THE CONSTRUCTION OF A VIDEO CODING PROTOCOL TO ANALYZE INTERACTIVE INSTRUCTION IN CALCULUS AND CONNECTIONS WITH CONCEPTUAL GAINS

Matthew Thomas
University of Central Arkansas

Instruments called concept inventories are being used to investigate students' conceptual knowledge of topics in STEM fields, including calculus. One interactive instructional style called Interactive-Engagement has been shown to improve students' gains on such instruments in physics. In this paper, we discuss the development of a video coding protocol which was used to analyze the level of Interactive-Engagement in calculus classes and investigate the correlation with gains on the Calculus Concept Inventory.

Key words: Calculus, instruction, interactive teaching, conceptual learning

Conceptual understanding has been a recent area of interest in undergraduate mathematics and in other STEM fields. This interest has manifested in the construction of instruments called concept inventories to measure students' conceptual understanding. Previous studies have found correlations between interactive instructional techniques, particularly one called Interactive-Engagement (IE), and gains on concept inventories. In this study, we investigate possible connections between conceptual gains on one such instrument, the Calculus Concept Inventory, and Interactive-Engaged instruction. To measure IE, we constructed a coding protocol that quantitatively measures the level of IE in a classroom. This study serves two purposes: (1) to develop a coding protocol to quantify IE instruction, and (2) to connect the results of the coding protocol with scores on the CCI.

Background

Conceptual Knowledge

Historically, there has been a division between the teaching of computational and conceptual material (Rittle-Johnson, Siegler, & Alibali, 2001). These "sharply contrasting orientations" (A. G. Thompson, Philipp, T. Thompson, & Boyd, 1994, p. 1) can be seen in the recent "math wars," where proponents of traditional mathematics typically emphasize procedural fluency and proponents of reform-based (or standards-based) mathematics emphasize conceptual understanding (Schoenfeld, 2004).

Rittle-Johnson et al. (2001) define *procedural knowledge* as "the ability to execute action sequences to solve problems" (p. 346). In contrast, *conceptual knowledge* is defined as "implicit or explicit understanding of the principles that govern a domain and of the interrelations between units of knowledge in a domain" (p. 346). For example, conceptual knowledge might be indicated by a student's understanding of the relationships between algebraic and graphical representations of functions. One way that conceptual knowledge can be demonstrated is by applying known principles or techniques in new situations. For example, recognition of the same topic, such as optimization, in a different subject area or context gives credence to the claim that conceptual understanding has been obtained (Hughes Hallett, 2006, p. 4).

Concept Inventories

Conceptual understanding may be measured through instruments called concept inventories.

Concept inventories are tests designed to measure the most basic knowledge in a field (Epstein, 2007). Typically, the tests are given in a multiple choice format, and involve no computation. When given as a pretest and posttest, the instruments measure the change in conceptual knowledge students undergo during a course. Many studies of conceptual understanding in physics education use concept inventories (e.g. Hake, 1998; Halloun, 1985; Malone, 2008; Rhoads & Roedel, 1999), and other disciplines are using them with increasing frequency (Libarkin, 2008).

The first concept inventory was the Force Concept Inventory (FCI), written by Hestenes, Wells, and Swackhamer (1992) to measure students' conceptual knowledge in introductory mechanics courses. Drawing upon the FCI, Epstein wrote a 22-question concept inventory for introductory calculus in 2007 (Epstein, 2007, 2013).

Interactively-Engaged Instruction

Interactively-Engaged (IE) instruction has been linked to gains in conceptual understanding as measured by concept inventories. IE was defined by Hake (1998) as a collection of methods designed, at least in part, to promote conceptual understanding through "heads-on (always) and hands-on (usually)" (p. 1) activities which lend themselves to immediate feedback through discussion with peers and/or instructors. In this study we operationalize the concept of an IE classroom in a way that allows IE to be quantitatively measured and explore potential correlations between Interactive-Engagement and gains on the CCI.

Previous studies that consider correlations between concept inventory scores and interactive instruction have relied on instructor and/or student self-reporting to quantify levels of IE in classrooms. We eliminated the need for self-reporting by developing a protocol and coding videos ourselves. Our protocol also allows for the examination of IE as a continuum rather than a dichotomous variable, as has been done before. For example, a study by Prather, et al. (2009) relied on instructor self-reporting of interactivity levels, where questions were designed to determine how frequently "interactive learning strategies" (p. 322) were implemented, and how often students made predictions or were asked questions during class. Rhea's (n.d.) study relied on student and instructor reporting of interactivity levels.

