Observing a Culture of Student Independence: The COSI protocol

Carlos Castillo-Garsow
Eastern Washington University

Jeff Crawford
Mead School District 354

Raymond Dempsey
Eastern Washington University

Becky Sommers
Eastern Washington University

Our cohort

Carlos Castillo-Garsow works as an assistant professor at Eastern Washington University, specializing in secondary mathematics education. He has also taught high school algebra and precalculus at schools in Arizona and Washington. Carlos primarily works in the area of conceptual semantics, studying how students, teachers, and mathematicians use mathematics as a language to convey their ideas, and what their choices of words tell us about the mathematics that they are imagining. He focuses on mathematical modeling and the development of a mathematical perspective on non-mathematical problems.

Jeff Crawford works with secondary mathematics teachers in the Mead School District. As a curriculum and instructional specialist, Jeff supports teachers and administration with implementing best practices, curriculum, and instructional tools that reflect current research. Having taught for over 20 years at university and high school levels, Jeff combined his experience in teaching mathematics with his passion for quality materials in endeavors that led him to work with groups throughout Washington State and the nation. His current work with EdReports, Student Achievement Partners, and Illustrative Mathematics inspires him to continue growing personally and increase the effect of his impact on the mathematical community. Jeff hopes that his work will further the awakening of mathematics in our schools and communities.

Raymond Dempsey has six years experience as a university developmental mathematics lecturer and currently teaches at Eastern Washington University. As a graduate teaching assistant, he wrote a master's thesis entitled "Determining the Alignment of Math 105 - Intermediate Algebra at the University of Wisconsin-Milwaukee to the Goals of the Common Core State Standards." He also creates supplemental learning materials for students on YouTube.

Becky Sommers is currently a mathematics lecturer at Eastern Washington University and was previously a mathematics teacher at Deer Park High School. Becky’s classroom experience with both college and high school math students has lent itself well to this project and the investigation of how to improve mathematics education.

CCSS

The Standards for Mathematical Practice expect that educators at all levels bring about attributes in all students around important “processes and proficiencies.” Notably at the heart of the practice standards is student ownership.

MP.1 states “Mathematically proficient students start by explaining to themselves the meaning of a problem and looking for entry points to its solution. They analyze... They make conjectures... They consider analogous problems... They monitor and evaluate…” MP.6 expects “Mathematically proficient students consider the available tools when solving a mathematical problem…are sufficiently familiar with tools appropriate for their grade…make sound decisions about when each of these tools might be helpful…”

This type of student ownership surfaced as a central theme in problems of practice found by most cohorts in this project. Our cohort chose to develop the Culture of Student Independence (COSI) observation protocol to measure student and teacher actions.

Observing a Culture of Student Independence: The COSI protocol

The transition from high school to college can be difficult for students. Students have more independence and more independence is expected from them. In particular, there is much less face time in college classes, and students are expected to make sense of material on their own, outside of class. During the initial brainstorming sessions for this Successful Transitions project, all of the workgroups expressed a concern for how students were developing as independent learners. These working groups then developed interventions designed to either foster skills or attitudes that would in some way help students progress toward more independence.

To measure independence, our cohort began work on an observation protocol. We were curious to identify how much students rely on the teacher as a source of knowledge in the classroom, and how much the teacher promotes a culture independence vs. a culture in which all knowledge comes from them. In particular, we wondered, when students are stuck on a problem, do they stop work and wait for the teacher to help them, or do they turn to other resources, such as classmates, notes, or calculator?

In response to this question, our cohort developed an observation protocol to measure how often student used resources other than the teacher in learning, both with and without a prompt from the teacher. We also measured the teacher's contribution to the classroom culture of independence: How often the teacher encouraged students to use other resources, vs. how often the teacher acted as the source of knowledge. Unlike more general observation protocols designed to evaluate a teacher as a whole, the protocol proposed in this paper is narrowly focused to a singular question of classroom culture: How to measure students using resources other than the teacher (with or without a prompt)?

The tool is designed to be used by instructors who are interested in measuring the independence of students' in their classrooms, or to evaluate the effectiveness of instruction designed to foster student independence. Because the Successful Transitions project involves both Mathematics and English/Language Arts instructors, the design of the observation protocol is intentionally content agnostic, focusing only on student independence.

The basic structure of the protocol involves an external observer attending class and recording observed behaviors in an iPad app or on a paper form. The exact categories of behaviors to be observed have changed over several design iterations of the observation protocol. In this paper, we describe the design history of the tool, why changes were made, and the current form of the tool.

Literature Review

Because this observation protocol is designed to observe classroom culture, the primary theoretical lenses through which we viewed the classes we observed is sociocultural. These perspectives provide tools for observing and interpreting how students use external resources such as classmates and calculators.

