The Inquiry-Oriented Instructional Measure

George Kuster[1] · Estrella Johnson[2] · Rachel Rupnow[2] · Anne Garrison Wilhelm[3]

Abstract

Here we report on the development process of the Inquiry Oriented Instructional Measure (IOIM), an instrument for scoring a lesson along seven inquiry-oriented instructional practices. The development of the IOIM was a multi-phase, iterative process that included reviewing K-16 research literature, analyzing videos of classroom instruction, and pilot testing. This process resulted in the identification of instructional practices that support the successful implementation of inquiry-oriented instruction (IOI) at the undergraduate level. These practices, which comprise the IOIM, provide an empirically grounded description of IOI. In addition, the IOIM provides a rubric for evaluating the degree to which an instructor's classroom instruction is reflective of these practices. As a proof of concept for the IOIM, we present the results of a pilot test – in which data from a large professional development program designed to support undergraduate mathematics instructors in implementing inquiry-oriented instruction was scored using the IOIM.

Keywords Inquiry-oriented instruction · Instructional measure · Teaching

Student-centered forms of instruction have been shown to support positive outcomes for undergraduate mathematics students (see Freeman et al.'s 2014 meta-analysis for a synthesis of this research). Additionally, empirical studies demonstrate that Inquiry Based Learning (IBL) is a more equitable form of instruction and leads to greater affective and cognitive gains when compared to non-IBL teaching methods (Laursen et al. 2014; Kogan and Laursen 2014). These outcomes directly align with the calls for improving undergraduate Science, Technology, Engineering, and Mathematics (STEM)

✉ Estrella Johnson
strej@vt.edu

[1]   Department of Mathematics, Christopher Newport University, Newport News, VA, USA

[2]   Department of Mathematics, Virginia Tech, Blacksburg, VA, USA

[3]   School of Education, Southern Methodist University, Dallas, TX, USA

education. Ferrini-Mundy and Güçler (2009) noted all of the calls for reform in STEM education focused on increasing student understanding of concepts, providing equitable access to students, and transitioning away from traditional teaching approaches to those that are student-centered and involve strategies that encourage active learning. With the push to increase the quality of STEM education, helping instructors understand, conceptualize, and implement such forms of instruction is an important step.

Instructional measures are tools that can be used by various groups within the education community to support the successful reform of undergraduate education. In particular, measures provide a shared vernacular and specific descriptions of instructional practices that can promote instructional change. Additionally, researchers can utilize measures to assess the impact of instructional interventions, and practitioners can use measures to guide their reflection and focus on improving specific aspects of their instruction.

Our work on a National Science Foundation funded project,[1] *Teaching Inquiry-oriented Instruction: Establishing Supports* (TIMES), necessitated and supported the development of such an instructional measure - one that could be used to to characterize, communicate, and evaluate instruction. TIMES was a research and development project with the goal of designing, investigating, and evaluating a system of supports for mathematicians interested in instructional change. Specifically, this system of supports was aimed at exploring what is needed to aid interested instructors in learning how to successfully implement inquiry-oriented instruction (IOI) with existing inquiry-oriented curricula materials. As we developed materials and protocols for our instructional support model, communicated with our participant instructors about the intended implementation of IOI materials, and evaluated our participants' instruction (and the efficacy of our instructional support model) it became paramount to be able to clearly articulate, communicate, and measure IOI.

There are two purposes of this article: (1) to report on the development process of the inquiry-oriented instructional measure (IOIM), a measure for evaluating the degree to which a lesson consists of practices that reflect IOI, and (2) to explicate the role of empirical research in the development of the IOIM. Before discussing the development of the measure, we start by situating and characterizing IOI.

Situating and Characterizing IOI

As mentioned, the IOIM was borne out of necessity as we worked to characterize, communicate, and evaluate instruction as part of the TIMES project. There were three curricula at the center of the TIMES project: Inquiry-Oriented Abstract Algebra (IOAA) (Larsen et al. 2013), Inquiry-Oriented Differential Equations (IODE) (Rasmussen et al. 2017), and Inquiry-Oriented Linear Algebra (IOLA) (Wawro et al. 2017). All three of these curricula were developed in accordance with the same instructional design theory: Realistic Mathematics Education (Freudenthal 1973). As a result of this shared theoretical grounding, the IOI materials share similarities in regard to instructional goals and intended implementation. It is this shared theoretical grounding, instructional goals, and implementation intentions that are the foundation of what we refer to in this paper as IOI. In this

section we broadly discuss the influence of the instructional design heuristics of RME on IOI and situate IOI in reference to IBL. We conclude with a characterization of the principles and practices of IOI that emerged as part of our development of the IOIM.

Realistic Mathematics Education (RME) conceives of mathematics as a human activity (Freudenthal 1973) and argues that instruction should be guided by a "hypothetical learning trajectory along which students can reinvent formal mathematics" (Gravemeijer and Doorman 1999, p. 126). These hypothetical learning trajectories, often presented as task sequences in the IOI curricula materials, are designed to use students' informal and intuitive ways of reasoning as starting points from which to build more sophisticated and formal mathematical understandings. While the RME design heuristics provided a theoretical basis for the development of these learning trajectories, the IOI instructional materials are also well grounded in empirical research.

The task sequences have been developed, tested, and refined through classroom teaching experiments (Cobb 2000). These teaching experiments entail extensive collection and analysis of data to document student reasoning in the form of classroom videos, artifacts of student work, and individual problem-solving interviews with students (Cobb 2000). Thus, the inquiry-oriented task sequences are built on research, in which refinements are informed by the nature of students' mathematical reasoning about the tasks, with the goal of maximizing students' opportunities to engage meaningfully in reinventing important mathematical ideas. The cycles of inquiry and formalization that comprise the hypothetical learning trajectories, supported by the task sequence and guided by the instructor, are usually carried out with students working in collaborative small-groups and whole-class discussions.

The decision to prioritize student engagement in authentic mathematical activity speaks to some of the ways in which IOI is similar to other forms of "active learning", such as inquiry-based learning (IBL). The characterizations of the students' participation within IOI and IBL classrooms are largely the same: both forms of instruction usually entail students working in groups on purposefully constructed task sequences, participating in whole-class discussions where they communicate and refine their ideas with other students through argumentation (explaining, justifying, etc.), and developing their own internalized understandings of important mathematical ideas by engaging in authentic mathematical activity (Laursen and Rasmussen 2019). In addition, progression through the curriculum for both forms of instruction is largely driven by student generated ideas and reasoning (Laursen et al. 2014; Speer and Wagner 2009; Wawro 2015).

While the nature of what students do in IOI and IBL classrooms are markedly similar, the role of an IOI *instructor* is much more specifically characterized in the literature. IBL instructors are responsible for "guiding, managing, coaching and monitoring student inquiry" (Laursen and Rasmussen 2019). In contrast IOI requires that, as students inquire into the mathematics, instructors inquire into students' thinking about the mathematics and draw on student work in specific ways to move forward the mathematical agenda of the class (Rasmussen and Kwon 2007). Starting with research-based tasks designed to promote rich mathematical thinking and reasoning, IO-instructors elicit student generated contributions, and through inquiry interpret them, decide which are useful, and then determine how to use them to move the classroom toward developing the lesson's intended mathematical idea (Speer and Wagner 2009). Thus within IOI, the notion of inquiry is as crucial an activity for the instructors as it is for the students, and makes up much of an instructor's in the moment work in the classroom.

