Journal of Mathematics Teacher Education
https://doi.org/10.1007/s10857-023-09595-2

Comparing how college mathematics instructors and high-school teachers recognize professional obligations of mathematics teaching when making instructional decisions

Inah Ko[1] · **Patricio Herbst**[2] · **Mollee Shultz**[3]

Accepted: 30 July 2023

Abstract
This paper investigates how mathematics instructors' recognition of the professional obligations of mathematics teaching varies based on their institutional environment, specifically whether they teach high school or college mathematics. Using an instrument that measures instructors' recognition of four hypothesized professional obligations, we surveyed 471 US high school mathematics teachers and 239 university mathematics instructors to measure the extent to which they recognized professional obligations when evaluating the appropriateness of certain instructional actions. After testing measurement invariance of four item sets, each of which measures one of the four hypothesized professional obligations—disciplinary, institutional, interpersonal, and individual obligations-, we compared the instructors' recognition of each of the four obligations between the two groups. We found that university instructors recognized the institutional obligation more than high school teachers, while recognizing the individual and interpersonal obligations significantly less. This investigation provides insight into the variation in the nature of mathematics teaching practice across different levels of schooling.

Keywords Professional obligations · High school mathematics teaching · College mathematics teaching · Measurement invariance

This work was done with the support of grants DRL-0918425 and DUE-1725837 to Patricio Herbst. All opinions are those of the authors and do not necessarily represent the views of the foundation.

✉ Inah Ko
inahko@umich.edu

Patricio Herbst
pgherbst@umich.edu

Mollee Shultz
mshultz@psesd.org

[1] Office of Graduate and Postdoctoral Studies, University of Michigan Medical School, Ann Arbor, MI, USA

[2] Marsal Family School of Education, University of Michigan, Ann Arbor, MI, USA

[3] Puget Sound College and Career Network, Puget Sound Educational Service District, Renton, WA, USA

Published online: 30 August 2023

Introduction

What kind of practice is mathematics teaching? Our field has seen an increase in studies concerned with the work of teaching mathematics over the last thirty years. This owes in no small part to theoretical advances in education research writ large that overcame prior focus on the individual characteristics of teachers or on the generic behaviors of teachers in the classroom and put more emphasis on the activity of instruction (Cohen et al., 2003; Hiebert & Stigler, 2017). International studies like TIMSS have addressed that question across countries for a given level of schooling (Hiebert et al., 2003; Stigler & Hiebert, 1999). As a contribution to our community's understanding of what kind of practice mathematics teaching is, we demonstrate how to investigate differences in mathematics teaching practice across different levels of schooling. Specifically, we examine differences in terms of the extent to which high school and university level mathematical instructors in the U.S. recognize obligations to the multiple stakeholders to the profession.

Cohen et al.'s (2003) *instructional triangle* highlights that interactions among teachers, students, and content occur in environments. Conceptualizing what those environments are like and how they might differentially influence instruction is an important goal for fundamental research on mathematics teaching. Chazan et al. (2016) proposed that mathematics teachers' work in institutional environments is subject to four professional obligations. These are (1) an obligation to the interpersonal dynamics and relationships instantiated in social groups; (2) an obligation to individual students; (3) an obligation to the schooling institutions; (4) an obligation to the discipline of mathematics. The professional obligations are resources instructors can use to justify their actions deviating from what is customary in instruction. For example, a teacher who recognizes the obligation to care for individual students' needs might use it to justify the action of adjusting lessons to engage a particular student.

Our present study uses the hypothesis that these obligations apply to mathematics instructors across different levels of schooling and that average recognition of those obligations by various groups of instructors may vary. We use that hypothesis to create a measure of recognition of each obligation and ask whether the average amount of recognition of each of those obligations is different for instructors at different levels of schooling: We compare groups of high school and lower division undergraduate mathematics instructors in the U.S. as a case of a particular society where one can expect the obligations to apply similarly. Our comparison illustrates one way to study how the environment of different levels of schooling could matter in instruction.

Research on the work of teaching across different instructional systems is important not only as a phenomenon of interest on its own but also as a resource for teachers who are often given recommendations based on research done in other contexts. In particular, university instructors are often asked to change their instructional practices based on research that has largely been conducted in K-12 settings (e.g., Rasmussen & Kwon, 2007). There is a growing body of research that recognizes the importance of the study of teaching in efforts toward improving practice (i.e., practice-based teacher education; Ball & Forzani, 2009) and that focuses on the work of teaching in K-12. But so far, the practices of mathematics teachers at the collegiate level have largely been understudied (Speer et al., 2010). Though there exists scholarship about teaching mathematics in college that documents the role of activity structures (e.g., lecture, small group work) on students' achievement and other outcomes (e.g., Sofroniou & Poutos, 2016; Sonnert et al., 2015), very few studies focus on understanding the practice of college mathematics teaching, including the rationale for specific teaching actions and decisions during instruction (Speer et al., 2010, p. 101). The lack of fundamental understanding of the practice of college mathematics teaching makes it difficult to assess whether the evidence for recommendations for practice developed on the basis of K-12 research is appropriate for the college level.