Vygotsky (1978) provides the idea of the Zone of Proximal Development (ZPD), which is often used to justify collaborative approaches in classrooms. The theory is that a student can solve a larger set of problems in collaboration than they can independently. Vygotsky (in translation) defines the Zone of Proximal

Development as "The distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86).

Vygotsky had a much larger learning theory about how work in the ZPD eventually becomes internalized so that the student can work a larger set of problems independently, but the larger theory of internalization is not relevant to this project. Instead, if we take as fact that students can solve a larger set of problems in collaboration than independently, then the ZPD defines our area of study. When a student working independently reaches a point in a problem that they cannot solve on their own, they must turn to another for help. The purpose of the COSI observation protocol is to record what the student chooses to do in these circumstances: Do they turn to a peer or to the teacher?

A classmate might not be the only resource available. The theory of distributed cognition or distributed intelligence (Pea, 1997) extends the ZPD to include interactions with artifacts in addition to interactions with other people. Artifacts in distributed intelligence is quite broadly defined to include tools such as calculators, but also symbolic representations such as pictures and text. Choice of artifact can greatly impact the types of problems that students can solve. The observation protocol both tracks and distinguishes between different sources of help: the teacher, another student, or an object (artifact) such as a calculator or notes.

Design of observation protocols

In our review of literature, observation protocols fell into two primary categories. Observation protocols in mathematics education tend to observe the class as a whole, and rate the teacher and or class in various dimensions using a Likert scale (Boston et al., 2015; Judson, 2013). Observation protocols in science education tend toward a different format of counting observable student or teacher behaviors and tracking those behaviors in time (Erdogan et al., 2010; Smith et al., 2015). Forbes et al. (2013) took a hybrid approach of rating observable student behaviors on a Likert scale.

For the development of our protocol, we decided to choose a more science education based approach of recording and tracking observable student and teacher behaviors in time. This is due in part to the narrow focus of the protocol question. Having a narrow focus allows us to track only a few behaviors, and tracking these behaviors in time gives better resolution into the particular structure of a class.

Depth of knowledge

When measuring for depth of knowledge (DOK), the cohort employed the Cognitive Rigor Matrix (Hess, 2009). In brief, the DOK levels can be described as follows:

Level 1	Recall and Reproduction	Recall of a fact, term, principle, concept; perform a routine procedure; locate details
Level 2	Skills and Concepts	Use of information; conceptual knowledge; select appropriate procedures for a given task; two or more steps with decision points along the way; routine problems; organize/display data; interpret/use simple graphs; summarize; identify main idea; explain relationships; make predictions

Level 3	Short-term Strategic Thinking	Requires reasoning, or developing a plan or sequence of steps to approach problem; requires decision making or justification; abstract, complex, or non-routine; often more than one possible answer; support solutions or judgments with text evidence
Level 4	Extended Thinking	An investigation or application to real world; requires time to research, problem solve, and process multiple conditions of the problem or task; non-routine manipulations; synthesize information across disciplines/content areas/multiple sources
Adapted from the Cognitive Rigor Matrix (Hess, 2009)		

Methods

The observation protocol has been developed using a combination of design research (Cobb et al. 2003) and grounded theory (Strauss and Corbin, 1998). Because an observation protocol involves assigning categories and numerical values to qualitative data, grounded theory seemed to be a natural choice for developing the categories to be observed. Grounded Theory distinguishes between three types of coding. In open coding, the raw qualitative data is analyzed and organized into natural categories (the open codes) that the researchers use to describe and organize the concepts they observe in the data. In axial coding, the codes themselves are the object of study. Axial coding involves organizing the open codes into a structure, and identifying categories of open codes. Selective (or closed) coding occurs at the point where no new categories are found by the open and axial coding process. At this point, the codes are finalized, a theory is constructed, and closed coding is used primarily to convert qualitative data to quantitative data. Closed coding is a natural fit for an observation protocol that relies on recording observed behaviors in time, so we followed the full procedure for developing open, axial, and closed codes as a way of constructing the protocol.

A design experiment (Cobb et al. 2003) is primarily an approach to theory building, by using cycles of invention, testing, revision in order to refine a theory. In the core design experiment methodology, a theory of learning or teaching suggests a particular intervention. This intervention is then implemented and tested. Unexpected results are used to refine the theory and design new cycles of intervention and testing until some sort of equilibrium is reached. The final product would be an improved theory. We did not use a full design research methodology, because the development of a protocol lacks an educational intervention, and because the final product is an assessment tool, rather than a theory. However the protocol has been developed using a design cycle of invention, testing, and refinement. In keeping with design research methodology, we value each cycle for the contribution to our greater understanding of the problem, and report on all of the cycles in the results.