We argue that this is one key instructional aspect that distinguishes IOI from other forms of inquiry instruction. Furthermore, the RME notion of reinvention – in which students' informal reasoning is elicited and leveraged in order to develop the formal mathematics – is a defining characteristic of IOI that may or may not be present in IBL.

Starting with the three IOI curricula (Inquiry-Oriented Abstract Algebra, Inquiry-Oriented Differential Equations, and Inquiry-Oriented Linear Algebra) allowed us anchor our work of developing the IOIM within the RME literature and a robust research base on the implementation of the IOI materials. Additionally, by considering the similarities between IOI and other instructional approaches (such as IBL) we were able to reflect on which aspects of IOI were critically necessary to capture on our instructional measure. However, we were only able to glean broad strokes by situating IOI in relation to RME and IBL. It was not until we finalized development of the IOIM that we were able to characterize IOI as a set of four instructional principles, which are supported by seven instructional practices. In the following sub-section we present this characterization. Then, in the methods, we detail the IOIM development process.

Inquiry-Oriented Instruction: Principles and Practices

Broadly speaking the research literature describes the role played by an IOI teacher as: inquiring into student mathematics, with regard to both individual student thinking and the overall learning trajectory (Rasmussen and Kwon 2007; Johnson and Larsen 2012); being an active participant in the development of the classroom's mathematics, both in terms of the mathematics of the moment and the mathematical trajectory intended by the curricular materials (Johnson 2013; Johnson and Larsen 2012); and bridging the gap between where the students are and the mathematical goals of the lesson (Wagner et al. 2007; Speer and Wagner 2009). Taking this research as our base, our work developing the IOIM resulted in the identifcation of four key instructional principles underpinning the successful implementation of IOI: *generating student ways of reasoning, building on student contributions, developing a shared understanding,* and *connecting to standard mathematical language and notation* (Kuster et al. 2017).

Generating student ways of reasoning includes engaging students in mathematical tasks so their thinking is explored and their ideas made public. Building on student contributions consists of engaging students in each others' ideas and using them (in potentially unforeseen ways) to direct class toward a mathematical goal. Developing a shared understanding highlights the importance of supporting each individual in developing commensurate ways of thinking, reasoning, and notating specific ideas. Connecting to standard mathematical language and notation involves transitioning students from the idiosyncratic mathematical notation and terms used in class to standard definitions and notation, such as "groups" or "phase planes." The four instructional principles are supported through the enactment of seven instructional practices. Additionally, these instructional practices provide detail about how teachers inquire into and leverage student thinking and reasoning to progress the mathematical agenda. A description of each of these practices follows.

Practice One

Teachers facilitate student engagement in meaningful tasks and mathematical activity related to an important mathematical point. This practice includes engaging students in

cognitively demanding tasks that promote authentic mathematical activity such as conjecturing, justifying, and defining (Jackson et al. 2013; Speer and Wagner 2009). It is important to note that while the mathematical topics the students are engaged in are important, the quality of the mathematical activity in which the teacher engages the students is the focus of this practice. To clarify, practice one is not intended to assess the quality of tasks themselves but rather to assess *the quality of the mathematical activity* (e.g., conjecturing, defining, arguing, etc.) promoted by the teacher. Engaging students in discipline specific practices is an essential aspect of inquiry; Laursen and Rasmussen (2019) note that engaging students in authentic mathematical activity is a key principle across all characterizations of inquiry.

In IOI, engaging students in meaningful tasks provides a context for the intended mathematical ideas to develop. Specifically, the tasks and mathematical activity provide a medium through which the teacher can elicit rich student thinking and reasoning, assess student thinking and reasoning, and subsequently create a path for the continued development of important mathematical ideas. Smith and Stein (2011) note that tasks requiring a high-level of cognitive demand – *procedures with connections* and *doing mathematics* – will elicit rich student thinking and reasoning. Thus, it is important for teachers to promote student engagement at these levels with appropriate questions and tasks. In this case we are using the notion of cognitive demand to characterize the quality of the mathematics the instructor is promoting. Therefore, facilitating student engagement in meaningful tasks and mathematical activity supports eliciting student contributions.

Practice Two

Teachers elicit student thinking and contributions. Before one can build on student contributions, one must first have student thinking and reasoning to build with. Therefore, once student ideas are generated they must enter a public space so the teacher can analyze their utility and, subsequently, support other students in engaging in ideas that will likely further the development of important mathematical ideas. With this in mind, IOI teachers encourage students to explain their thinking and reasoning in ways that "uncover the mathematical thinking behind the answers" (Hufferd-Ackles et al. 2004, p. 92). By seeking mathematically rich contributions from their students, teachers promote continued and deeper engagement in the tasks, which supports generating student ways of reasoning. In addition, once these contributions enter a public space, they provide opportunities for students to make sense of each other's thinking as well as opportunities for the teacher to build on student thinking (Hufferd-Ackles et al. 2004; Leatham et al. 2015). Therefore, eliciting student thinking and contributions also supports building on student contributions.

Practice Three

Teachers actively inquire into student thinking. Successful IOI teachers purposely and intently inquire into student thinking to promote student engagement in mathematics and to identify and better understand students' thinking and reasoning. By asking students to explain their reasoning, following up with clarification-type questions, rephrasing, etc., students are prompted to reflect on their thinking (Borko 2004; Hiebert and Wearne 1993). Thus, inquiring into student thinking supports generating student ways of reasoning. At the same time, teachers use various forms of inquiry to generate models of student

thinking and learning (Rasmussen and Kwon 2007); that is to understand how their students are making sense of the mathematics at hand. IOI teachers ask questions and engage in conversations in ways that indicate they are trying to figure out what ideas students are using, how they think about those ideas, and how they piece them together. In other words, they participate in conversations that indicate they are engaging in what Davis (1997) calls *interpretative listening* – listening to students with the desire of "making sense of the sense they are making" (p. 365*)*. This stands in contrast to engaging in discussion to evaluate the correctness of student thinking and reasoning. The models of student thinking and reasoning built through this inquiry process aid the instructor in figuring out plausible ways the intended mathematics may develop.

Practice Four

Teachers are responsive to student thinking and use student contributions to inform the lesson. IOI teachers listen to students' contributions (e.g., reasoning, methods, and justifications) and, when appropriate, use these contributions as a springboard for follow up questions and further exploration by the students. Once contributions have been made, teachers must determine which student ideas are productive, which ones to utilize, and how to utilize them to reach the mathematical goals of the lesson (Speer and Wagner 2009). These determinations are aided by models of student thinking and learning formed through the inquiry process and, importantly, may require reconsidering the development of the mathematics (Rasmussen and Kwon 2007). Responsive teachers generate instructional space in the form of tasks or questions directly related to students' existing ways of reasoning that aid students to further progress toward the intended mathematics of the lesson (Johnson and Larsen 2012; Johnson 2013; Rasmussen and Kwon 2007). The formulation of such new questions and tasks that stem directly from student generated ideas supports building on student contributions. Further, in response to a particular student's contribution - which the teacher determines contains a key idea necessary for continuing the development of the mathematics - a teacher may pose a new question or task to the entire class so they can build a commensurate understanding of that contribution. In such a case, being responsive to student thinking and using student contributions to inform the lesson creates instructional space and can also support developing a shared understanding.