While we assume that the four obligations apply to mathematics teaching both at the high school and at college level, we expect that the extent to which individual instructors recognize each obligation will vary. As an initial foray distinguishing between environments, and to illustrate the methodology proposed, we consider the high school level and the university level, though surely more in-depth studies could look at distinctions within those (e.g., private vs. public) and wider studies should also look at differences among elementary, middle, and high school teachers. When we started this study, we did so with some conjectures. We suspected that high school teachers, expected to teach a wider range of students and to interact with students in a wider range of activities, would be more likely to recognize the individual and the interpersonal obligations than university instructors (who could expect to treat students as adults with individual responsibility). We suspected that university mathematics instructors, being in more frequent contact with mathematical research, might be more likely to recognize the obligation to the discipline. And we questioned the institutional obligation: While university instructors' academic freedom made us suspect that they might feel less obligated to the institution and the demands from district and school administrators suggested that high school teachers might be more obligated to the schooling institutions, we also realized that some practices at the department level (e.g., textbook selections, common final exams) particularly affect university instructors of lower division (freshman and sophomore) mathematics courses (Rasmussen & Ellis, 2015).

With such hypotheses, we ask the research questions: Are college mathematics instructors, on average, more or less likely than high school mathematics teachers to recognize each of the obligations? In earlier work we had constructed measures of recognition of each obligation based on the extent to which practitioners agree with instructional decisions that depart from an instructional norm on account of each obligation. In this study we inspect whether these measures can be used to compare groups and then make such comparisons. To answer these questions, we gauge the extent to which high school teachers and college instructors recognize each of the four obligations. To measure instructors' recognition of the obligations, we presented instructors with classroom scenarios, represented as storyboards using cartoon characters, and in which an instructor or teacher had done an action that responded to opportunities or demands of the moment and disrupted what the instructor had avowedly planned to do. In each of those scenarios, the teacher's action might be justified as a response to one of the obligations, and respondents were asked to evaluate what the teacher had done. The participants' responses to the items were then analyzed by employing psychometric methods including item factor analyses and tests of measurement invariance between the two groups of instructors.

In this paper, we offer a tool for researchers to measure mathematics teachers' recognition of obligations related to discipline, institution, individual, and interpersonal aspects of their professional practice. By providing empirical evidence of the measures' validity and invariance across different teacher populations, researchers interested in measuring recognition for other teacher groups may consider using these measures. Additionally, we present findings that demonstrate the differing recognition of professional obligations between high school teachers and college instructors. Readers could relate these differences to the distinct institutional environments to understand how teaching practices may differ between high school and college mathematics classrooms.

This comparison has practical implications for supporting students' transition from high school to university. Specifically, understanding the differences in the extent to which high school teachers and college instructors recognize the four domains of professional obligations could shed light on some of the challenges students face when transitioning from high school to college. This transition can already be a difficult experience for many students, as they face new environmental, financial, social, and academic changes (Briggs et al., 2012; Cheng et al., 2015) for which students may not be well prepared (Perry & Allard, 2003; Upcraft & Kramer, 1995). Understanding the extent to which college instructors could be counted on to be similar or different from high school teachers could be useful for counselors to prepare their students for college and help them adjust to the transition.

Theoretical framework

This study is inscribed in the approach to the study of mathematics teaching that Herbst and Chazan (2012) have called *practical rationality*—the study of the resources available to teachers to construct a practice that is rational or sensible (see also Leatham, 2006). The practical rationality concept and methods are influenced by Bourdieu's (1990) simultaneous critique of observer-centered and participant-centered approaches to the study of practice. Inspired by Bourdieu's concept of the *habitus*, practical rationality seeks to capture the teacher's sense of being disposed to do something in a given instructional context and to represent it as a response to both the resources the context makes available to the teacher and the teacher's individual resources that the context calls upon. Thus, practical rationality studies these resources not as static elements in place, but rather as affordances and constraints for teachers to use as savvy actors in that place. What might appear to practitioners as the appropriate thing to do at a given moment (given other things happening) may not need an explicit representation for the practitioner (e.g., as a policy), but inasmuch as it can be observed recurrently it can be useful for observers to represent in the form of statements and special vocabulary. The constructs of *norms* and *obligations* are ways in which practical rationality represents what appears to practitioners as the appropriate thing to do, but these representations are valid as long as they are taken not as descriptions of the practitioner's reality but as models of this reality constructed by an observer, or statements that everything happens *as if* what is stated is the case.

Instructional norms and professional obligations

Practical rationality posits that to understand the actions and decisions instructors make we need to account for the resources available to them in practice. Hence, their actions not only express their individual traits but also respond to affordances and constraints of practice. Chazan et al. (2016) describe those conditions as issued from two distinct aspects of the work of teachers. On the one hand, there are conditions associated with the *role* a teacher plays in mathematics instruction, enabling students' mathematical learning and work: To describe the affordances and constraints of instruction, practical rationality uses the notion of *instructional norm*. Brousseau's (1997) notion of didactical contract is especially useful to make more precise what is meant by instructional norm: The teacher and student are bound to each other through mutual, differentiated responsibilities vis-à-vis the content of studies, which function like a contract. This didactical contract is tacit and relies on implicit expectations developed from earlier experiences in school. These expectations can be modeled by statements representing for an observer what the teacher or the students appear to experience as the appropriate thing to do. Examples of such instructional norms are that it is appropriate for (1) the teacher to decide what problems students have to work on in order to learn something and (2) students to expect the teacher has chosen such problems on account of intended instructional goals, even if these are implicit. While we rely on the hypothesis that every class has a didactical contract, we surmise that variability must exist in the extent to which the same norms apply across classrooms. But, while an individual teacher may endeavor to establish idiosyncratic norms in their own class, we make the assumption that such negotiations happen against a background of shared contractual norms that students have been socialized into from earlier schooling. In our present contribution it is less important to establish what specific contract is in place in a classroom than to use the hypothesis that a contract exists and that it can be represented as a system of norms, regulating the role of the teacher in instruction.