The First Design Cycle

Initial design of the observation protocol began by watching videos of high school mathematics classrooms that had high amount of in-class student work. As we watched the video, we began to record student behaviors that we found relevant to the question of measuring students' use of resources to resolve a problem. This list was then supplemented with behaviors that we had observed in our own experience. This list of behaviors became our open codes:

Teacher identifies resources
Teacher provides instruction on use of resources
Teacher does not create opportunities for other resources (lecture/quiz/etc)
Teacher answers question that could have be redirected to class/student
Student asks for more student-student interaction
Student asks for more work time
Student requests information on resources
Student requests instruction on use of resources
Student asks question of teacher
Teacher creates opportunities for students to ask questions of each other
Teacher asks open-ended question
Students ask questions of each other
Students share observations with other students
Students observe and learn from each others' approaches to problems
Student builds on another student's comment
Student challenges another student's comment
Student expresses uncertainty to another student
Student waits for teacher to come along to solve the problem
Internet access restricted
Student passively uses assigned resource (read text/watch video)
Internet access made available
Teacher provides students with control of reference (textbook/video/etc).
Teacher assigns creation of reference (journal/wiki/etc).
Student reviews textbook
Student reviews video tutorial
Student flips back in notes/handout
Student reviews student created reference
Students create connections in student created reference
Student uses website
Student suggests or brings in resource
Student asks metacognition/reflection questions
Student connects problem to out of classroom experience.
Student demonstrates intrinsic motivation to learn material
Student creates a sub-problem for themselves

OBSERVING STUDENTS INDEPENDENCE

- Student reorganizes or re-represents the problem (ex: draws a diagram)
- Student categorizes problem by underlying mathematical principle
- Student makes their own observations beyond what is planned by the teacher
- Student makes their own problems beyond what is planned by the teacher
- Student proposes an experiment

Initial axial coding and first trial coding

Our initial organization of the data involved classifying the open codes into major categories based on who the actors were, and sub-categories based on the type of interaction:

- Student-Teacher
  - Teacher identifies resources
  - Student requests instruction
- Student-Student
  - Teacher opens collaboration
  - Students share observations
  - Students ask questions of each other
  - Students observe each other
  - Students (build on/challenge) each other's comments
  - Student expresses uncertainty to another
- Student-Reference
  - Student uses external reference
- Student-Internal

This initial coding scheme was used to code classroom video from the Trends in International Mathematics and Science Study (TIMSS) website. The videos were coded using a form designed in the iCoda iPad app. iCoda is software tool for field observation in which a form consists of “buttons.” Whenever an event corresponding to a code is observed to start, a button is pressed to activate it, and then pressed again to deactivate the button. iCode records the start time, end time, and title for all buttons pressed during an observation. Early versions of our form no longer exist, but a sample later version can be seen in Figure 6. The resulting data was studied using analysis software written in-house by the authors.

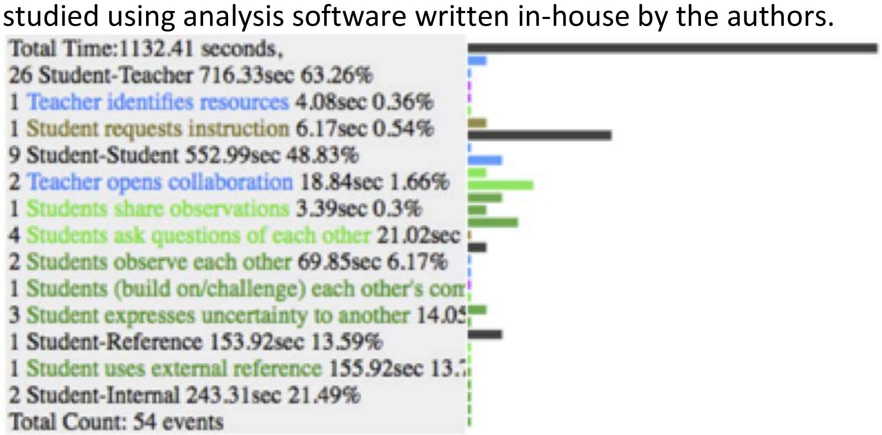

Figure 1: Example code histogram of code frequencies (counts)

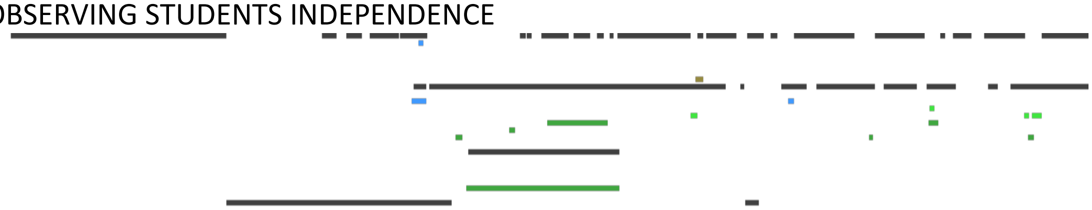

Figure 2: Sample timeline showing the time of occurrence of each code during the class video. Vertical axis is code (in the same order as the histogram in figure 1). Horizontal axis is time.