Practice Five

Teachers engage students in one another's reasoning. IOI teachers regularly engage students in each other's ideas, thinking and reasoning (see Larsen and Zandieh (2008) or Johnson and Larsen (2012) for examples of this) by asking students to make sense of, to analyze, and to sometimes utilize the contributions of other students; this goes much beyond assessing for correctness. When students engage in each other's reasoning, it can prompt them to evaluate and revise their own ways of thinking and reasoning (Brendehur and Frykholm 2000; Engle and Conant 2002). In addition, when students wrestle with the ideas and ways of thinking and reasoning that originated with other students, normative ways of thinking within the classroom can develop (Tabach et al. 2014). In this way, when teachers engage students in one another's reasoning they support developing a shared understanding.

Practice Six

Teachers guide and manage the development of the mathematical agenda. The instructional units that make up the various IOI curricula were all developed to reach specific instructional goals, though the actual path one takes to reach those goals is somewhat flexible (see Lockwood et al. 2013 for more information about how such flexibility is built into the IOAA instructional materials). Given the mathematics develops from the students, this often means the teacher must guide student exploration and discussion in ways that promote their development of the lesson's intended mathematical ideas. Since the mathematics is being co-constructed in the classroom (Yackel et al. 2003), the teacher must figure out or recognize which ideas will be useful and then productively focus student activity around them. That is to say, the instructor must ensure that student activity is leading them to the discovery of the intended mathematical ideas. Teachers can use this practice to support building on student contributions through the use of pedagogical content tools (Rasmussen and Marrongelle 2006); (re)focusing explorations on particular student contributions; or assigning new tasks that build toward a mathematical goal. In addition, once an important idea or way of reasoning has been developed, teachers can focus the class's activity on making sense of that idea, which supports the development of a shared understanding.

Practice Seven

Teachers support formalizing of student ideas and contributions, and introduce formal language and notation when appropriate. The intent behind IOI is to build formal and standard mathematics, which includes not only skills and concepts but mathematical practices as well, from (informal) student-generated ideas and ways of reasoning. This means helping students develop their own mathematics by notating, refining, and generalizing their own ideas and activity, and then helping connect their mathematics to that of the broader mathematics community. Teachers can support the development of the class's (more formal) mathematics by marking (e.g. writing an idea on the board, naming an idea, saying "that is important", etc.) key student contributions as important and prompting their utilization or subsequent refinement in future problem-solving activities. This can be achieved through the use of transformational records and generative alternatives (Rasmussen and Marrongelle 2006) or by introducing language and notation in reference to student-generated ideas. These actions help formalize student ideas and support developing a shared understanding of those ideas within the classroom. In addition, by prompting students to use their now-developed ideas to figure out formal "textbook" mathematics – introducing formal mathematics after students have developed a commensurate understanding and prompting them to make sense of it – teachers can support connecting to standard mathematical language and notation.

Summary

A summary of how each of the practices is related to the four instructional principles is provided in Table 1. Note that practices may support more than one principle.

Table 1 Principles and Practices of IOI

Principles	Practices Supporting Principle
Generating student ways of reasoning	1) Teachers facilitate student engagement in meaningful tasks and mathematical activity related to an important mathematical point. 2) Teachers elicit student reasoning and contributions. 3) Teachers actively inquire into student thinking.
Building on student contributions	2) Teachers elicit student reasoning and contributions. 3) Teachers actively inquire into student thinking. 4) Teachers are responsive to student contributions, using student contributions to inform the lesson. 6) Teachers guide and manage the mathematical agenda.
Developing a shared understanding	4) Teachers are responsive to student contributions, using student contributions to inform the lesson. 5) Teachers engage students in one another's reasoning 7) Teachers support formalizing of student ideas and contributions and introduce formal language and notation when appropriate
Connecting to standard mathematical language and notation	7) Teachers support formalizing of student ideas and contributions and introduce formal language and notation when appropriate

Method for the Development of the Inquiry-Oriented Instructional Measure

The IOIM was developed in five phases using a combination of research literature, videos of classroom instruction from both expert and novice IOI instructors, expert validity checks, and notes and memos from pilot training sessions. Broadly consistent with thematic analysis (Braun and Clarke 2006), the overall process began with codes and categories that were developed from data and iteratively refined in a process resulting in a descriptive framework of IOI. In each iteration, new data was purposefully sought to address specific questions and hypotheses that arose in previous iterations of data analysis and to further analyze and refine the emerging characterization of IOI. This process shares characteristics with Charmaz's (2006) description of *theoretical sampling,* a research practice aimed at identifying gaps in the existing categories and codes that require further exploration and elaboration. Other research methods were evident in each of the phases as well, such as aspects of Lesh and Lehrer's (2000) iterative video analysis. In the following sections we outline the work completed in each phase and how it influenced the development of the IOIM. The five phases of development can be grouped into two main categories: *creating the prototype* and *pilot testing*.

Creating the Prototype

Logistically speaking our first goal in the development process was to create a reasonable prototype that we could then subject to validity and reliability testing. This required identifying the salient aspects of IOI, determining which aspects needed to (and could) be reasonably measured, and then creating a meaningful measurement scale along which these aspects could be communicated and assessed. The process through which this prototype was developed occurred in three phases, which are discussed in the following paragraphs.

Phase 1: Defining the task - what is IOI and how do we measure it?

This phase resulted in a general understanding and vocabulary for characterizing IOI and information on how to measure teaching in general. In this phase, we searched through IOI research literature related to the development and implementation for the focal curricular materials (e.g., Johnson 2013; Johnson and Larsen 2012; Rasmussen and Blumenfeld 2007; Rasmussen and Marrongelle 2006; Stephan and Rasmussen 2002; Wagner et al. 2007; Wawro 2015) in order to identify the defining characteristics of IOI. After reading though the literature we created the following very general, framing of IOI: "In inquiry-oriented teaching, the students, task sequence, and the teacher each have an important and interactive role for advancing the mathematical agenda." With this framing we then returned to the research literature and coded for "teacher", "students", and "tasks." We then analyzed the data within each code, which resulted in a starter list of instructional practices of IOI. Further, we illustrated each practice with justifications and examples from the literature.

We then turned our attention to existing instructional measures to determine: 1) if they adequately captured our emerging characterization of IOI, and 2) what aspects of instruction were included in other measures. Generally, evaluation tools used for research and practice are designed to serve specific purposes and to work within certain constraints, and the purposes and constraints align with the goals and limitations of the research being performed. For the TIMES project, our goal was to focus on the instructional practices in which the *teacher* engaged while in the classroom and we needed to be able to do this with video recordings of lessons. In addition, it was necessary for the IOIM to have the flexibility to be utilized across an array of undergraduate mathematics courses. Not only does the content differ across the undergraduate mathematics curriculum but the mathematical goals are often vastly different. For instance, an introductory differential equations course is often intended to develop an understanding of a set of solution methods, whereas introductory abstract algebra is often utilized to develop notions of formal mathematical proof. Instead of being overlooked, this difference in goals needed to be flexibly built into the measure.

After encountering numerous assessment tools, we focused our attention on three common observation protocols and instructional measures, as they seemed to best fit with our working definition of IOI: the instructional quality assessment (IQA), the mathematical quality of instruction (MQI), and the reformed teaching observation protocol (RTOP). In the following paragraphs we explain why we determined these tools and others like them did not fit with the specific needs of the TIMES project.