On the other hand, there are conditions associated with the *position* the instructor occupies in the institution that makes room and allocates resources for instruction. Chazan et al. (2016) describe the position of mathematics instructor as accountable to four stakeholders: Knowledge, the Client, Society, and Organization. Each of those stakeholders obligates the teacher in particular sorts of ways. The teacher is obligated to knowledge, particularly to the discipline of mathematics (disciplinary obligation). The teacher is obligated to their students, as they have to care for students as whole individuals (individual obligation). The teacher is obligated to society, as they represent and cultivate the values of a society (interpersonal obligation). And the teacher is obligated to organizations such as the department or school, as they need to perform functions that satisfy expectations of efficiency and legality (institutional obligation). Thus, professional obligations describe other elements of the context within which a teacher's decisions and actions are made. Chazan et al. (2016) also noted that for each obligation there is a need to distinguish on the one hand the obligations themselves (e.g., manifest in the discourse addressed to teachers in professional literature), and on the other hand how teachers recognize the obligation in specific instances in their practice, which provide evidence of individual variability. This paper, however, focuses on charting the distribution of this individual variability in the recognition of obligations, and comparing groups of individuals in regarding the recognition of those obligations.

Practical rationality uses instructional norms and professional obligations to account for the decisions teachers make. At moments when a teacher has to make a decision, instructional norms might describe the default decision to make in circumstances like the one at hand. Yet, an individual teacher in the specific circumstances at hand might see other decisions as conceivable, too. These alternative, possible decisions may depart from the norm but may be sensible as well. One or more of the professional obligations may provide possible justification for choosing to do something that departs from the norm. Consider, for example, a scenario in which, after introducing the topic of the day, the teacher is ready to assign a problem to the class and needs to decide how to pose the problem. Among the instructional norms activated in the moment we would hypothesize that (1) the problem should make use of the knowledge just introduced and (2) the problem should contain all the information students would need to solve it. But the instructor could conceive alternatives. For example, the instructor could consider providing less than the needed information to solve the problem. The disciplinary obligation might justify doing so as it would provide an opportunity for students to consider different possibilities for the missing information and possibly distinguishing cases in which the problem is solved in one or another way—valuable aspects of mathematical practice that might not have been among the goals for the lesson but that could be targeted incidentally if a norm was breached (see also Herbst & Chazan, 2020).

Clearly, any one of those decisions might open up different conceivable alternatives to the normative action, some alternatives might be justifiable on account of an obligation, but discourageable on account of another obligation. We surmise that whether an obligation is perceived as a justification for an action that departs from the norm depends on the extent to which the individual teacher recognized that obligation. If it is the case that different schooling levels might shape the position of the teacher differently, the notion of recognition of an obligation might provide one way of describing how mathematics instruction is different: Different obligations might have different weights in justifying (or discouraging) the same instructional decisions. For example, Marston (2010) documents that elementary teachers tend to prefer working with young people more than college instructors and college and high school instructors prefer teaching a subject matter more than elementary teachers. Could the notion of professional obligation and their instantiation as justifications of possible decisions in instruction be used to ground claims like that?

While individuals might differ in the extent to which they find particular decisions justifiable, we contend that scenarios staging decisions like those exemplified above can be used to construct a measure of the extent to which practitioners recognize a given obligation.

Constructing a measure of recognition of an obligation

To construct a measure of recognition of an obligation, we would need to identify aspects of each of the obligations that were recognizable by practitioners of different levels of schooling. Then, we would need to use those aspects to construct items in which the practitioners' recognition of the obligation played a role in explaining some of the variability in their evaluations of the appropriateness of certain instructional actions teachers do in classrooms. While we would need to show that items measure the same construct (i.e., recognition of an obligation) in the same way across populations, we would hope that the scales established from the responses from the samples of these populations provide a basis for noting differences in the extent to which the obligation is recognized. This consideration also suggests that while multiple scenarios could be used to explore what considerations are involved in recognizing one or another obligation among a group of practitioners, some scenarios might not provide common grounds for validly eliciting recognition of a given obligation across populations. If common grounds can be established across populations, they would allow for examining differences between populations regarding the same obligation.

