TAMI-OP

Toolkit for Assessing Mathematics Instruction – Observation Protocol

Charles N. Hayward[1], Sandra L. Laursen[1], and Timothy Weston[2]

[1]Ethnography & Evaluation Research, [2]ATLAS Institute
University of Colorado Boulder

Updated July 26th, 2021

TAMI-OP should be cited as:
Hayward, Charles N., Laursen, Sandra L., & Weston, Timothy. (2017). *TAMI-OP: Toolkit for Assessing Mathematics Instruction – Observation Protocol*. Boulder, CO: University of Colorado Boulder, Ethnography & Evaluation Research

What is TAMI?

TAMI, Toolkit for Assessing Mathematics Instruction, is a suite of tools designed by Charles Hayward, Sandra Laursen, and Timothy Weston at the University of Colorado Boulder. TAMI is for researchers, evaluators, or anybody else who may be interested in characterizing and measuring instructional practices in college mathematics courses. Currently, it includes an observation protocol (TAMI-OP) and instructor survey (TAMI-IS). Work is underway to expand it to include other tools as well. This document explains the observation protocol in detail and answers some common questions about it.

How is TAMI-OP different from existing observation protocols?

Every protocol is designed for a specific purpose. Some are designed just to describe what is happening in a class, while others aim to evaluate the quality of instruction. Some are granular and measure repeatedly in short intervals, while others are based on holistic ratings of the entire class period. The main purpose of TAMI-OP is to describe what practices are being used in mathematics classrooms in 2-minute intervals throughout the class, and not to consider the effectiveness of these practices or the expertise with which they are implemented. However, it also includes some evaluative items, some holistic items, and space for qualitative descriptions.

In each two-minute interval throughout class, both student and instructor behaviors are recorded. Additionally, frequency and types of student and instructor questioning are also coded. TAMI-OP also incorporates the ICAP framework (Chi & Wylie, 2014) as an evaluative component for the effectiveness of those practices. TAMI-OP is adaptable – it allows users to add one custom code of their choosing, which can be used to focus on a specific target activity. For example, users could adapt this feature to measure when instructors provide sufficient wait time after a question, use Think/Pair/Share activities, or employ a participation strategy such as calling on a randomized student. There is also a space to take notes, which can be used to record additional detail not captured in the coding choices. At the conclusion of the interval-based classroom observation, TAMI-OP also includes some holistic end-of-class rating items.

While this may sound like a lot for an observer to accomplish, TAMI-OP is quite easy to use for real-time coding, either in-person or through video recordings. Training can be accomplished in a few hours, and

sufficient inter-rater reliability can usually be achieved in just a handful of practice sessions. Conducting a classroom observation takes only a minute or two longer than the actual class session.

Where did TAMI-OP come from?

TAMI-Observation Protocol draws heavily on some existing descriptive, segmented protocols but was adapted specifically for use in college mathematics classes. TDOP: Teaching Dimensions Observation Protocol (Hora, Oleson, & Ferrare, 2013) is the intellectual 'grandfather' of TAMI-OP. It captures various dimensions of what is happening during a class, measured in 2-minute intervals. Smith, Jones, Gilbert, and Wieman (2013) shortened and modified the TDOP to produce a similar observation protocol specifically for use in undergraduate STEM courses. Their instrument is called the COPUS: Classroom Observation Protocol for Undergraduate STEM. We attempted to use the COPUS in college mathematics classes, but found that mathematics instructors' practices did not quite align with the COPUS codes, perhaps because they were developed for science courses. We adapted the COPUS to better capture the practices we were seeing in college mathematics courses, and tested our items through multiple rounds of interviews and observations with college mathematics instructors. Our work was funded by a grant from the National Science Foundation (*DUE 1245436: Development of a Validated Self-Report Instrument for Measuring the Classroom Impact of Student-Centered Professional Development for College Instructors)*

The main portion of TAMI-OP is descriptive – it simply describes what instructor and student behaviors are present in each two-minute interval and does not evaluate their quality. This was by design, because we were interested in measuring instructors' efforts to incorporate particular instructional practices rather than their skill in doing so. Moreover, evaluative protocols often require extensive training over multiple days (e.g. RTOP; Sawada, et al., 2002). We added the ICAP framework (Chi & Wylie, 2014), a simple categorization of students' cognitive engagement, to TAMI-OP so that we could include an evaluative component that does not require extensive training. Additionally, at the end of the class, TAMI-OP contains 16 holistic, evaluative questions to rate the overall quality of the class. Again, these are simple evaluative measures that do not involve complicated rubrics or extensive training. We developed and used these items in an earlier study (Laursen, Hassi, Kogan, Hunter, & Weston, 2011).