The IQA was developed with a focus on "opportunities for students to engage in cognitively challenging mathematical work and thinking" (Boston et al. 2015, p. 160). The IQA revolves around assessment of cognitive demand (Boston 2014) and, in general terms, it assesses the quality of instruction based largely on what *students* are doing and saying during the lesson. In practice this requires observing classrooms and analyzing assigned tasks and students' completed work. We determined the IQA was not a good fit as its focus on students' cognitive challenge and classroom discussion was not fine-grained enough for the purpose of TIMES.

The MQI was developed to aid in drawing connections between teacher knowledge and classroom instruction (Hill et al. 2008) and focuses on evaluating the quality of the mathematics available to students during instruction. The MQI was promising, as its development was informed by an awareness of "interactions among teachers, students, and content" (Boston et al. 2015, p. 161), an aspect that directly aligned with our

working definition of IOI. However, the MQI is pedagogically-neutral in the sense that it focuses on what mathematics the lesson consists of as opposed to how the lesson is conducted. Along these lines, the measure is very much tied to mathematical content. We determined the MQI was not suited for our needs because of its focus on the mathematical content instead of the instructor's actions. The combination of being content focused and pedagogically-neutral was the polar opposite of the needs of the TIMES project; we needed to be able to use a single rubric to measure lessons in different courses that had disjoint mathematical goals while also being sensitive to pedagogically-focused practices.

The RTOP was designed to measure the degree to which a mathematics or science lesson is reform-oriented (Sawada et al. 2002). The main goal of the RTOP was to serve as a tool for professional development aimed at improving instruction. Of the three assessment tools, the RTOP was most aligned with the goals of the TIMES project and was composed of various sub-scales consistent with the practices supporting IOI. For this reason we piloted the protocol with lessons from IOI instructors but found that the protocol, much like the IQA, focused on what students were doing as opposed to what the instructor did to promote those activities. In addition we found some aspects of IOI, such as guiding the students to a specific mathematical goal or inquiring into student thinking, were not evident in the RTOP. Thusly we decided the RTOP would not adequately capture IOI in its entirety and was not suited for our purposes.

Though it was ultimately determined existing measures were not applicable in their entirety, they did influence the refinement of practices and provided useful information that guided the underlying structure of the measure. For instance, the IQA influenced our descriptions for the various scores regarding eliciting student contributions, and the RTOP provided language describing what it means to be responsive to students. In addition, we determined that a rubric which delineated each of the practices across various levels of implementation quality over an entire lesson was appropriate. Because the role of the teacher changes depending on where the students are in the process of re-inventing a mathematical idea, we determined that a holistic score generated by considering the instructional practices over an entire instructional unit was most suitable. Thus, for our purposes, the unit of analysis for a lesson was typically a multi-day sequence of class meetings extending from the beginning of an instructional unit to the end.

Phase 2: Examining data - Addressing measure limitations

Although the analysis of the research literature in Phase 1 led to the identification of numerous practices specific to IOI, from a practical standpoint, the utilization of a measure required being able to *observe* these practices. In this phase, we cycled between analyzing videos and checking existing literature to verify that the practices identified from the literature in Phase 1 were also evident and observable in classroom instruction. In the first pass through the video data, we watched two expert teachers (IO curriculum developers) and three novices. The variation in experience level was purposeful. We felt the difference between experience levels would accentuate key aspects of instruction while also allowing us to capture their developmental range, which would be useful when later delineating implementation quality. While watching these videos, we documented the classroom events with content logs, coded for

practices, and wrote narratives for each of the practices based on what was observable in the videos. We then used the narratives to inform the descriptions of the practices. We repeated this process using new and previously watched videos (as a check that our descriptions were growing as opposed to shifting) until the narratives we were writing no longer provided new language for the descriptions. Thus, the characterizations of the practices are grounded in both supporting literature and video data.

Once the practices were defined and descriptions of their enactment were created, we set out to delineate each practice across five levels of implementation quality (i.e., high, medium-high, medium, medium-low and low). Using the video data, we created a rubric for scoring the quality of the implementation of each practice by ranking each of the various teachers' implementations of each practice. Watching videos of experts and novices, and writing narratives for their implementations of the practices made this process much smoother. Specifically, by comparing each narrative within each practice to the general description of that practice, we were able to rank the narratives in terms of which best matched the description. We then used the highest ranked narratives as the high scores, the next highest as the medium high, and so on. To make the rubric more cohesive, we then identified themes within the various levels of quality by comparing across the practices within each of the scores. Two important factors emerged: the quality of the mathematical activity and who was engaged in that activity (i.e., the teacher or the students). For example, a lesson in which a teacher consistently explores student-generated ideas while the class is watching will be scored lower than a lesson in which a teacher prompts students to explore another student's idea. Similarly, a lesson in which a teacher who regularly draws connections between student ideas and standard mathematical language and notation will be scored lower than a lesson in which a teacher engages the students in completing this activity themselves.

The process of characterizing the instructional practices evident in the videos from low to high raised important questions regarding how these practices connected to each other and how they fit within the broader role of the teacher. During this phase we began positing categories, what we called instructional principles, under which these practices could be placed. Understanding how the practices related to each of the principles impacted the instructional support we provided to instructors, helped inform descriptions of the principles and practices, and also guided some of our actions in Phases 3 and 4.

Phase 3: Refinement using outside sources

In this phase, we began seeking resources from beyond IOI research literature and feedback from mathematics education researchers not directly involved in the development of the measure or familiar with IOI. Our intentions here were to uncover key aspects of IOI that we did not notice because of our familiarity with it and to gauge if the descriptions we created communicated our ideas as intended, even to those not familiar with IOI. First, we asked a mathematics education researcher not familiar with IOI to code two videos with the drafted rubric. We then met with them, discussed areas of confusion, and worked out discrepancies between scores. This resulted in a list of practices and qualitative descriptions that needed further elaboration or refinement. To help with the refinement process we began searching through K-12 research literature looking

for aspects of K-12 instruction that were commensurate with the practices we identified in IOI. We then used the language and descriptions from the K-12 literature to refine and elaborate our descriptions as needed. This also led to a better understanding of the four instructional principles and their supporting practices.

For instance, one area of confusion - how much say the teacher has over the direction of the developing mathematics - was clarified by the notion that teachers must not just have class discussions, but that these discussions must be used by the teacher to advance the mathematical agenda (Jackson et al. 2013). This played a key role in quality descriptions of *guiding and managing the development of the mathematical agenda*; one's instruction should indicate the existence of a clear mathematical goal, as opposed to leaving the direction completely in the hands of the students. As another example, Practice 1, *engaging students in meaningful tasks and mathematical activity,* borrows heavily from Stein et al. (2008) notion of *doing mathematics* - the extent to which students engaged in cognitively demanding tasks and used mathematical argumentation to support or refute any claims. Specifically, lessons were scored higher when teachers typically asked questions eliciting justifications or explanations from their students as opposed to eliciting responses focused on the correct answer or memorization.

At the end of Phase 3 we felt confident that we had identified the instructional practices necessary for successfully implementing IOI and that our rubric communicated an accurate characterization of the practices as well as the various degrees at which they could be enacted. As such we moved on to validity testing.