To construct a measure that could provide for that comparison, we had to operationalize the notion of professional obligation by using its relationship with the notion of instructional norm so that specific decisions and justifications could be envisioned. To gauge the recognition of obligations that might justify departures from norms, we sampled norms that are very general. In other words, we focused on instructional norms that could apply across courses of mathematical study and across levels of schooling, including that teachers should teach the content in the syllabus or textbook, do not digress from the material being studied or make it much more general than needed for the class, and take responsibility to teach students what students will be accountable for learning. We also considered it normative that the teacher should assign problems related to what students are studying, answer students' questions about the content being studied, evaluate the correctness of students' solutions to problems, and use assessments to ascertain students' learning. Of course, we recognize that individual instructors might depart from those norms in justifiable ways, and that current educational policies and teacher development programs try to move instructors away from some of those practices. Therefore, we hasten to clarify that when we say "normative," we do not necessarily mean "correct": Rather, we mean that such behaviors would be expected in the sense that practitioners would not see it as remarkable if they were followed.

As we thought of the obligations as providing grounds for departure from instructional norms, we looked into how the different obligations might be triggered for the teacher in their daily work. The disciplinary obligation could be triggered for example by noticing unmarked incorrect knowledge (e.g., typos in textbooks), opportunities for important mathematical practices in material to be taught or in work assignments, or examples of mathematical ideas in out-of-school contexts, etc. The institutional obligation could be triggered by opportunities to follow institutional policies (e.g., pacing charts, exam schedules, course prerequisites, prescribed curriculum) and institutional rules (e.g., academic calendar, bell schedule). The interpersonal obligation could be triggered by opportunities to attend to socially desirable values (e.g., sharing, empathy) or to fulfill socially expected functions (e.g., work hard, be accountable). The individual obligation could be triggered by all sorts of students' individual idiosyncrasies, including physical or cognitive traits, emotions, etc. The specific item format that we used to operationalize these professional obligations, as well as each item statement, is described in the Methods section and Appendix A.

Professional obligations in different teacher populations

A number of empirical studies have used the obligation construct across different teacher populations. Bieda et al. (2015) used obligations to distinguish the sources of justification used by groups of experienced and novice teachers; and a comparable use was demonstrated by Lande and Mesa (2016; see also Mesa, 2014), who inspected the discourse of groups of full-time and part-time community college instructors. The obligation construct has also been used by other empirical researchers who interviewed individual teachers and identified their difficulties implementing reform practices (Kosko & Gao, 2016; Webel & Platt, 2015). In all these uses, a common outcome has been the suggestion that different kinds of practitioners recognize the multiple obligations to varying degrees.

To empirically examine teachers' recognition of the obligations, we have been developing a set of instruments, each of which is associated with one of the four obligations described above:disciplinary, institutional, interpersonal, and individual. This instrument is a scenario-based questionnaire, which we call PRofessional Obligation Scenario Evaluation (PROSE). Initial reports of the development of this instrument and validation work including focus group review for high school teachers is provided in Herbst et al. (2014), Herbst et al. (2016), and Herbst and Ko (2018). The instrument developed for college instructors is described in Shultz (2020; see also Shultz & Herbst, 2021). In the present paper, we complement that information by providing results on the differences in the professional obligations between high school teachers and college instructors, using PROSE.

Methods

Participants

This manuscript offers secondary data analysis from two different datasets. The PROSE set of instruments were completed by a nationally distributed sample of 471[1] U.S. high school mathematics teachers located across 47 states and 239[2] college instructors located across 37 states. The high school teachers were recruited from over 12,000 public high schools across 47 states using a stratified systematic probability proportional to size sampling method based on geographical region and urbanicity. One secondary mathematics teacher was randomly selected from each school, and then recruited via email. Of the 767 high school teachers who agreed to participate in the project, 471 teachers responded to at least one PROSE item. Data for college instructors came from a national sample representing 94 mathematics departments across 37 states. Participants were recruited by email to department secretaries who were asked to forward the recruitment to their faculty. The sample was representative of undergraduate mathematics instructional faculty with respect to race (Blair et al., 2018) but contained 20% more women than the national population of university mathematics instructors (Golbeck et al., 2018).

Participants used an online platform to respond to instrument items as well as to survey questionnaires asking about their educational background and teaching experience. High school teachers had been teaching mathematics for an average of 14.6 years (min = 1, max = 40) and 59.7% of them were female. College instructors taught for an average of 8.1 years (min = 1, max = 30) and 49.2% of them were female. Of the college instructors, 44.6% of the instructors were graduate student instructors, and the remaining 55.4% were part-time or full-time faculty members.

PROSE instrument

The version of the PROSE scenario-based instrument used in this study consists of four sets, one for each obligation (disciplinary, institutional, interpersonal, and individual obligation). At the beginning of each set, participants are presented with an introduction screen indicating the obligation under consideration. Figure 1 below presents the introduction screen of the disciplinary instrument for high school teachers.

As shown, items describe actions departing from the norm in a generic way saying that the teacher "deviates from a lesson" and allude to each obligation with more specific layperson's language (e.g., "attend to an issue of mathematical importance" cued the disciplinary obligation). Each set consists of multiple items, each presenting a scenario in which a teacher deviates from their plan for a reason related to the obligation being considered. After viewing each scenario, participants rate the extent to which they agree with a statement that indicates what the teacher should have done if they were following a norm. The statement says that "the teacher should have [done what was normative], rather than [what he or she was seen doing]" (for example, "The teacher should keep to what is in the

[1] 441, 448, 435, 471 for disciplinary, institutional, interpersonal, and individual PROSE instruments, respectively.