Methods

All students taking introductory calculus in the fall semester of 2010 at a large southwestern university took the CCI as a pretest and posttest. Instructors teaching introductory calculus again in the spring semester of 2011 were invited to participate in the study. Of the ten instructors who taught introductory calculus in both Fall 2010 and Spring 2011, five agreed to be videotaped in the classroom three times during the semester. The student scores and instructor videos were collected during different semesters for logistical reasons, however instructors indicated that they were using the same instructional style both semesters. This difference in timing of data collection should be considered when interpreting the results of the study.

Coding Process

We developed a set of interaction types including descriptions of what would constitute each type of interaction. We then used three videos to refine the descriptions of the interactions, develop key examples, and add categories of interactions that were not previously anticipated.

The final coding protocol was applied to the 12 videos not used for the development of the coding protocol. It is important that the results of a video coding protocol are not dependent upon the individuals coding, so that the coding protocol can be used by other researchers to reach similar results. Two researchers independently coded one video from each instructor and created a master code to resolve any disagreements. The independent codes of the two researchers were over 80% reliable for each interaction type in each of 3 videos. The remaining 9 videos were coded by one of the researchers.

CCI Analysis

Traditionally, concept inventory scores are analyzed using a measure called the normalized gain, which is the fraction of gain achieved out of the total that could be obtained, defined as:

$$\langle g \rangle = \frac{Posttest\ Score - Pretest\ Score}{Maximum\ Possible\ Score - Pretest\ Score} \quad . \qquad (1)$$

This score is typically defined at the classroom level, though the effect of computing these scores at the student level has been addressed as well (Bao, 2006; Coletta & Phillips, 2005). Bao found that the differences could largely be attributed to differences between classes where all students gained uniformly and those where the rank order of students changed. This change in rank order might occur in situations where an instructional style is particularly effective for a subset of the population, like students with initially lower ability. We considered the effect of using both student-level and instructor-level normalized gain scores.

Results

Final Coding Protocol

Videos were coded by classifying each interaction. Only interactions around non-routine problems were considered admissible. For the purposes of this study, we considered routine problems to be those that were completely procedural; they required no interpretation and were algorithmic in nature, such as finding the derivatives of a list of functions. In the classrooms observed, wholly procedural problems were uncommon. Most problems included a real-world context or were building towards a discussion of underlying concepts. For example, all related rates problems observed were considered non-routine because they included an interpretation, such as determining how to model the problem or how to interpret a solution in real-world terms. A problem involving a conical sand pile might include a conversation about the shape of a sand pile, or the interpretation of the sign of the rate of change of the radius with respect to time.

The scope of an interaction was determined by the framing of the question or comment which initiates the interaction. For example, an instructor might ask "what is the value of x in this problem?". In this case, the question marks the beginning of the interaction, and the end of the interaction occurs when the value of x is determined. If the instructor instead asked "how would we set this problem up?", the interaction would be considered to conclude when the setup for the problem has been addressed. Though not frequent, this allows for a single interaction to include multiple exchanges and/or multiple students.

All interactions were categorized as either public or private. Private and public interactions may contribute to student gains in different ways, and the literature does not currently distinguish between these types of interactions in an IE classroom. By dividing interactions in this way, we can investigate whether public or private interactions encourage greater gains. Then, both private and public interactions were categorized by the initiator of the interaction. The initiator of an interaction is the person who introduces the content of a conversation. In public conversations, this was very clear, as the instructor typically initiates interactions unless a student specifically asks a question or proposes an idea. The only category of interactions that was not initiator dependent was called Developing Concepts. These episodes consisted of a sustained discussion on the conceptual content on a topic. For example, an instructor might develop the idea of L'Hopital's rule by appealing to notions of derivatives and rate of change to motivate the statement of the rule, or a student might ask whether L'Hopital's rule has anything to do with rates of change.