Pilot Testing for Validity and Reliability

Though we were confident the characterizations provided in the rubric accurately reflected the teaching practices we identified through the analysis of literature and video data, we still needed to verify the measure contained (and was not missing) salient practices of IO-instructors. Additionally, we needed to confirm the measurement scales were accurately capturing *meaningful* differences in the implementation quality of each of the practices. Once this step was complete we could then move on to pilot testing.

Phase 4: Sharing to clarify (validity testing)

In this phase, our intent was to pilot the use of the rubric with IOI experts and novices, and this piloting was completed in two steps. Our goals were to validate the measure with experts, clarify connections between the principles and practices, and understand how to effectively communicate our interpretation of them. We first asked four experts in undergraduate mathematics education who were already familiar with IOI but not familiar with the measure to use the rubric to score the same lesson. In addition, we asked them to provide separate scores indicating how well they felt each of the four instructional principles were reflected in the lesson. This step led to another iteration of refinement of the elements in the rubric (e.g., we better delineated between characterizations of implementation quality based on expert feedback), and had two immensely important outcomes. Most importantly, despite no training, the scores across all six researchers (the four experts and two rubric developers) were all within one point.

Thus, the descriptions in the rubric were generally meaningful to researchers familiar with IOI. Secondly, having scores at the level of both the principles and practices allowed for better determining which practices supported which principles. Specifically, we looked for instances in which principles were scored high but (proposed) supporting practices were not or vice versa. Cases of mismatched scores between principles and practices alerted us to potential instances in which we had incorrectly attributed a certain practice to supporting a certain principle.

We then engaged in a pilot training process where we trained two relatively novice mathematics education graduate students, each having no background in IOI on how to use the rubric. During this process we asked the scorers to take careful notes of issues that arose for them as they utilized the rubric. We also recorded the meetings when we met to discuss the scores they assigned. From this process we concluded that two of the initially developed practices (teachers introduce language and notation when appropriate and teachers support formalizing of student ideas and/or contributions) were, in practice, capturing the same aspects of instruction. Therefore, we condensed them into a single practice, captured by practice seven, *Teachers support formalizing of student ideas and contributions, and introduce formal language and notation when appropriate*. We also created resources within the rubric for scorers, including guiding questions, "evidenced by" descriptions for each practice, and we boldfaced certain words in the rubric for emphasis.

Phase 5: Sharing to use (pilot testing)

In this phase, our goal was to train and utilize multiple scorers to consistently score 42 TIMES videos. To this end, we created resources supplimentary to the rubric, including: a frequently asked questions document, a document listing tips to remember while watching and while scoring videos, a definitions page where frequently used rubric terms like "instructional space" could be explained in detail, and examples of justifications and scores given to the videos used in training.

Additionally, based on discussions during training, a few small changes were made to the rubric to clarify possible times for scores to be given: (1) in practice 5, additional text and boldfacing were added to scores low through medium-high to highlight distinctions from one score to the next, and (2) in the medium score for Practice 7, the format of the description was changed from two implicit cases to three explicit cases in which the score would be given. In what follows, we describe the IOIM and present the pilot testing.

Description of the Inquiry-Oriented Instructional Measure The IOIM consists of a rubric (see supplementary materials and http://times.math.vt.edu) that measures the degree to which a lesson is inquiry-oriented by examining the quality of the enactment of each of the seven practices. Within the rubric, each of the practices is scored on a 5 point Likert-scale from low to high which is done holistically over the lesson. These "lessons" can range from just one class session to an entire instructional unit, (e.g., 2.5 h of instructional time spread across multiple class sessions). For each of the practices, the quality of the activity promoted by the instructor distinguishes a low score (1) from

a high score (5) on the rubric. Take for example Practice Two: teachers elicit student thinking and reasoning. If an instructor evokes solely procedural contributions from students, the lesson would score much lower, medium-low (2), than if they routinely have students share their thinking, reasoning, and justifications, which would earn a score of 5. In the following paragraph we outline the key attribute(s) being captured within the rubric for each practice.

For Practice One the rubric captures the extent to which the teacher engages students in "doing mathematics," or the extent to which students are engaged in cognitively demanding tasks and mathematical activity, such as arguing, conjecturing and justifying (Stein et al. 2008). The key factor for delineating Practice Two concerns the degree to which the teacher elicits rich mathematical reasoning and explanations from students, as opposed to eliciting memorized facts. When measuring Practice Three, the scores are separated by the level at which the teacher seems to probe students' statements and reasoning in order to improve their own understanding of the students' mathematics. Being responsive to student thinking (Practice Four) is measured based on how much the teacher uses students' questions and ideas as the building blocks for the developing mathematics. Scores for Practice Five are broken up based on the extent to which the teacher prompts and allows students to make sense of each others' reasoning without acting as a filter that interprets ideas for the students. The scores for Practice Six exhibit the level to which the teacher guides and manages a lesson that reaches a clear mathematical goal using student reasoning and contributions. Lastly, the rubric for Practice Seven gauges the degree to which the teacher transitions from students' (perhaps task specific) language and notation to standard mathematical language and notation.

Pilot Testing the IOIM As mentioned above, our final phase of measure refinement involved training scorers to consistently score 42 lessons from the TIMES project. In what follows, we first describe the TIMES study participants and the resulting data that we analyzed using the IOIM; then we describe the training process; and, finally, we address the results from the coding of videos. We present this use of the TIMES data, in part, as a proof of concept for the instructional measure. The IOIM was designed to capture and assess the instructional approach promoted by TIMES instructional supports. Thus, scoring the TIMES fellows allowed us to see if there was consistency between the instruction and the instructional measure – allowing us to see if the measure was capturing the intended instructional approach and if the supports were having the intended outcome.

TIMES instructional support model and participants As part of the TIMES project, 42 mathematics instructors (13 Abstract Algebra, 13 Linear Algebra, and 16 Differential Equations instructors) participated as fellows. The 42 fellows were from various locations throughout the United States, from a variety of institution types, and a wide range in classroom contexts. These fellows received three forms of professional development support: curricular support materials, summer workshops, and online working groups. The curriculum materials are formatted as task sequences that include rationales, examples of student work, and implementation suggestions. (See Lockwood et al. (2013) as an example.) In addition to the IO instructional materials, TIMES fellows were also supported through a summer workshop and online working groups. These summer workshops spanned three days and were designed to help instructors

develop an understanding of IOI and their role as the teacher; develop a shared vision of instruction and student learning goals; and develop a familiarity with the curriculum materials, task sequences, and online resources. The format of these workshops cycled between cross-content sessions with broader discussion about IOI and content specific break-out sessions where the focus was on the curricular materials.

Then, during the semester of implementation, instructors participated in weekly online working groups facilitated by project personnel. These meetings had 3–5 participants and each meeting had two main components. Roughly twenty minutes of each meeting were dedicated to providing "just in time" support for the instructors. This included discussions around managing small-group work, sharing homework and assignment ideas, and debriefing about how the instructors were progressing through the materials. The other forty minutes of the meeting were structured to be a lesson study (Lewis et al. 2006). Each online working group selected two instructional units as their focus. For each of the two focal instructional units, the working group would first discuss the mathematics of the lesson, followed by a discussion of student learning goals and implementation considerations. After instructors taught the unit, they would share clips of their instruction for group reflection and discussion. A common thread throughout each of these sessions was attending to the principles of IOI – generating student ways of reasoning, building on student contributions, developing a shared understanding, and connecting to standard mathematical language and notation.