[2] 211, 217, 205, 236 for disciplinary, institutional, interpersonal, and individual PROSE instruments, respectively.

In the following storyboard we invite you to consider a scenario in which a high school teacher **deviates from a lesson in order to attend to an issue of mathematical importance**. We are interested in the extent to which you think the teacher's action is justifiable.

You can move through the storyboard at any rate you like. Use the arrows at the bottom left of the window to move between slides. Be sure to view the entire story.

When you finish viewing the storyboard, we will ask you to answer some questions about the scenario. You will be asked to rate the extent to which you agree with the actions taken by the teacher, and to comment on your rating. As you answer each question you will have the storyboard available to review.

Fig. 1 Introduction screen of disciplinary PROSE instrument. © 2014, The Regents of the University of Michigan, all rights reserved, used with permission

textbook, rather than require students to use information that is different from what is in the textbook"). The responses were recorded on a 6-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." Items were reverse coded so that high ratings indicated that the teacher would favor departing from the plan on account of the obligation at stake, implying that the participant strongly recognized the given obligation.

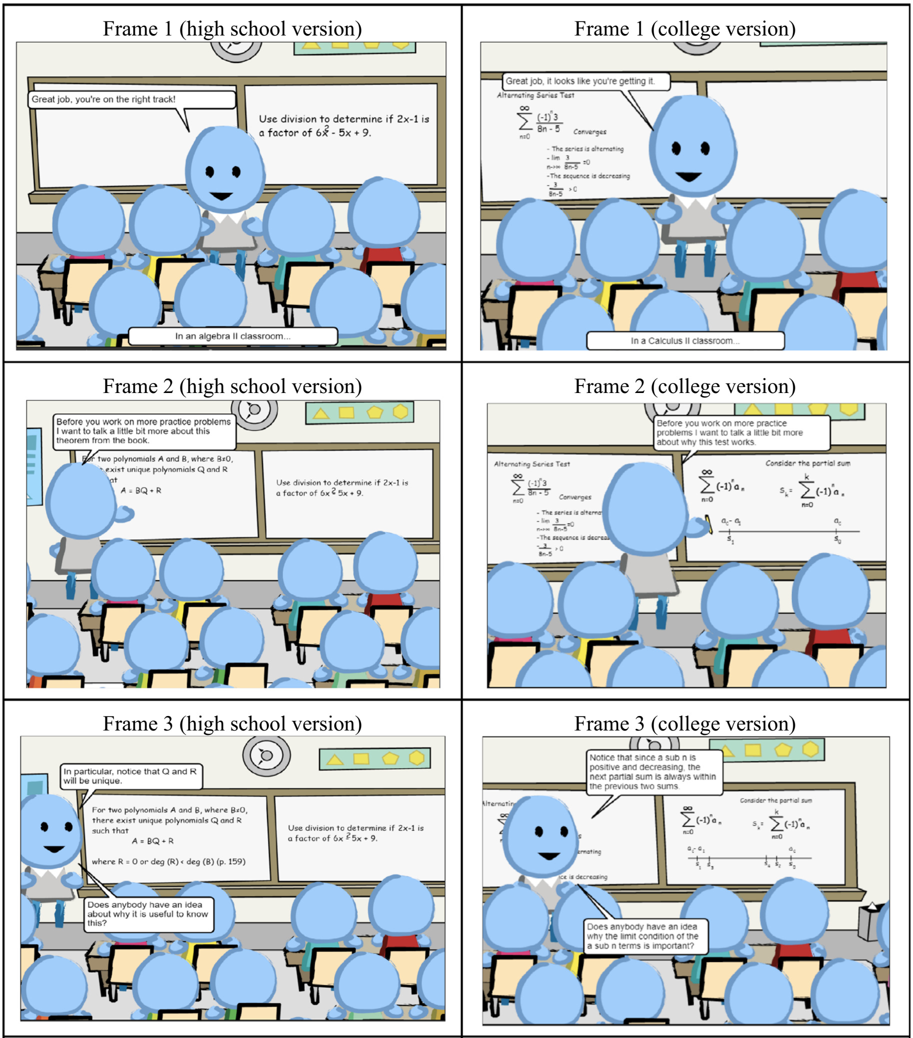
Content adjustment

To approximate scenarios presented in the items to the actual contexts of high school and college classrooms, respectively, we made adjustments for some of the items. This is similar to the process of translating questionnaires into another language in that we adapted the scenarios in the high school version episodes to scenarios happening in college classrooms by changing high school specific contents to comparable contents that were specific for college level (specifically, lower division undergraduate courses such as calculus or linear algebra). For example, as shown in the example item presented in Fig. 2, we changed the algebra content on the board (division of polynomials) to content appropriate in a calculus course (convergence of series), but maintained the same instructional decision (namely to deviate from problem review to attend to an issue of mathematical theory). The statement participants were asked to agree with at the end of the item was the same in high school and college versions.

This process enabled us to ask participants to consider the given statement in close connection with their actual practice of teaching, but this adaptation necessitated an evaluation of the degree to which the items measure the same construct across groups. To ensure that the items measure each type of obligation in the same way for both high school teachers and college instructors, we tested measurement invariance before comparing latent means and variances between the groups.