Second axial coding and first trial coding

During this design cycle and discussion, a second way of organizing the codes emerged, which led to a second axial coding of the codes. We decided that scoring a class based on student actions wasn't reasonable in situations where the instructor didn't give the students an opportunity to act. For example, in the timeline above (Figure 1), you can see a long period of time at the beginning of the class when the teacher is introducing the problem (The leftmost black bar representing "student-teacher" interaction).

In this second organizing scheme, we placed the observed behaviors in two dimensions. The dimension was our original top-level category of the actors, while the second dimension was based what kinds of opportunities the students had to use resources, and whether or not they took those opportunities. The new axial coding scheme looked something like (Figure 3).

	Student-Teacher	Student-Student	Student-Reference	Student-Internal
Opportunity Created	1	2		1
Not Created				
Initiative Taken		5		
Opportunity Taken	4	6	2	
Not Taken	1	1		

Figure 3: Table of results for two-dimensional axial coding

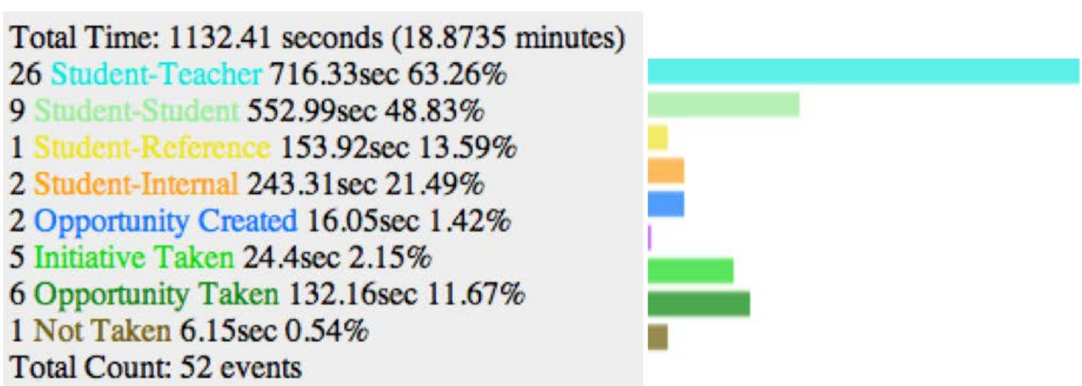

Figure 4: Count histogram under two-dimensional axial coding scheme

Figure 5: Example timeline under two-dimensional axial coding scheme.

The Second Design Cycle

After the initial design cycle, we found that there was a dissonance between focusing on pairs of actors in one dimension, and opportunities in the second dimension because opportunities involved only a single actor. We redesigned the codes to combine these two categories into a single resources dimension.

Teacher is resource
Teacher points to resource
Student uses teacher
Student uses student
Student uses object

However, we still wanted to make a commentary on quality of content, so we added a new second dimension based on Depth of Knowledge (DoK) levels in an attempt to track when students were relying on the teacher for simple recall situations or when their need required more analysis or deeper thinking. The resulting iCoda form can be seen in Figure 6.