Public Interactions

In order to be coded as a public interaction, an interaction must be visible and audible to the majority of the class, the content of the interaction must be calculus-based, and must access students' knowledge, not students' perception of their knowledge. Accessing students' perceptions occurred frequently when an instructor asked "does that make sense?". An answer to this question does not provide the instructor with any information about students' understanding, only whether they think they understand. Similarly, choral response questions, where the answer was clear from the instructor's question, almost never provided substantial information to an instructor and were inadmissible. These questions typically only assessed student perception of understanding, and never provided opportunities for discussion to continue. If a choral response question did lead to a substantial conversation, this conversation was eligible to be counted as an interaction.

Public interactions typically took place when a student asked a question or made a suggestion during class by raising their hand. If the student made a suggestion that extended the conversation beyond the scope of the current conversation, this was considered a new, student-initiated interaction, as opposed to a continuation of the occurring interaction. Student-initiated interactions can include incomplete attempts, such as an incompletely formed question or suggestion. For student-initiated interactions, a student attempting to contribute to the discussion was the key factor in identifying the student as the initiator.

Public interactions were further divided by the initiator, and then by type of interaction. The student-initiated interactions consisted of developing strategies, sensemaking, and checking for correctness (see Table 1 for descriptions and examples). These types of interactions were derived from the descriptions of IE classrooms given by Hake (1998) and Epstein (2007).

Table 1: Public Student-Initiated Interaction Categories

Category Name	Description	Examples
Developing strategies	A student suggests or asks a question about how to solve a problem. This may be a suggestion or question specific to the problem at hand or about a class of problems.	Suggesting a new step in a problem, or asking whether a different solution path would be successful.
Sensemaking	A student makes a comment or raises a question about interpreting content in the course.	Interpreting answers, units, magnitudes, or signs of answers in the work being discussed.
Checking for correctness	A student makes a comment which corrects or asks about the correctness of a solution or step in a solution process.	A student asks why a particular step in a process was justified, or points out a mistake.

Instructor-initiated interactions were those in which the instructor specifically asked a question or began an interaction where the instructor determined the topic of the conversation. These interactions were divided into the categories: promotes sensemaking, promotes checks / connections to previous material / extensions beyond current material, encourages revisions from students, check procedures for sense-making, and presentation of problems worked on by students (see Table 2 for descriptions and examples).

Table 2: Public Instructor-Initiated Interaction Categories

Category Name	Description	Examples
Promotes sensemaking	Making a suggestion about how to think about a problem or type of problem.	Drawing attention to notation, such as noting where a parameter is being used in a new way.
Promotes checks / connections to previous material / extensions beyond current material	The instructor extends the discussion outside of the immediate context.	Connecting the immediate material to material that has already been covered or will be covered in the future, referencing a prior problem or prior comment made by a student.
Encourages revisions from students	Explicitly suggesting a revision from the students in the class.	A revision of work the instructor has written himself/herself, or a suggestion to improve upon or work presented by a student.
Check procedures for sense-making	Checking whether the steps of a specific solution process make sense.	Asking why a particular step was done as opposed to a different step, or asking what justifies a particular step of a solution.
Presentation of problems worked on by students	Instructor provides direct and immediate feedback to students immediately after work is completed.	Instructor presents the solution to a problem on the board after students had worked on the problem either individually or in groups, and had completed work on the problem.

Private Interactions and Work Times

Private interactions occurred whenever students were working with each other or discussed content with an instructor when the majority of the class could not hear or see the exchange. When private work time occurred, the number of interactions was counted and the total amount of time students spent actively working was recorded. This time was counted separately depending on whether the private work was groupwork or individual work because groupwork allows students to provide each other with immediate feedback and individual work time provides students opportunities to engage with content. The amount of time devoted to groupwork varied greatly among the five instructors, and has the potential to be another characteristic of an IE classroom. The amount of time in private work was only considered if the private work lasted at least two minutes. Shorter interactions did not allow students to engage with each other sufficiently, or the questions beginning the private work were not of sufficient difficulty to encourage in-depth, conceptual conversations.

In addition to time being provided for groupwork, many of the private work times also included instructor-student interactions as the instructor circulated the room. The number of these interactions was recorded, then further categorized by who initiated the interaction. Instructor-initiated interactions were those in which the instructor asked a specific question of a student, instead of a question that invited conversation but did not initiate discussion of the content.