Item selection

The initial sets of items were selected from those analyzed in our prior study examining the dimensionality of each obligation using high school teachers' responses to the items (Herbst & Ko, 2018). Our ultimate goal was to examine the recognition of obligations across levels of schooling using measures representing each of the four obligations as a



Q. Please rate how much you agree or disagree with the following statement: **"The teacher should give students additional practice problems, rather than elaborate on mathematical theory."** 1 - Strongly Disagree; 2 - Disagree; 3 - Slightly Disagree; 4 - Slightly Agree; 5 - Agree; 6 - Strongly Agree	Q. Please rate how much you agree or disagree with the following statement: **"The instructor should give students additional practice problems, rather than elaborate on mathematical theory."** 1 - Strongly Disagree; 2 - Disagree; 3 - Slightly Disagree; 4 - Slightly Agree; 5 - Agree; 6 - Strongly Agree

Fig. 2 Example disciplinary item (left: high school version; right: college version). © 2014, 2018 The Regents of the University of Michigan, all rights reserved, used with permission

unidimensional construct. To accomplish this, we looked for a set of items that were highly correlated with each other and indicated a key feature of an obligation.[3] In other words, for each obligation, we selected items that loaded into one factor to estimate the degree

[3] This presents a threat to content validity that we address in the limitations.

of teachers' agreement on departing from the plan in terms of a single scale representing one unidimensional obligation under consideration. For example, in selecting interpersonal items that had shown to load onto three distinguishable dimensions in the previous study, we chose only the items that loaded onto a factor that the most interpersonal items were loaded onto. Thus, the interpersonal obligation construct represents teachers' recognition of an obligation to promote students' interpersonal relationships. As a result, 12 disciplinary, 7 institutional, 7 interpersonal, and 5 individual items were selected for the items used in this study.

Given the small number of items within each instrument, internal consistency of each instrument was evaluated in terms of not only Cronbach's alpha but also average inter-item correlation. All the instruments yielded Cronbach's alpha values greater than 0.60, a criterion considered acceptable for a small number of items such as less than 10 items (Loewenthal, 2001). Average inter-item correlations also suggested that all four instruments have acceptable internal consistency given that the suggested range of average inter-item correlation is between 0.15 and 0.25 (Clark & Watson, 1995; see Appendix B, Table 8 for more details). The acceptable internal consistency of Likert scale item responses within each instrument indicates that items within the same instrument are coherently related to each other for both groups. Given that the common characteristics of the items within the same instrument is the type of obligation presented in the items, acceptable internal consistency warrants our assumption that the targeted obligation is commonly associated with teachers' decisions departing from the planned actions.

Tests of measurement invariance in multiple-group item factor analyses

As we scale the level of an unobservable (or latent) construct (i.e., participants' recognition of an obligation) using a set of item responses, it is important to demonstrate that the way in which items are related to the targeted construct is equivalent across the compared populations. This is commonly described as testing measurement invariance; we describe here how we proceeded to do this with our four item sets.

A series of multiple-group item factor analysis was conducted to ensure comparability of the latent factor means and variances between high school teachers and college instructors. The first test conducted for measurement invariance in this study is a test for configural invariance that evaluates whether the factor structure of the item responses is the same for high school teachers and college instructors. As we started with high school items that loaded onto one dimension each (one for each obligation), to specify the configural invariance model to be tested, we constrained the item factor structure to be the same as a one-factor model across the two groups. This configural invariance model was compared, in subsequent analyses, to more restrictive models (metric invariance and scalar invariance models described below) using statistics for comparing nested statistical models.

Provided that all items measuring a single obligation load onto one factor, the second test, a test of metric invariance, was conducted to test the equality of factor loadings between the groups by constraining the factor loadings to be the same between the two groups in addition to the constraints specified in the configural invariance model. This second test attempts to demonstrate that the rate of change in an item score to the change in the construct score is equivalent across the two groups. Finally, a test for scalar invariance that evaluates the equality of item thresholds (Brown, 2006, p. 268) was conducted after establishing partial invariance in the first and the second invariance tests. Scalar invariance, if established, would add to metric invariance assurance that item scores are equivalent at a given level of the construct being measured. The scalar invariance model was specified by adding constraints on the item thresholds between the groups. When the fit of the scalar invariance model was significantly worse than that of a less constrained model (e.g., partial metric invariance model), we identified the source of misfit based on modification indices and the item contents and conducted a partial measurement invariance test (Byrne et al., 1989) in which some constraints of the items suggested by the indices are freed.

We evaluated the invariance not only using the nested model comparisons with the DIFFTEST option, which relies on adjusted chi-square values for the WLSMV estimator, but also using the model fit changes suggested by Cheung and Rensvold (2002) to supplement the comparison test which might be sensitive to sample size. In terms of model fit, we examined the change in CFI and considered that the invariance hypothesis can be retained if the change is equal to or less than −0.01. After ensuring that the same construct is being measured in the same way between the groups of high school teachers and college instructors, we proceeded to examine latent means and variance differences between the two groups of participants. In the specification of all the multiple-group models, theta parameterization was used with the WLSMV estimator using Mplus v.7.4 (Muthén & Muthén, 1998–2015).