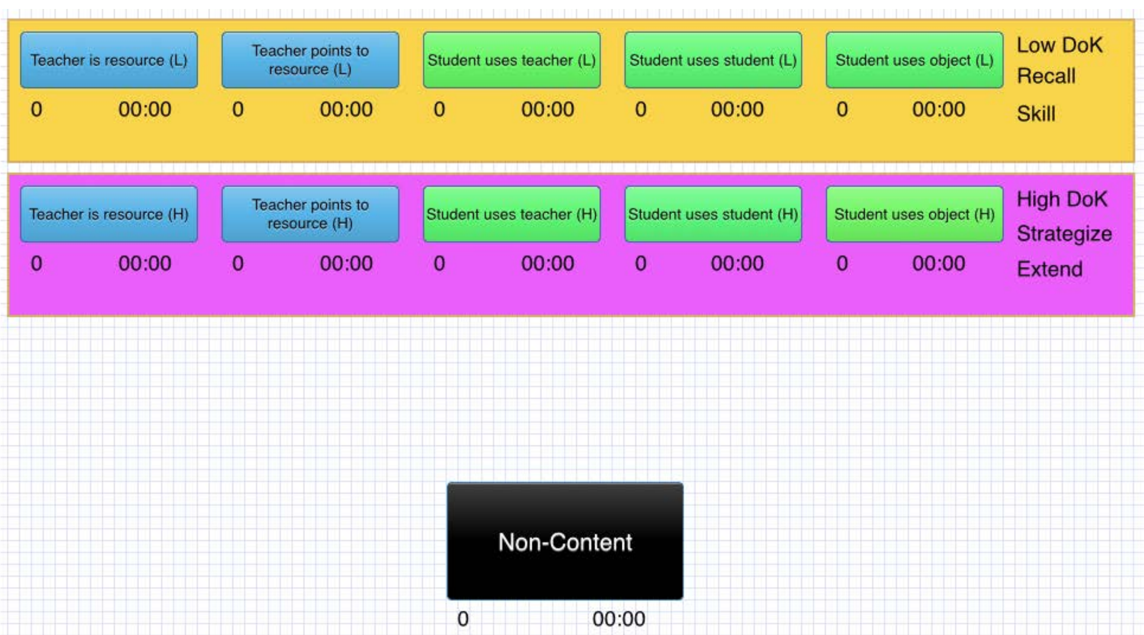

Figure 6: iCoda form for cycle 2. Each button corresponds to a code. The observer presses a button to activate the code, and again to release the button, and iCoda records start and end times.

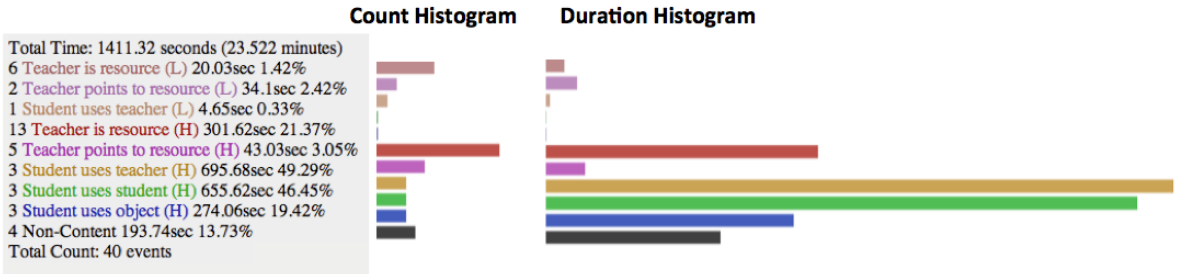

Figure 7. Comparison of count and duration histograms for the same class, showing very different pictures of the classroom.

We began to observe a large number of overlapping event instances; this meant that our initial approach of counting code instances was not viable. Because each button only had an on-off state, overlapping instances would be recorded only as a single event. For example students asking for help often overlapped with each other, leading to a misleadingly low count of student's asking for help. We compensated for this by making a move to tracking the duration of a code rather than counting instances (Figure 7). This showed us a very different picture of the classroom, where the students spend a lot of time waiting for the teacher to come help them.

A second problem was the interface of the redesigned form. Decisions identifying actors could be made very quickly, (as soon as the actor began speaking), while decisions about depth of knowledge required more insight into the content. This led to a time lag where the observer would often know which code to choose in the resources dimension long before they knew which DoK level was appropriate to code for. After this iteration and the challenges presented by the additional coding of DoK, the decision was made to remove this feature from the coding.

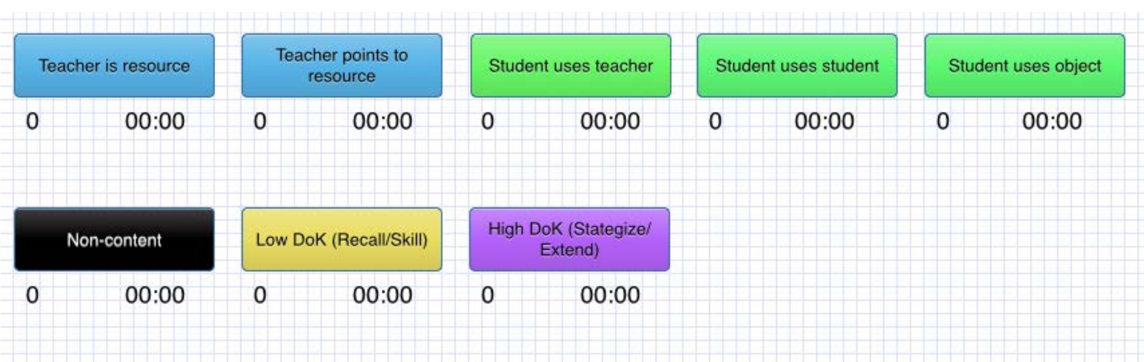

Figure 8: Redesigned iCoda form separating the two dimensions to account for time lag.