Miscellaneous (Uncategorized) Interaction Count

A final category was created to capture the interactions which did not fall into any of the other predefined categories. These included interactions where the topic was precalculus material or may not have qualified as any other particular type of interaction.

Results of Coding

The counts of each type of interaction are given in Table 3 along with the associated normalized gain scores. Among these five instructors, the coding protocol distinguished instructional activities in meaningful ways, both by indicating which instructors were more interactive and by quantifying differences between interactive instructors.

Table 3: Counts of Types of Interactions by Instructor

	Instructor	A	B	C	D	E
	Developing concepts	3	1	0	0	0
Student work-time, including private interactions	Groupwork time (seconds)	1872	1249	0	865	0
	Individual work time (seconds)	0	504	0	2939	0
	Total work time (seconds)	1872	1753	0	3804	0
	Instructor initiated private interaction	11	18	0	0	0
	Student-initiated private interactions	16	14	0	38	0
Instructor-initiated public interactions	Promotes checks	6	7	6	5	2
	Encourages revisions from students	4	8	0	5	6
	Promotes sense making	9	5	1	3	3
	Feedback on questions answered by students	22	9	14	2	18
	Problem presented which students have worked on	1	0	4	8	1
Student-initiated public interactions	Student initiated developing strategies	4	2	1	4	4
	Student initiated sensemaking	2	0	3	8	10
	Student initiated check correct	4	2	4	4	10
	Misc. (Uncategorized) interaction count	52	85	40	51	58
	Normalized gain	0.239	0.271	0.190	0.246	0.259

CCI Gains

There were 26 sections of the course, with a maximum capacity of 35 in each section. Most classes were near capacity, and on average 18.5 students per section participated in the study, with a range of 10 to 26 participating students. The classrooms of the 5 instructors who agreed to participate represented a spectrum of normalized gain scores on the CCI ranging from 0.19 to 0.27, near the national average. The mean normalized gain for the entire participant group at the large, southwestern university where our study was conducted was 0.25, meaning that 25% of the previously unknown concepts was learned during the course. Normalized gain scores for the entire 26 sections ranged from 0.14 to 0.36.

Correlations Between Counts and Gains

The total number of classroom interactions (including those considered "Miscellaneous") was significantly related to student gains, as demonstrated in Figure 1 and reported as Model 1 in

Table 4. Among the specific categories, “Encourages Revisions from Students” was significantly related to student gains, shown in Figure 2 and reported as Model 2 in Table 4.

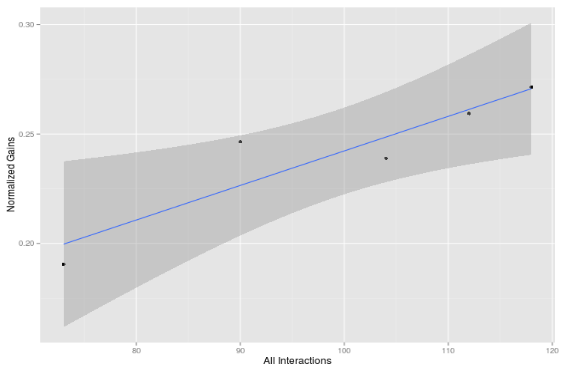

Figure 1: Normalized gains versus all interactions

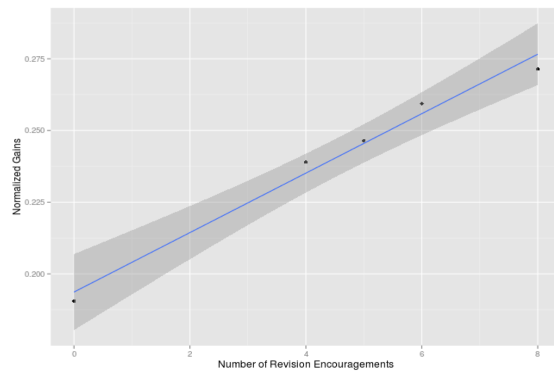

Figure 2: Normalized gains versus number of revisions encouraged

Table 4: Regression Results for Exploratory Analysis

Variable	B	SE(B)	β	*t* (df = 3)	Sig(*p*)	R^2
Model 1: All Interactions						
Constant	0.084	0.039		2.181	0.117	0.849
Interactions	0.002	0.0003	0.922	4.108	0.026	
Model 2: Number of Revisions Encouraged						
Constant	0.194	0.004		46.49	< 0.001	0.983
Interactions	0.010	0.001	0.992	13.22	< 0.001	

Note: B indicates the unstandardized regression coefficient. β indicates the standardized regression coefficient.