How does the TAMI-OP electronic protocol work?

The TAMI-OP electronic protocol is an Excel-based coding protocol. It relies heavily on the use of macros to accomplish helpful features. It requires a fully featured version of Excel running on a regular computer. Functionality may be limited on tablet-type PCs, like the Microsoft SurfacePro. It will not run on tablets or computers unable to run Excel (iPads and Chromebooks, for example.) Questions about the electronic protocol should be directed to chuck.hayward@colorado.edu

Why are the end-of-class items and RTOP included in TAMI-OP?

Both of these instruments are holistic items. The end-of-class are more descriptive and the RTOP is largely evaluative. While these instruments are not officially part of TAMI-OP, they are included in the template as they can easily be coded at the conclusions of the real-time TAMI-OP coding. This makes it possible to make comparisons across various types of coding protocols used on the same classes.

The Excel-based design overcomes problems of both paper and web-based protocols:

- It is a stand-alone Excel file that doesn't require internet access or any registration.
- There is a built-in timer. Macros visually help to align the coder with the current time interval, but do not force the coder to advance at the end of the two minutes. Coders are free to continue coding or may go back and edit information for a previous interval if needed. There is no loss of data this way or need to edit the data at the end of the session.
- Files are stored locally and automatically using unique identifying information from the observation (course, time, and coder). So, even with multiple coders, the files will not be accidentally overwritten. ***NOTE: Due to changes in Apple's sandboxing processes, this feature is no longer completely automated on Macs. Instead, the user has to manually set a location and perform a few simple actions when prompted by dialog boxes.**
- Programming is written to be responsive to adjustments. For example, if the class runs longer than anticipated, the coder can continue coding - up to a maximum of 190 minutes. The length of class will automatically be adjusted to the completed time on the timer.
- The timer can be started with any time on it for observations that start after class begins.
- Automatic recoding: filled cells are converted to '1' whereas blank cells are converted to '0' in the DatabaseTransfer sheet to prepare for data analysis. Any character in the cell is coded as a 1 so coders are free to code with any value, and double-keyed typos do not result in errors.
- Additional options for coders, sites, etc, can easily be added for various contexts.
- Filled cells are colored blue to help visually interpret the flow of the class.
- It can easily be customized and creates new templates with one extra code for a target behavior or by creating pre-filled cells. These may be helpful in repeated observations of the same course.

What does TAMI-OP measure?

Activities are coded if they are present in each two-minute interval, so the code may indicate the activity happened throughout the full interval or for only a portion of it. Thus, within any given time interval, there may be two seemingly independent codes if activities switch within the 2-minute interval. For example, a sheet like the one below might result if the instructor started class with some quick announcements ("Adm – administrative"), then spent the first 10-11 minutes of class lecturing ("Lec") while writing on the board ("RtW" – real time writing). During this time, the instructor asked 2 informational questions ("QIn"), which students answered ("AnIn"), possibly by just providing answers to calculations. After this lecture, the instructor had students work on a problem individually ("Ind"). Students spent about 2-3 minutes working individually, and then the instructor reviewed the problem ("Rvw") for a couple of minutes before resuming lecturing. During the review, the instructor answered ("AnQ") one student question ("Q"). Looking at coding patterns over time helps to reveal the flow of the class and general trends in how class time is spent.

	1. Students doing														2. Instructor doing												
Start minute	SP	GP	RtW	WG	OG	Ind	Q	AnIn	AnRs	WC	C/V	T/Q	W	O	Rvw	Lec	RtW	MG	1o1	QMd	QIn	QRs	AnQ	Adm	W	O	TA	Bth
0																x	x							x				
2																x	x											
4								2								x	x				2							
6																x	x											
8																x	x											
10						x										x	x											
12						x	1								x								1					
14															x	x	x											

The cells shaded in light green indicate the number of different types of questions & answers exchanged between instructor and students. Q&A are only marked during times of full-class activity. During activities like groupwork, the noise makes it too difficult to reliably code numbers or types of questions and answers occurring across multiple groups. In general, if instructors engage in question and answers with students during group or individual activities, it is coded as "MG" for moving and guiding.