The Third Design Cycle

At this point, our codes became relatively fixed. The third design cycle was primarily about negotiating the meanings of individual codes. Due to the limited number of TIMSS videos, we moved to coding classroom video from two high school algebra classes in Arizona.

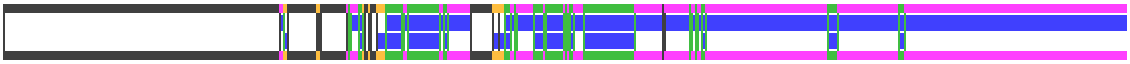

Figure 9: Sample inter-rater reliability analysis for code “Student uses Student.” Blue bars are the codes by the two raters. Black represents times where raters agree there is no code. Green represents times when raters agree there is a code. Orange and magenta represent times of disagreement when only one rater coded. This particular example shows 49.082% agreement between coders.

Initial inter-rater reliability tests with the form had low ratings (Figure 9). We returned to the video to explore times when the coders disagreed and began negotiating the meanings of codes.

Teacher is resource:

The primary disagreement here was whether teacher is resource was only lecture, or whether a teacher responding to a question from a student counted as teacher is resource. Essentially the disagreement was about whether teacher as resource must always be teacher initiated. We decided for the later option: That any time a teacher acted as a source of information, the teacher was acting as a resource, regardless of who initiated it.

Teacher points to resource:

The primary disagreement here was whether assigning an activity that required a tool (such as a calculator) counted as teacher points to resource, because students were expected to use (and be introduced to) the tool. Unlike the case above, we decided that teacher points to resource needs to be defined as the teacher suggesting the use of a tool that was not part of a lesson. For example, if a lesson is set up in such as way that the student is expected to use a graphing calculator as a means to solving a problem, we would not code this as “teacher points to resource.” We would, however, code the use of a graphing calculator as “teacher points to resource” if the calculator was not part of the lesson plan.

Fourth Design Cycle

In our third trial, we tested both the electronic form and a new paper form (Figure 10). The paper form was used as a checklist. During observations, the observer kept a timer, and every minute would check the boxes corresponding to any code events during that minute. The paper form also had a section for the observer to write notes on their observations.