Operationally, the research questions presented in the Introduction can be re-stated as follows:

- Does each of the four factors reflecting teachers' recognition of the four professional obligations (disciplinary, institutional, interpersonal, individual) satisfy measurement invariance across high school teachers and college mathematics instructors? In other words, does each unidimensional factor allow for valid statistical comparison of latent factor mean and variance across groups?
- How do the variances and means of these latent factors differ between high school teachers and college mathematics instructors?

Results

In this section, we first report the results of measurement invariance tests. After ensuring that the instruments measure the same constructs between high school instructors and university instructors, we present the results of the structural invariance tests examining the equality of factor means and factor variances between the two groups. The series of measurement invariance and structural invariance were conducted separately for each instrument measuring one of the four professional obligations. The summary model fit statistics and comparison results derived from measurement invariance tests are presented in Table 1.

Measurement invariance

Recognition of the disciplinary obligation

First, we tested whether our hypothesized unidimensional model with the items representing disciplinary obligation fits each group. The item factor analysis conducted with high school teachers and with college instructors shows that one-factor model where all the 12 items are loaded on one factor provided good model fit with each group, respectively

Table 1 Results of the tests of measurement invariance

Instrument	Model	CFI	TLI	RMSEA	ΔCFI	$\Delta\chi^2$	Δdf	p
Disciplinary	Configural	0.963	0.954	0.060				
	Metric	0.968	0.964	0.054	0.005	18.327	11	0.074
	Scalar	0.934	0.950	0.063	−0.034	189.998	59	<0.001
	Scalar_P1 (A114x freed)	0.950	0.961	0.055	−0.018	126.495	54	<0.001
	Scalar_P2 (A114x, A111x freed)	0.966	0.973	0.046	−0.002	63.031	49	0.086
Institutional	Configural	0.989	0.983	0.036				
	Metric	0.993	0.991	0.026	0.004	4.780	6	0.572
	Scalar	0.928	0.954	0.059	−0.061	101.293	34	<0.001
	Scalar_P1 (A217x freed)	0.957	0.971	0.047	−0.036	66.668	29	0.001
	Scalar_P2 (A217x, A205L freed)	0.982	0.986	0.032	−0.011	36.268	24	0.052
Interpersonal	Configural	0.983	0.973	0.057				
	Metric	0.976	0.969	0.061	−0.007	16.972	6	0.009
	Metric_P1 (A303x freed)	0.977	0.983	0.053	− 0.006	9.214	5	0.101
	Scalar	0.913	0.944	0.082	−0.064	147.547	34	<0.001
	Scalar_P1 (A311x freed)	0.957	0.970	0.059	−0.020	71.391	29	<0.001
	Scalar_P2 (A311x, A301x freed)	0.968	0.975	0.054	−0.009	50.030	24	0.001
Individual	Configural	0.984	0.969	0.055				
	Metric	0.988	0.983	0.041	0.004	4.127	4	0.389
	Scalar	0.858	0.925	0.085	−0.130	110.838	24	<0.001
	Scalar_P1 (A403L freed)	0.920	0.952	0.068	−0.068	64.987	19	<0.001
	Scalar_P2 (A403L, A409x freed)	0.960	0.971	0.052	−0.028	33.146	14	0.003
	Scalar_P3 (A403L, A409x, A402L freed)	0.975	0.978	0.045	−0.013	17.938	9	0.036

(RMSEA = 0.062, CFI = 0.960, TLI = 0.950 for high school teachers; RMSEA = 0.057, CFI = 0.969, TLI = 0.962 for college instructors). The model included correlated errors for the two items that have to do with how closely to follow the textbook (items A102x and A109x in Appendix A, Table 4). As shown in the fifth and sixth columns of the Table, all the standardized item factor loadings were greater than 0.3 and significant at the $p < 0.001$ level.

After ensuring that the unidimensional model fit well with each of the groups, we proceeded to investigate the model of configural invariance that tests equality in factor structure between the two groups. In the test, factor variance and factor mean were set to 1 and 0 in each group and all factor loadings and thresholds were set to be freely estimated. The model showed good model fit (Disciplinary Configural in Table 1), implying that it is reasonable to assume the same factor and pattern of factor loadings in the measurement model representing the relationships between the items and the latent factor across the groups of participants.

After confirming the equality of the factor structure, we tested the model of metric invariance that tests equality of factor loadings. In the model, all item thresholds were set to be estimated and the factor mean was fixed to 0 in both groups. The factor variance was fixed to 1 in the group of high school teachers (reference group), whereas it was freely estimated in the group of college instructors. The DIFFTEST comparing configural and this metric invariance model suggested that the metric invariance model is not significantly worse than the configural invariance model (DIFFTEST $\chi^2(11) = 18.327$, $p = 0.074$), indicating the equality of factor loadings was held. Other model fit statistics were also improved in the metric invariance model compared to the configural model (Disciplinary Metric in Table 1).