TAMI-OP Code Definitions

1. Students are doing	
SP	Student presenting solution or proof to the class (code until student sits down) • *Continue to code until the student sits down or is no longer "on" (for example, continue coding if students are asking questions to the presenter even if the presentation part has ended.)*
GP	Group presenting solution or proof to the class (code until students sit down) • *Continue to code until the group sits down or is no longer "on" (for example, continue coding if students are asking questions to the presenters even if the presentation part has ended.)*
RtW	Real-time writing on the board, doc projector, etc. • Code while students are presenting to help indicate if it is prepared or in the moment. Do not code while students do individual work or take notes.
WG	Working in groups on structured group work (worksheet, whiteboards, etc.) – activities that were part of instructor's lesson plan
OG	Other group activity (such as a Think/Pair/Share) – activities that are designed in the moment to respond to student difficulties or needs that become apparent during class
Ind	Individual thinking/problem solving. • Mark when instructor explicitly asks students to think about question/problem on own. • Can also be used if instructor gives no explicit directions but the norm or assumption seems to be that students work individually.
Q	Student asks question (use tick marks in box to count the number of questions) • Code when a student asks a question to an instructor or student presenter.
AnIn	Student answers question/makes comment by providing specific information (usually a short answer) • Code when students answer questions from instructors or another student by providing information (commonly a computational result or reciting something they have previously learned.)
AnRs	Student answers question/makes comment by providing reasoning or justification • Code when students provide a reasoning why or how to do something, or explain their thinking. • Do not code justifications that are just recitations of procedures, e.g. Instructor asks "How do we find this?" and student responds "Factor". Those can be coded as AnIn instead.
WC	Engaged in whole class discussion by offering explanations, opinions, judgment, etc. to whole class (often facilitated by instructor) • Code when students respond to each other's comments or questions. Can be moderated *through* the instructor, but students should be responding to each other's ideas, not just back and forth with instructor.
C/V	Students interacting with computers/simulations (code student interaction, even if instructor is operating the computer) • Can code for students *using* manipulatives to help understand mathematical concepts. Do not code if the students are just watching the instructor model them.
T/Q	Taking a test or quiz
W	Waiting (instructor late, working on fixing AV problems, instructor occupied, etc.) • Code when students could be doing something mathematical, but are not.
O	Other • Code for unique situations that don't fit other categories.

2. Instructor is doing

Rvw Instructor reviewing students' thinking or student contributions (presentations, homework, test, in-class, etc.)
- Code only when reviewing student contributions or student thinking. Do not code if instructors are just reviewing a topic they've already covered without attending to student work.

Lec Lecturing (presenting content, deriving mathematical results, presenting a problem solution, etc.)
- Code when instructor is presenting mathematical content to the entire class.

RtW Real-time writing on the board, doc projector, etc. (often checked off along with Lec)
- Code while lecturing/reviewing to help indicate if it is prepared materials (slides) or in the moment.

MG Moving through class guiding on-going student work during active learning tasks
- Code when instructor is interacting with students during active learning by answering or asking questions, providing help, etc.

1o1 1-on-1 extended discussion with one/few individuals, not paying attention to rest of the class
- Can be coded along with MG or AnQ.

QMd Questions intended to moderate or invite student participation (may also be done in the form of a comment)
- Code for things such as instructor asking a student to present, checking for understanding, asking students if they want to see another example, etc. Non-content related questions.

QIn Question requesting information (looking for a specific answer)
- Only code when instructor waits for or expects an answer from students. Do not code hypothetical questions instructor asks him or herself as a teaching strategy.

QRs Question requesting reasoning (looking to understand why)
- Only code when instructor waits for or expects an answer from students. Do not code hypothetical questions instructor asks him or herself as a teaching strategy, for example, "Why would we want to find this?" immediately followed by instructor answering the question.

AnQ Listening to and answering student questions with entire class listening

Adm Administration (assign homework, return tests, general announcements about deadlines or grading, etc.)

W Waiting when there is an opportunity for an instructor to be interacting with or observing/listening to student or group activities and the instructor is not doing so
- Do not code during tests or quizzes.

O Other
- Code in unique situations (such as instructor leaving the room to get materials, working on grading while students are busy). Often used when explaining instructions for an activity to differentiate from Adm code.