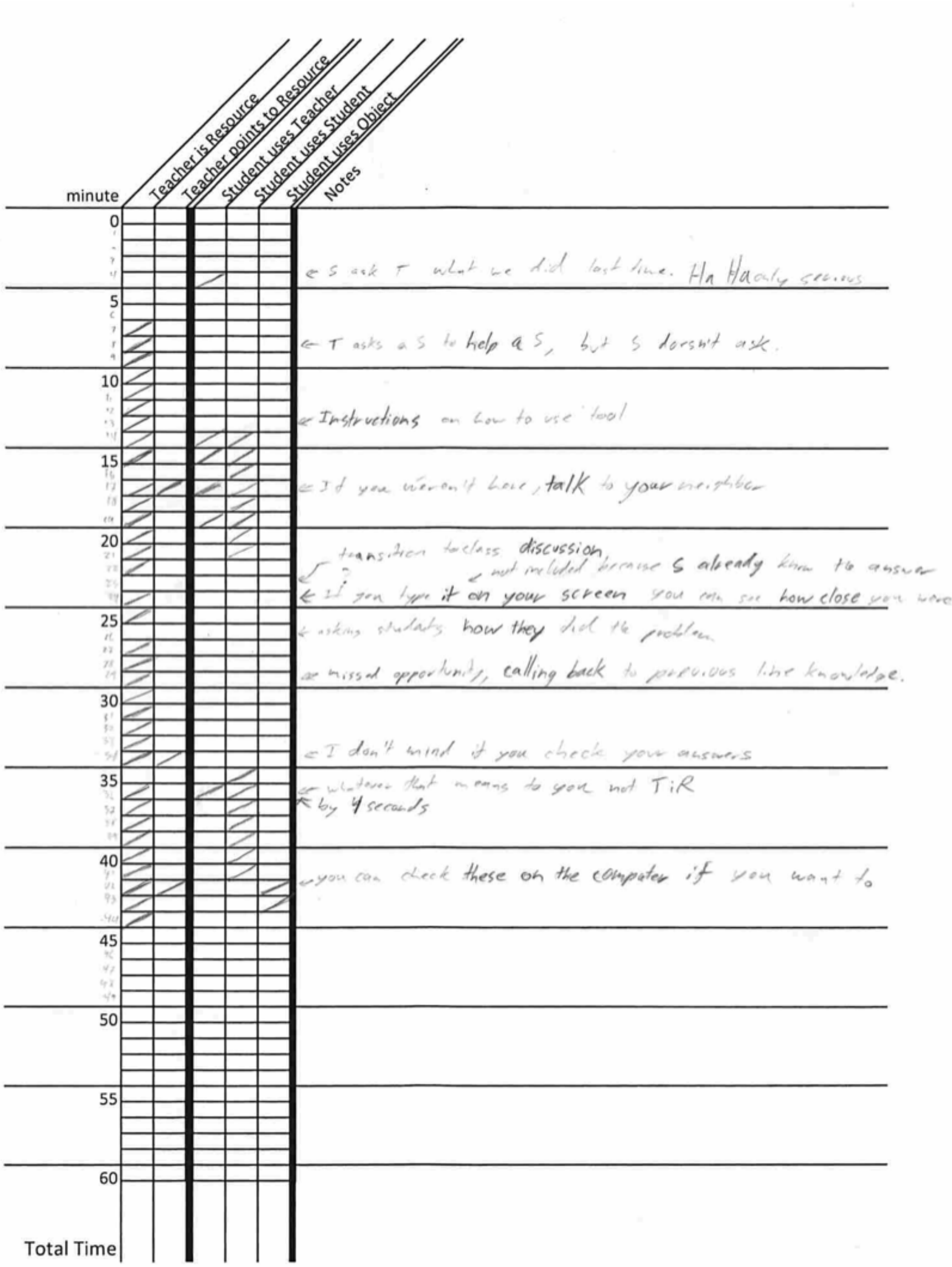

Figure 10. Sample paper form, showing timeline and accompanying notes.

The team universally preferred the paper form for three reasons. First, coding minute by minute instead of second by second gave the team more time to think and analyze while coding. Second, the paper form allowed for the creation of a timeline without specialized software; and third, the team liked the opportunity to add notes to the timelines. These notes served as both explanations for an observer's' particular choice of code, and reminders to the observer of what happened for future discussion. The paper form became the sole tool used to gather date during the final inter-rater reliability run.

Definitions of Code

Throughout the development of the COSI observation protocol, the cohort regularly revisited the meanings of each code. Discussions and adaptations were made at each iteration. The efficacy of each code definition was tested with sample videos of classrooms. During the last design cycle, the following definitions were decided.

Code	Meaning	Clarifiers
Teacher is resource	A teacher provides information (student initiated or not) in order to assist students in completing a mathematical problem, task or investigation.	Examples of this code include: ● teacher lectures ● teacher explains how to solve a problem ● teacher says "On #14 you will want to…" ● teacher using socratic dialogue to guide students through a process
Teacher points to resource	A teacher suggests resources other than an instructional authority in the room in order to further the completion of a mathematical problem, task or investigation.	Teachers assigning a resource broadly to an entire group is different than teachers offering choice through suggestion or personalized advice. Examples of this code include: ● teacher points student to notes, books, internet, etc in response to or in anticipation of student request ● teacher says "Where do you think you might find that out?" ● teacher says "If you get stuck, you might think of asking your partner." ● teachers says "You can find the answer to that in your notes." Non-examples of this code include: ● teacher assigns a specific tool to be used by the whole class ● teacher prompts "Ok class, get your notes out." ● teacher prompts "Groups A & B, you will need your calculator." ● teacher prompts "Alright everyone: Think, Pair, Share."

Student uses teacher	A student takes action to solicit input from a teacher (or another instructional authority) in order to further the completion of a mathematical problem, task or investigation -- whether the teacher provides aid or not.	Examples of this code include: ● student asks clarifying questions about the mathematical problem, task or investigation ● student asks for strategies or procedures ● student waits for teacher to do the problem or give answers
Student uses student	A student takes action to solicit input from a classmate in order to further the completion of a mathematical problem, task or investigation.	Students volunteering information to other students without being prompted would not be an example of this code.
Student uses object	A student takes action to seek information from non-human resources in order to further the completion of a mathematical problem, task or investigation.	Examples of this code include: ● using notes, books, internet to find information ● going to wolframalpha.com to perform mathematical operations ● using tools (e.g. calculators, graph paper, etc.) of their own choosing to explore a problem or work out thoughts ● student follows up (by choice) on a suggestion to use a tool made by a teacher. Non-examples of this code include: ● teacher assigns a specific tool to be used by the whole class

Final Consensus of Reliability

Inter-rater reliability for the fourth design cycle was generally high, although there still appear to be areas where definitions need to be refined. Overall, the observers had an average agreement of 81.39%. Areas of strong agreement where in the “Teacher is Resource” and “Student uses Student” codes. These codes were the most common in all videos and were the most heavily negotiated (Figures 11, 12).