Student-Level Analysis of Student Scores

While the previous results were conducted at the classroom level, one can use a statistical technique called Hierarchical Linear Modeling, also known as multi-level modeling, to analyze scores at the student level (Gelman & Hill, 2007). Using this technique, we can analyze student-level gain scores and the results of the video coding protocol at the instructor-level. We calculated normalized gain scores for students using the same formula as was used for classrooms, and constructed a model called a null model which partitions the variance between the student-level and instructor-level. We found that over 99.9% of the variance lies at the student-level, suggesting that, at the university where the study was conducted, nearly all the variation in student-level normalized gain scores can be attributed to differences between students rather than differences between instructors. This suggests that university-level factors, such as department culture, or choice of textbook, may be affecting student gains, and future studies including video analysis of classrooms from multiple universities may provide further insight as to whether this protocol can help understand the relationship between IE instruction and gains in conceptual knowledge. The discrepancy between the instructor-level analysis and the student-level analysis suggests that this relationship is perhaps more complicated than previously thought, and warrants further investigation.

Conclusions

The video coding protocol developed in this study provides a means for analyzing additional classrooms to further investigate the connections between IE instruction and gains in conceptual learning as measured on a concept inventory. When analyzed at the classroom-level, as is traditionally done, our data indicated that despite the small sample size, this coding protocol may describe IE behaviors which are tied to gains on the CCI. While the analysis of the Hierarchical Linear Model suggests that care needs to be taken in interpreting the results at the classroom level, the qualitative differences between classrooms demonstrated by the use of the protocol suggest that this tool can be useful in further investigations of IE instruction.

References

Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, *74*(10), 917. doi:10.1119/1.2213632

Coletta, V. P., & Phillips, J. a. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*, *73*(12), 1172. doi:10.1119/1.2117109

Epstein, J. (2007). Development and validation of the Calculus Concept Inventory. In *Proceedings of the Ninth International Conference on Mathematics Education in a Global Community* (pp. 165–170).

Epstein, J. (2013). The Calculus Concept Inventory - Measurement of the effect of teaching methodology in mathematics. *Notices of the AMS*, *60*(8), 1018–1026.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. doi:10.1119/1.18809

Halloun, I. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, *53*(11), 1043. doi:10.1119/1.14030

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158. doi:10.1119/1.2343497

Hughes Hallett, D. (2006). What have we learned from calculus reform? The road to conceptual understanding. *MAA NOTES*, *69*(July), 43.

Libarkin, J. (2008). Concept inventories in higher education science. In *National Research Council Promising Practices in Undergraduate STEM Education Workshop 2*. Washington, D. C.

Malone, K. (2008). Correlations among knowledge structures, force concept inventory, and problem-solving behaviors. *Physical Review Special Topics - Physics Education Research*, *4*(2), 1–15. doi:10.1103/PhysRevSTPER.4.020107

Prather, E., Rudolph, A. L., Brissenden, G., & Schlingman, W. M. (2009). A national study assessing the teaching and learning of introductory astronomy. Part I. The effect of interactive instruction. *American Journal of Physics*, *77*(4), 320. doi:10.1119/1.3065023

Rhea, K. (n.d.). The Calculus Concept Inventory at a large research university. *Unpublished manuscript*.

Rhoads, T. R., & Roedel, R. J. (1999). The Wave Concept Inventory - A cognitive instrument based on Bloom's Taxonomy. In *Proceedings of the 29th ASEE/IEEE Frontiers in Education Conference* (Vol. 3, pp. 14–18). IEEE.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, *18*(1), 253–286. doi:10.1177/0895904803260042

Thompson, A. G., Philipp, R. A., Thompson, T. W., & Boyd, B. A. (1994). Calculational and conceptual orientations in teaching mathematics. In A. Coxford (Ed.), *1994 Yearbook of the NCTM* (pp. 79–92). Reston, VA: NCTM.