Data Video classroom data of the implementation of the second focal unit was scored using the IOIM. This video data was collected from a single iPad generally located at the back of the room and focused on the instructor. Thus, there are some limitations with the data; we were not able to capture small group discussions and during whole class discussion it was not always possible to identify or clearly hear the students as they made contributions. As a result, the training documents directed scorers to take whole class discussion as the primary source of evidence when scoring the lesson.

Here, the instructor enacted lesson is the unit of analysis. The length of these lessons varied from one to four and three-quarter hours, and often spanned multiple class meetings. All 13 IOAA fellows recorded the same unit, as did all 13 IOLA fellows. Each IODE online working group was provided the opportunity to select the focal units that best met the needs of that particular group of instructors (e.g., a unit every fellow was planning on implementing or one that the participants felt would be particularly difficult to implement). As a result, the lessons collected and scored from the 16 IODE fellows were not all from the same instructional unit.

As we only scored one lesson from each instructor, we acknowledge that we cannot make claims about an instructor's "typical teaching practice", nor are we trying to document change that may have occurred through the participation in the TIMES program. Instead we scored these videos to provide a "snapshot" of the TIMES fellow's instruction.

Training and scoring We implemented the full-scale training of five graduate students from various mathematics education backgrounds (including undergraduate, K-12, and pre-service teacher education). Training started with having the scorers watch sample video clips from instructors using the IOAA, IOLA, and IODE curricular materials that exemplified the different levels of IOI for each of the practices. As training progressed, scorers were given more opportunities to watch longer segments of classroom video with a partner or on

their own each evening and to justify their own scores using the rubric. In group meetings, scorers engaged in facilitated debates of their scores, which allowed misunderstandings of terms and weaknesses in justifications to be resolved. The scorers were also asked to write in their own words what each practice would look like at high, medium, and low levels as another check of their understanding of the rubric and synthesis of group discussions. At the end of a week of training, scorers were given a test video to determine their readiness to code independently. All trainees gave scores within 1 level of the trainer's scores, which allowed them to be released to independently score videos gathered from TIMES instructors.

Following the training period, the six scorers (five newly trained scorers and the trainer) each scored eight to twenty-one new videos from the sample of 42. All videos were scored by at least two scorers, and scoring responsibilities were distributed across courses. Scorers were allowed to choose their own pace for scoring videos. In order to maintain continued reliability, the trainer double-coded each scorer after every fifth video to make sure all scores remained within 1 of the trainer's scores. When comparing other scorers' scores with trainer scores, they were within 1 point 87.4% of the time and in exact agreement 42.8% of the time. In cases where the scorer was off by two levels, they were asked to rewatch and rescore the video in light of the discussion with the trainer before being allowed to continue scoring videos. Because all videos were double-coded, we were also able to assess inter-rater reliability by comparing the two sets of scores for each video. This comparison resulted in slightly higher inter-rater reliability: 95.2% agreement within 1, 57.1% exact agreement, average kappa of .357, and average Spearman rho of .753. According, to Landis and Koch (1977), this average kappa falls near the top of the "fair" agreement range and Gwet (2010) suggests that kappa scores are often negatively skewed when the actual scores are not well distributed across the potential scoring range. We list all of these indicators of inter-rater reliability to provide a more robust picture of the inter-rater reliability (Wilhelm et al. 2018). One set of scores (i.e., one score for each of the seven practices) was created for each lesson by averaging scores between the two scorers.

Analysis Scores for the TIMES video data were analyzed descriptively overall and by course assignment. First, consistency was assessed using correlations. Then, summary statistics and a number of visual representations were used to describe the data. Finally, Wilcoxon Rank-Sum and Wilcoxon Signed-Rank Tests were used to describe the statistical significance of qualitative differences between courses or practices.

Results of pilot testing First, with respect to consistency, we found that scores for the seven practices were highly correlated (Rupnow et al. 2018). Correlations ranged from .683 (Practice Five and Practice Seven) to .935 (Practice One and Practice Three), with a mean of .840. Further, with respect to the trends in the scores, we found that, across all 7 practices, the median score for TIMES fellows' lessons was between medium (3) and medium-high (4). (See Fig. 1.)

Comparing the practices, lessons scored lowest on Practice Three, "Teachers actively inquire into student thinking." In fact, Wilcoxon Signed-Rank Tests revealed that Practice Three scores were statistically significantly lower than any of the other practices (see Table 2). In contrast, lessons received their highest scores on Practice Six, "Teachers guide and manage the development of the mathematical agenda." Again, Wilcoxon Signed-Rank Tests revealed that Practice Six scores were statistically significantly higher than all of the other practices.
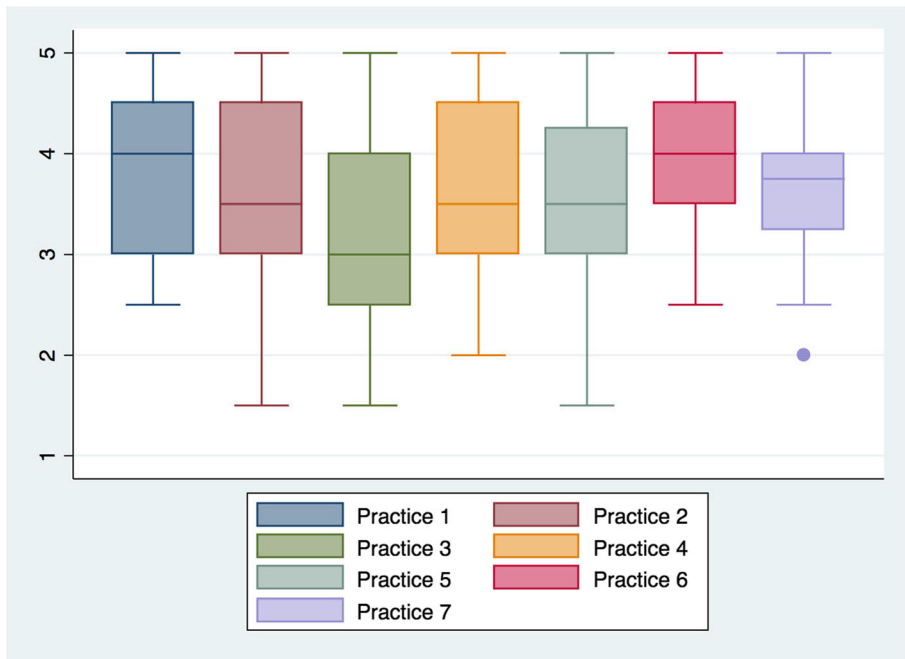
Fig. 1 Box plot of seven practices across all 3 courses

When comparing scores across courses there were some interesting contrasts. Figure 2 displays the score distributions for the seven practices across the three different courses. Immediately, scores for IOLA stand out as higher than IODE and IOAA. Recall that the number of instructors in each of these samples was relatively small. When comparing the scores for the seven practices across the different courses, Wilcoxon Rank Sum tests revealed few statistically significant differences. The only two even marginally statistically significant results are in comparing Practices Four and Five between IOLA and IOAA. Scores for IOLA lessons were significantly higher than scores for IOAA lessons on both Practice Four ($p < .05$) and Practice Five ($p < .10$). We discuss these findings in greater detail below.