Figure 11. Sample comparison of “Teacher is Resource” for two observers, showing 95.6% agreement.

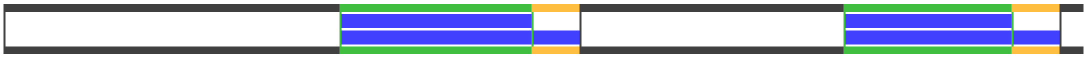

Figure 12. Sample comparison of "Student uses Student" for two observers, showing 94.6% agreement

There were some areas of disagreement in "Student uses Teacher" but the disagreements appear to be durational. Overall, observers agreed on the beginnings of "Student uses Teacher" events. The difference appears to be whether or not the observer coded the teacher's response as "Student uses Teacher" or only the incidence of the student asking the question as "Student uses Teacher" (Figure 13).

Figure 13. Sample comparison of "Student uses Teacher" showing the longer codes due to including the teacher's response. This particular example has 71.1% agreement.

The greatest disagreement between observers was for the rarely occurring codes, "Teacher points to Resource" and "Student uses Object." A sample of the disagreement on "Teacher points to Resource" can be found in Figure 14. Only one observer coded "Student uses Object" in the final run. It appears that these rare codes still need to be better negotiated and defined.

Figure 14. Sample comparison of "Teacher points to Resource" showing low agreement. This particular example shows 57.7% agreement. Counting only the durations coded, there is 5% agreement given the existence of a code.

Conclusions/Recommendations

This tool was created to help an observer better track student behavior and student/teacher interaction in a classroom setting. The tracking of student behavior is needed to address the problem of lack of independence in students. The COSI protocol helps track student behavior by allowing the observer to empirically record how students respond to needing extra help with a problem, or needing more information. With the support of this tool, data and trends regarding classroom culture can be more easily identified and the effects of an intervention can be measured.

COSI was designed to be easy to use in real time with basic instruction on its use and designed with clear definitions so multiple observers should have reasonable consensus on the data recorded. This work is not complete, as some of the definitions still need some refinement, but COSI is developed enough to be viable as place to start a conversation about someone's classroom. It is also adjustable so observers can use it to reflect other ways in which learning takes place in the classroom. Though not designed to be an intervention by itself, it could be used to support interventions focused on the interactions between the instructor and the students.

Comparing the electronic to the paper form lead the counterintuitive conclusion that the paper form was easier to use. The ability to code after the event gave the observers more time to observe, think, evaluate, and adjust their conclusions. This means that the paper form may solve our initial problem with coding DoK levels. The primary problem with coding DoK levels was the time needed to make a decision. Using a paper form to code after the event has completed may make it possible to evaluate of the depth of the content in addition to the students' independence. This is an area to explore as we continue to refine the protocol.

References

Boston, M., Bostic, J., Lesseig, K., and Sherman, M. (2015). A comparison of mathematics classroom observation protocols. Mathematics teacher educator, 3(2):154 – 175.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., and Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32(1):9–13.

Erdogan, I., Campbell, T., and Abd-Hamid, N. H. (2010). The student actions coding sheet (SACS): An instrument for illuminating the shifts toward student-centered science classrooms. International Journal of Science Education, 33(10):1313 – 1336.

Forbes, C. T., Zangori, L., and Biggers, M. (2013). Investigating essential characteristics of scientific practices in elementary science learning environments: The practices of science observation protocol (P-SOP). School Science and Mathematics, 113(4):180 – 190.

Judson, E. (2013). Development of an instrument to assess and deliberate on the integration of mathematics into student-centered science learning. School Science and Mathematics, 113(2):56 – 68.

Hess, Karin K. (2009), A New Lens for Examining Cognitive Rigor in Standards, Curriculum, and Assessments. National Center for the Improvement of Educational Assessment, NCIEA.org.

Pea, R. D. (1997). Practices of distributed intelligence and designs for education. In Salomon, G., editor, Distributed Cognitions, pages 47 – 87. Cambridge University Press.

Smith, M. K., Jones, F. H. M., Gilbert, S. L., and Wieman, C. E. (2015). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. CBE – Life Sciences Education, 12:618 – 627.

Strauss, A. and Corbin, J. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. SAGE Publications, Thousand Oaks, second edition.

Vygotsky, L. (1978). Mind in Society. Harvard University Press, Cambridge, MA.