TA Teaching Assistant is doing the activity marked in "Instructor doing" portion of time slot

Bth Both instructor and Teaching Assistant are doing the activity in "Instructor doing" portion of time slot

3. ICAP Framework from Chi & Wylie (2014)

Examples of Learning Activities by Mode of Engagement

	PASSIVE *Receiving*	ACTIVE *Manipulating*	CONSTRUCTIVE *Generating*	INTERACTIVE *Dialoguing*
LISTENING to a lecture	Listening without doing anything else but oriented toward instruction	Repeating or rehearsing; Copying solution steps; Taking verbatim notes	Reflecting out-loud; Drawing concept maps; Asking questions	Defending and arguing a position in dyads or small group
READING a text	Reading entire text passages silently/aloud without doing anything else	Underlining or highlighting; Summarizing by copy-and-delete	Self-explaining; Integrating across texts; Taking notes in one's own words	Asking and answering comprehension questions with a partner
OBSERVING a video	Watching the video without doing anything else	Manipulating the tape by pausing, playing, fast-forward, rewind	Explaining concepts in the video; Comparing and contrasting to prior knowledge or other materials	Debating with a peer about the justifications; Discussing similarities & differences

End of Class Items

The End of Class items provide some holistic, descriptive ratings of the entire class session. To rate *What percentage of students (approximately) participated in class?,* pick the appropriate quartile based on general impressions of how many students participated AT ANY TIME throughout the class. If it is not possible to rate this, leave it blank. The next question asks if the participation was representative with either "Yes", "No", or "NA" choices. Uses NA when class is not visible, there is no diversity in class, or there is no participation in the class.

The remaining items are rated on a 5-point scale. In general, try to rate based on how much of the class is described by the particular item. It does not need to be every minute of class. Think of the class sessions divided into quarters and rate how many of them are described by the item.

1 - Never	The behavior NEVER occurred during the class. *(Do not use if the behavior happens even once.)*
2	The description happened at least once during the class, but only characterizes a small part of class (<25%).
3	The description characterizes some of class, but not the majority (25-50%).
4	The description characterizes the majority of class, but not all of it (50-75%).
5 – Very often	The description characterizes the entire class (>75%).

In general, stick to the literal description as much as possible. Try to avoid rating based on what you think "good teaching" should be characterized by. The items do not all behave linearly. For example, an instructor may receive a low rating for *offering help to students* because (1) the instructor is missing opportunities to provide help or (2) the students do not need help. Thus, it is possible that the same rating could apply to both an instructor who lectures without fielding student questions and an instructor who poses a topic for student inquiry and then observes groups that engage in rich discussion without further instructor assistance, even though the hypothetical classes are obviously very different. There are a few items or terms that benefit from further explanation:

Work together with other students	This item should reflect students actually working with each other. If a group activity is really independent worktime where students occasionally check in with each other just to verify answers, it may not be appropriate to give a high rating on this item. This item should reflect co-engagement in rich mathematics, whatever form that may take.
Set the pace or direction of class time	This actual instance of when this happens (for example, a student asking a question or an instructor providing directions for an activity) may be much shorter than how much of the pace/direction is set by these activities. Rate this item based on the amount of class that is decided by student/instructor. For example, if the content of the entire class is generated from a few student questions, students have set the pace or direction for the class. Conversely, if students are working in groups for most of the class but it is on a structured and well-defined activity, the instructor has largely set the pace or direction for the class.
Personal feedback	Personal is the key word for this item. General feedback such as "students struggle with this concept" is not personal. However, identifying common errors from a test given to this specific class could be considered personal. This item should be rated on how responsive the instructor's feedback is to the particular group of students they are working with, whether it be individual or group-based feedback.
Positive atmosphere	Positive classrooms generally mean that student input is listened to and valued. Marks are generally high for this item, but obvious examples that may result in low ratings for this are students being ignored (e.g. raised hands that are not called on, groups asking for help where the instructor never comes around) or under-valued (e.g. instructors cut off questions/answers before students finish them or instructors jump in too quickly if students are conjecturing during a discussion or presentation).
Summarize or place class work in a broader context	Broader context may include real-world applications or also placing the lesson's topic more broadly within the course or mathematics in general.