Table 2 Summary of findings from wilcoxon signed-rank tests in comparing scores for practices

	1	2	3	4	5	6	7
1		+**	+***	+***	+**	-*	+*
2			+***	=	=	-***	=
3				-**	-**	-***	-**
4					=	-***	=
5						-***	=
6							+*

Row compared with column (e.g, "+" in 1,2 indicates scores on practice 1 significantly greater than scores on practice 2) (*$p < .05$, **$p < .001$, ***$p < .0001$)
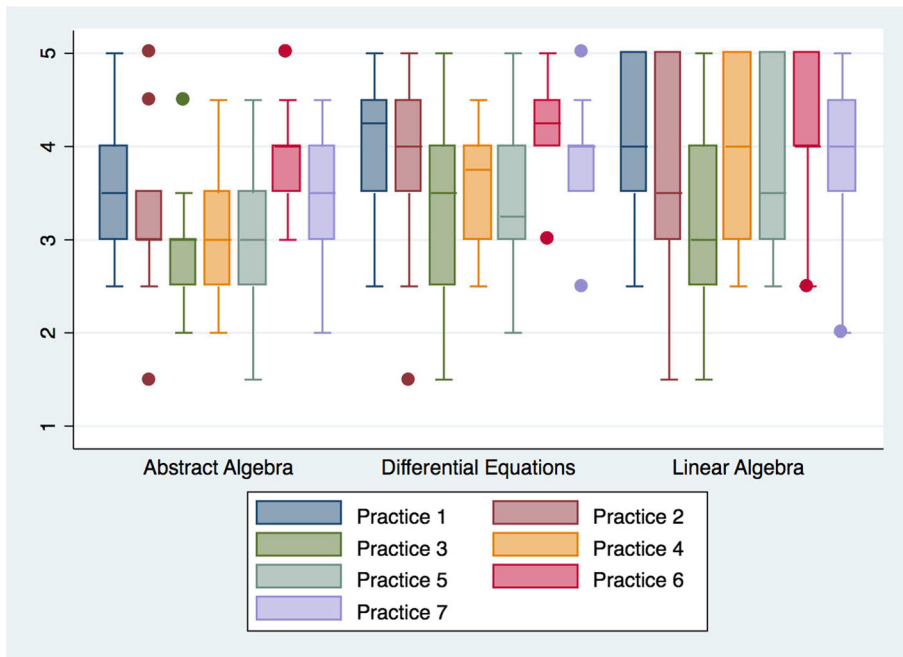
Fig. 2 Box plot of seven practices by courses

Discussion of pilot testing As a proof of concept, we see the successful application of the IOIM to the TIMES data as very promising for several reasons. First, the training protocol resulted in a reasonable inter-rater reliability (within 1 point). Suggesting that the IOIM can reliably be used to score enacted instructional lessons holistically. Second, the lesson scores were highly correlated, which offers some construct validity given that all the lessons scored came from instructors participating in the same IO instructional supports. Third, the IOIM was sensitive enough to capture differences between practices (i.e., Practice Three scores were statistically significantly lower than any of the other practices) and differences between content areas (i.e., scores for IOLA stand out as higher than IODE and IOAA). This sensitivity is notable given that all instructors were taking part in the same professional development program. As we were able to capture vairiance within this very narrow context, we are hopeful that this measure may also be able to capture differences between commensurate forms of instruction, such as IBL and IOI.

In and of themselves, these results offer several points worthy of discussion - particularly in regard to the differences found between the practices and between the content areas. The fact that Practice Three scores were statistically significantly lower than any of the other practices while Practice Six scores were statistically significantly higher than all of the other practices raises questions about IOI and the IOIM. For instance, without more lessons that represent a wider range of instructional approaches, it is impossible to know if these differences are present because Practice Three is more difficult to enact or because it is more difficult to find evidence for it when scoring lessons. In terms of the differences between the content areas, more research needs to be done to understand if these differences can be understood in terms of the written

lessons (e.g., did some lessons have a greater potential for high-level IOI?), in terms of the instructors in each content area (e.g., familiarity with content or prior experience with IOI), in terms of the scorers (e.g., familiarity with the content), or in terms of the instructional supports (e.g., differences in implementation of the online working groups). These additional avenues for research offer additoanl opprotunities to refine the IOIM and the training documents.

Discussion

In this paper, we outlined the development of a rubric for measuring IOI, and showed the role empirical research played in this development process. For TIMES, the specific research project that motivated this instructional measure, the IOIM was needed to characterize, communicate, and evaluate the specific type of instruction we were trying to support. More generally however, the development of the IOIM represents an effort to operationalize inquiry-oriented instruction. With the publication of some very prominent and visible studies showing the benefits of active learning (e.g., Freeman et al. 2014; Laursen et al. 2014), there is a growing consensus that active learning "works". While these results of student success and learning are certainly compelling, questions remain regarding how different groups of students may be benefiting (or not) in different ways from more active forms of instruction.

For instance, Laursen et al. (2014) found that, "in non-IBL courses, women reported gaining less mastery than did men, but these differences vanished in IBL courses" (p. 415). However, Johnson et al. (in press) found a gender performance gap in IOI courses, with men outperforming women, that was not present in non-inquiry oriented courses. This pair of studies is illustrative of Eddy and Hogan's (2014) argument that any classroom intervention will impact different groups of students in different ways. Furthermore, the disparate results emerging from two very similar instructional approaches (IBL and IOI) bolster Singer and colleagues' (Singer et al. 2012) call for the identification of critical instructional features in order to explore the ways in which particular instructional approaches may impact various groups of students. Without instruments that allow us to distinguish between what teachers are actually doing in their classroom, the field is unable to theorize, document, or understand how aspects of that instruction supports student learning. Thus, the IOIM represents our attempts to characterize instruction – with the idea that this characterization will allow for further research into the relationships between instructional practice and student learning.

Creating a measure for IOI at the undergraduate level presented non-trivial and unique challenges. First, it was necessary for the IOIM to have the flexibility to be utilized across an array of undergraduate mathematics courses. This was challenging due to the content differences and the differing mathematical goals. For instance, in our focal courses, mathematical goals range from developing an understanding of solution methods to developing notions of formal mathematical proof and definitions of abstract concepts. Instead of being overlooked, this difference in goals needed to be flexibly built into the measure

and was captured by including language such as "formalizing student ideas" and "students were engaged in generalizing their thinking." Such language is true to the tenets of IOI while not prescribing the sophistication of the students' mathematical work or the exact content that work needed to consist of. The extent that the IOIM can be used in other disciplines with commensurate instructional approaches, such as physics, warrants investigation.

Second, the IOIM needed to incorporate a wide variety of instructional strategies. From a theoretical standpoint, in IOI the teacher navigates along the continuum of pure telling and pure student exploration (Rasmussen and Marrongelle 2006). From a practical standpoint, flexibility across the range of IOI was necessary because of the nature of the TIMES project: supporting instructional change. Thus, the measure needed to provide information regarding how the participating instructors were incorporating aspects of IOI into their instruction and to what degree they were doing so. Based on the scores of the TIMES fellows' lessons, the IOIM appears to be sensitive enough to discriminate between inquiry-oriented instructors. More work needs to be done to verify that the IOIM can be used for different forms of instruction (such as interactive lecture and IBL), that the IOIM can discriminate between other forms of instruction, and that the IOIM can be used to assess change over time.

