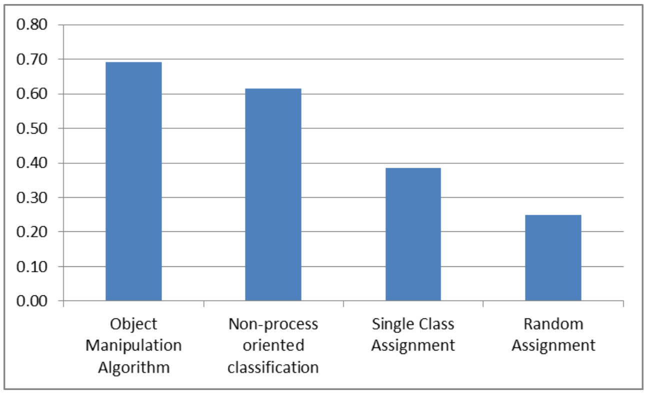

Figure 9 – Classifier Accuracy Based on Object Manipulation Algorithm as Compared to Other Techniques

Similarly, the confusion matrix derived from our 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 we 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 our 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 10 highlights these differences along the dimensions of the five general object manipulation action classes (the cluster centroids that we 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.

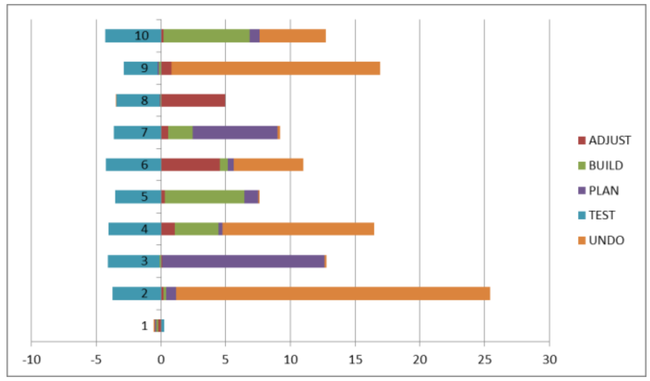

Figure 10 – Cluster Centroids from K-means Clustering

4.2.1 TEST Cluster

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

4.2.2 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.

4.2.3 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 we 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.

4.3 Gesture Analysis Results

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

Table 6 - 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 6 we 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 we were 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 we relax our levels of expertise to simply be binary, we see that the algorithm performs significantly better (see Table 7)

Table 7 - 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 we will describe later, future research will help us explore this theory in more detail.

5. DISCUSSION

5.1 Pedagogical Considerations

From a pedagogical perspective, we 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 we 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.

5.2 Object Manipulation Analysis Discussion

Moving now to the results of the object manipulation analysis, we see three primary contributions. On the whole, we 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. We 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. We 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 we 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 we 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, we were 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. Our 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.

5.3 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 our case we found that two-handed coordinated actions were correlated with expertise. It is our 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 we 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, we are looking to motivate the development of authentic, process-oriented assessments that can be enacted in minimally instrumented environments. Our 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 we plan to combine our data capture technique with a think-aloud protocol, as so we can begin to align user actions and user cognition more explicitly. We will also endeavor to study how collaboration influences the emergence of expert-like behaviors. Finally, we 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.

6. CONCLUSION

In this paper, we have presented a pair of techniques for analyzing and detecting expertise as recognized through object manipulation and gestures. In so doing, we 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, we 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 we hope will encourage other researchers to embark upon this promising, yet challenging, area of study.

7. REFERENCES

[1] Pea, R. (1987). Programming and problem-solving: Children's experiences with Logo. In T. O'Shea & E. Scanlon (Eds.), Educational computing (An Open University Reader). London: John Wiley & Sons.

[2] Barron, B., & Darling-Hammond, L. (2010). Prospects and challenges for inquiry-based approaches to learning: OECD Publishing

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

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

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

[6] 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.

[7] Siegler, R.S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. American Psychologist, 46, 606-620.

[8] Kirsh,D. (2009). Problem Solving and Situated Cognition. In Philip Robbins & M. Aydede (eds.), The Cambridge Handbook of Situated Cognition. Cambridge.

[9] Wagner, S., Nusbaum, H., & Goldin-Meadow, S. (2004) Probing the mental representation of gesture: Is handwaving spatial? Journal of Memory and Language. 50, 395-407.

[10] Siemens, G. and Baker, R. (2012). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12), Simon Buckingham Shum, Dragan Gasevic, and Rebecca Ferguson (Eds.). ACM, New York, NY, USA, 252-254.

[11] US Department of Education (2012) Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. Washington, D.C.

[12] 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. Eindhoven, Netherlands.

[13] Worsley, M. (2012). Multimodal Learning Analytics: Enabling the Future of Learning Through Multimodal Data Analysis and Interfaces. 2012 IEEE International Conference on Multimodal Interfaces (ICMI). Santa Monica, California, USA

[14] Blikstein, P. (2013). Multimodal Learning Analytics: a research agenda. In Proceedings of the 3rd International

Learning Analytics Knowledge Conference (LAK 2013), Leuven, Belgium. ACM, New York, NY, USA, 252-254.

[15] Needleman S. & Wunsch C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology

[16] Anderson, J.R. & Schunn, C.D. (2000). Implications of the ACT-R Learning Theory: No Magic Bullets in R. Glaser (Ed.), Advances in instructional psychology (Vol. 5). Mahwah, NJ.

[17] Hoppe, K.D. (1988). Hemispheric specialization and creativity. The Psychiatric Clinics of North America 11, 3, 303-315.