RTOP

The RTOP is a published scale (Sawada, et al., 2002) that uses holistic, evaluative ratings of the entire class session. It is commonly used in research about instructional practices and is included here so that comparisons can be made across various types of instruments (like the TAMI-OP and End of Class items) as well as with other studies that have used this common instrument. The RTOP requires significant training to achieve decent reliability ratings, and users should consult the Training (Sawada et al., 2000) and Reference Manuals (Pilburn et al., 2000) for full information. Included below are some tips we use to be as consistent as possible in our ratings. Like the End of Class items, we generally think of the ratings as describing quartiles of the class.

0 – Never	The behavior NEVER occurred during the class/ The item does NOT AT ALL describe the class. *(Do not use if the behavior happens even once.)*
1 – at least once	The description happened at least once during the class, but only characterizes a small part of class (<25%).
2	The description characterizes some of class, but not the majority (25-50%).
3	The description characterizes the majority of class, but not all of it (50-75%).
4 – Very descriptive	The description characterizes the entire class (>75%).

Like the End of Class items, we interpret them as literally as possible. Some RTOP items can be quite difficult to rate because they require the rater to infer about the instructor or student mindsets, require deep content knowledge, or may require knowledge of the course sequence outside of the individual session. There are some items that benefit from further explanation:

In this lesson, student exploration preceded formal presentation.	In order to receive a high rating on this item, there needs to be both (1) student exploration and (2) some type of formal presentation, whether that comes from the instructor or the student. Student exploration alone is not sufficient without then formally sharing those results in some form. Sometimes, this may happen in a later class. This can't always be inferred, but sometimes instructors will mention their intentions to do so in a later session.
Students made predictions, estimations and/or hypotheses and devised means for testing them.	In a mathematics context, we generally interpret this item to be about making conjectures and exploring them or proving them with mathematical processes like writing equations, making drawings or graphs, doing sample calculations, or using logic and other proof methods.
Students were involved in the communication of their ideas to others using a variety of means and media.	In a mathematics context, we generally interpret this item to mean using graphs, equations, drawings, gestures, or proofs to convey mathematical reasoning.
The teacher's questions triggered divergent modes of thinking.	These include types of questions that can have multiple correct answers. Usually, this means it is not simply the simple calculation of a value or recitation of the name of a procedure to use. We find that this item aligns quite well with the Qin/QRs distinction of the TAMI-OP.
There was a high proportion of student talk and a significant amount of it occurred between and among students.	Though the item does not stipulate it, we interpret this item to mean on-topic, mathematical student talk. We only provide high ratings for classes where students are communicating with each other about mathematics in a deep engaging way. We do recognize that mathematical communication may happen in forms other than just talking and may be full class, in small groups, or one at a time (e.g. during a presentation) as long as there is back-and-forth communication between students.
The metaphor "teacher as listener" was very characteristic of this classroom.	Again, we do not interpret this one completely literally. We assume the intent is that the instructor is listening to student communication about mathematics or "listening" to students communicate their thinking in other ways such as via a drawing or writing activity. Broadly, we interpret this item to mean that the instructor is receptive and responsive to student mathematical thinking and communication.

Works Cited

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist , 49* (4), 219-243.

Hora, M. T., Oleson, A., & Ferrare, J. J. (2013). *Teaching Dimensions Observation Protocol (TDOP) user's manual.* Madison, WI: Wisconsin Center for Education Research.

Laursen, S. L., Hassi, M.-L., Kogan, M., Hunter, A.-B., & Weston, T. (2011). *Evaluation of the IBL Mathematics Project: Student and instructor outcomes of Inquiry-Based Learning in college mathematics.* (Report to the Educational Advancement Foundation and the IBL Mathematics Centers. Boulder, CO: Unversity of Colorado Boulder, Ethnography & Evaluation Research, Assessment & Evaluation Center for Inquiry-Based Learning in Mathematics.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). Reformed teaching observation protocol (RTOP) reference manual. Tempe, Arizona: Arizona Collaborative for Excellence in the Preparation of Teachers. http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Chemistry%20Assessments/RTOP%20Reference%20Manual.pdf

Sawada, D., Piburn, M., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). Reformed teaching observation protocol (RTOP) training guide. Tempe, Arizona: Arizona Board of Regents. http://physicsed.buffalostate.edu/pubs/RTOP/RTOPTrgGd_IN002.pdf

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., et al. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics , 102* (6), 245-253.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE- Life Sciences Education , 12* (4), 618-627.