Another line of future work will include looking at the IOIM scores of the TIMES fellows' lessons as both a dependent variable, allowing us to investigate the influence of TIMES instructional supports, and as an independent variable, allowing us to investigate the impact of IOI instruction on student learning outcomes. This will provide the field with much needed information about supporting instructional change and the relationship between instructional practice and student learning.

Compliance with Ethical Standards

Conflict of Interest   On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15.

Boston, M. D. (2014). Assessing instructional quality in mathematics classrooms through collections of students' work. In *Transforming mathematics instruction* (pp. 501–523). Springer International Publishing.

Boston, M., Bostic, J., Lesseig, K., & Sherman, M. (2015). A comparison of mathematics classroom observation protocols. *Mathematics Teacher Educator, 3*(2), 154–175.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Brendehur, J., & Frykholm, J. (2000). Promoting mathematical communication in the classroom: Two preservice teachers' conceptions and practices. *Journal of Mathematics Teacher Education, 3*, 125–153.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative research*. London: Sage Publications Ltd.

Cobb, P. (2000). Conducting classroom teaching experiments in collaboration with teachers. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 307–334). Mahwah: Lawrence Erlbaum Associates Inc.

Davis, B. (1997). Listening for differences: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education, 28*(3), 355–376.

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education, 13*(3), 453–468.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399–483.

Ferrini-Mundy, J., & Güçler, B. (2009). Discipline-based efforts to enhance undergraduate STEM education. *New Directions for Teaching and Learning, 2009*(117), 55–67.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.

Freudenthal H. (1973) *Mathematics as an Educational Task*, D. Reidel, Dordrecht.

Gravemeijer, K., & Doorman, M. (1999). Cotext problems in realistic mathematics education: A Calculus course as an example. *Educational Studies in Mathematics, 39*, 111–129.

Gwet, K. L. (2010). *Handbook of Inter-Rater Reliability* (Second ed.). Gaithersburg: Advanced Analytics, LLC.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30*(2), 393–425.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511.

Hufferd-Ackles, K., Fuson, K. C., & Sherin, M. G. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education, 35*(2), 81–116.

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education, 44*(4), 646–682.

Johnson, E. (2013). Teachers' mathematical activity in inquiry-oriented instruction. *The Journal of Mathematical Behavior, 32*(4), 761–775.

Johnson, E. M., & Larsen, S. P. (2012). Teacher listening: The role of knowledge of content and students. *The Journal of Mathematical Behavior, 31*(1), 117–129.

Johnson, E., Andrews-Larson, C., Keene, K., Melhuish, K., Keller, R., & Fortune, N. (in press). Inquiry and inequity in the undergradaute mathematics classroom. *Journal for Research in Mathematics Education*.

Kogan, M., & Laursen, S. L. (2014). Assessing long-term effects of inquiry-based learning: A case study from college mathematics. *Innovative Higher Education, 39*(3), 183–199.

Kuster, G., Johnson, E., Keene, K., & Andrews-Larson, C. (2017). Inquiry-oriented instruction: A conceptualization of the instructional principles. *PRIMUS, 28*(1), 13–30.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Larsen, S., & Zandieh, M. (2008). Proofs and refutations in the undergraduate mathematics classroom. *Educational Studies in Mathematics, 67*(3), 205–216.

Larsen, S., Johnson, E., & Weber, K. (Eds.). (2013). The teaching abstract algebra for understanding project: Designing and scaling up a curriculum innovation. *Special Issue of the Journal of Mathematical Behavior, 32*(4), 691–790.

Laursen, S.L. & Rasmussen, C. (2019). More than Meets the I: Inquiry Approaches in Undergraduate Mathematics. *Proceedings of the Twenty-second Special Interest Group of the Mathematical Association of America on Research in Undergraduate Mathematics Education Conference on Research in Undergraduate Mathematics Education*. Oklahoma City, OK.

Laursen, S. L., Hassi, M. L., Kogan, M., & Weston, T. J. (2014). Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study. *Journal for Research in Mathematics Education, 45*(4), 406–418.

Leatham, K. R., Peterson, B. E., Stockero, S. L., & Van Zoest, L. R. (2015). Conceptualizing mathematically significant pedagogical opportunities to build on student thinking. *Journal for Research in Mathematics Education, 46*(1), 88–124.

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analyses of conceptual change. In *Handbook of research design in mathematics and science education* (pp. 665–708).

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3–14.

Lockwood, E., Johnson, E., & Larsen, S. (2013). Developing instructor support materials for an inquiry-oriented curriculum. *The Journal of Mathematical Behavior, 32*(4), 776–790.

Rasmussen, C., & Blumenfeld, H. (2007). Reinventing solutions to systems of linear differential equations: A case of emergent models involving analytic expressions. *The Journal of Mathematical Behavior, 26*(3), 195–210.

Rasmussen, C., & Kwon, O. N. (2007). An inquiry-oriented approach to undergraduate mathematics. *The Journal of Mathematical Behavior, 26*(3), 189–194.

Rasmussen, C., & Marrongelle, K. (2006). Pedagogical content tools: Integrating student reasoning and mathematics in instruction. *Journal for Research in Mathematics Education*, 388–420.

Rasmussen, Keene, Dunmyre & Fortune (2017) Inquiry-oriented Differential Equations. https://iode.wordpress.ncsu.edu/. Accessed 04 Arp 2019.

Rupnow, R., LaCroix, T., & Mullins, B. (2018). Building lasting relationships: Inquiry-oriented instructional measure practices. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, & S. Brown (Eds.), *Proceedings of the 21st annual conference on research in undergraduate mathematics education conference on research in undergraduate mathematics education* (pp. 1306–1311). San Diego.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102*(6), 245–253.

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Smith, M. S., & Stein, M. K. (2011). *5 Practices for Orchestrating Productive Mathematics Discussions*. Reston: National Council of Teachers of Mathematics.

Speer, N. M., & Wagner, J. F. (2009). Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *Journal for Research in Mathematics Education*, 530–562.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340.

Stephan, M., & Rasmussen, C. (2002). Classroom mathematical practices in differential equations. *The Journal of Mathematical Behavior, 21*(4), 459–490.

Tabach, M., Hershkowitz, R., Rasmussen, C., & Dreyfus, T. (2014). Knowledge shifts and knowledge agents in the classroom. *The Journal of Mathematical Behavior, 33*, 192–208.

Wagner, J. F., Speer, N. M., & Rossa, B. (2007). Beyond mathematical content knowledge: A mathematician's knowledge needed for teaching an inquiry-oriented differential equations course. *The Journal of Mathematical Behavior, 26*(3), 247–266.

Wawro, M. (2015). Reasoning about solutions in linear algebra: The case of Abraham and the invertible matrix theorem. *International Journal of Research in Undergraduate Mathematics Education, 1*(3), 315–338.

Wawro, M., Zandieh, M., Rasmussen, C., & Andrews-Larson, C. (2017). The inquiry-oriented linear algebra project. http://iola.math.vt.edu. Accessed 04 Arp 2019.

Wilhelm, A. G., Gillespie Rouse, A., & Jones, F. (2018). Exploring differences in measurement and reporting of classroom observation inter-rater reliability. *Practical Assessment, Research & Evaluation, 23*(4).

Yackel, E., Stephan, M., Rasmusen, C., & Underwood, D. (2003). Didactising: Continuing the work of Leen Streefland. *Educational Studies in Mathematics, 54*, 101–126.