The scalar invariance model was specified by constraining all the factor loadings and all the item thresholds to be equal across the two groups. In the model, the factor mean and variance were set to 0 and 1, respectively for the group of high school teachers, whereas they were freely estimated for the group of college instructors. The DIFFTEST comparing the previous metric invariance model (less constrained) and the scalar invariance model suggested that the scalar invariance model is significantly worse than the metric invariance model, DIFFTEST $(59) = 189.998$, $p < 0.0001$. Following the methods recommended by Byrne et al. (1989), we examined the source of non-invariance and continued invariance evaluation. To identify the non-invariant items, we used the modification indices (reflecting the improvement of model fit associated with freeing the constraints) and sequentially relaxed the constraints until an acceptable partial measurement invariance model could be established. As a result, the model allowing different thresholds of A114x and A111x between the groups (Disciplinary Scalar_P2) was shown to be not significantly worse than the metric invariance model (DIFFTEST χ^2 $(49) = 63.031$, $p = 0.086$). The change in CFI also suggested no meaningful difference between the partial scalar model P2 (Disciplinary Scalar_P2) and the metric model (Δ CFI < 0.01). This result implies that the item thresholds of these items (A111x, A114x) are different at the same degree of recognition of disciplinary obligation. Specifically, the differences suggested that college instructors are more likely to agree with a teacher's action taking time to make connections to other mathematical ideas (A111x) or elaborating on the mathematical theory (A114x), when controlling for the level of disciplinary obligation.

Recognition of the institutional obligation

Following the same steps conducted for testing measurement invariance of the disciplinary items, analyses were conducted for the institutional items. The unidimensional model fit the data well for each of the data from high school teachers (RMSEA = 0.028, CFI = 0.994, TLI = 0.990) and college instructors (RMSEA = 0.056, CFI = 0.960, TLI = 0.940). The model fit indices of the unidimensional factor structure suggested that overall factor structure is invariant across groups (Institutional Configural in Table 1). After confirming the configural invariance, with the correlated errors between the two items A202x and A211x that have the same beginning clauses, the equality of factor loadings was tested. Model fit indices suggested that the metric invariance model is not significantly worse than the configural model (DIFFTEST $\chi^2(6) = 4.780$ $p = 0.572$) (Institutional Metric in Table 1).

The equality of item thresholds (scalar invariance) was then examined across the groups. The scalar invariance model was significantly worse than the metric invariance model. However, the partial scalar invariance model where the thresholds of the items A217x andA205L are freed did not significantly decrease the model fit compared to the metric invariance model (DIFFTEST $\chi^2(24) = 36.268$, $p = 0.052$) (Institutional Scalar_P2).

Recognition of the interpersonal obligation

The same series of measurement invariance tests were conducted for the interpersonal items. The one-factor item analysis conducted within each group suggested that it is reasonable to assume that the seven items coherently measure one construct (RMSEA = 0.07, CFI = 0.971, TLI = 0.953 for high school teachers; RMSEA = 0.011, CFI = 0.999, TLI = 0.999 for college instructors). In the model, errors between the item A303L and A304L were set to be correlated given the similar language used in those items' scenarios ("move on to" another work) which point to the normative practice of progressing in instruction by completing new tasks.

The test of configural invariance suggested that the unidimensional structure is invariant across the groups (Interpersonal Configural in Table 1). Next, we achieved a partial metric invariance (Interpersonal Metric_P1 in Table 1) after allowing different factor loadings for the item A303x, which contributed less to the high school teachers' recognition of interpersonal obligation than that of college instructors. A partial scalar invariance was then achieved after allowing different item thresholds of the item A311x and A301x (Interpersonal Scalar_P2 in Table 1), as evidenced by the change in CFI. To evaluate the impact of the difference in the criterion we use, we conducted structural invariance models (factor mean and factor variance comparison) under the two different assumptions on the degree of partial invariance. The results are reported in the Structural Invariance section below.

Recognition of the individual obligation

The model fit indices of the one-factor model with the five individual obligation items suggested that the five items coherently explain a significant amount of the variance of the latent construct (RMSEA = 0.037, CFI = 0.992, TLI = 0.984 for high school teachers; RMSEA = 0.080, CFI = 0.973, TLI = 0.946 for college instructors). Next, the configural invariance was tested and the model provided good fit statistics, indicating that the factor structure is invariant across the groups (Individual Configural in Table 1).

The configural invariance model was then compared to the metric invariance model. Model indices suggested metric invariance between the groups (DIFFTEST $\chi^2(4) = 4.127$, $p = 0.389$), indicating the equality of item factor loadings on the latent construct between the groups (Individual Metric in Table 1). Next, the metric invariance model was compared to the scalar invariance model. The full scalar invariance model significantly decreased model fit (DIFFTEST χ^2 (24) = 110.83, $p < 0.0001$) relative to the metric invariance model. After a series of model comparisons, we could obtain a partial scalar invariance model where the thresholds of the items A403L, A409x, A402L are freely estimated (Individual Scalar_P3 in Table 1). Then, we compared the factor mean and factor variance under the scalar partial invariance model in Section Structural invariance.

In summary, following the recommendation to use a change of CFI to test measurement invariance (Cheung & Rensvold, 2002) and the recommendation from Byrne et al. (1989) that one completely invariant item is enough for group comparisons, we conclude that partial measurement invariance was established for the four targeted constructs across the groups of participants. However, it is important to bear in mind the possible sources of non-invariant items across groups in that we had to release more than a half of the total item parameters for items measuring the recognition of the individual obligation. This might imply that high school teachers and college instructors recognize this obligation in different ways to a certain degree or the scenarios represented in those non-invariant items may not be equivalently applicable across high school and college mathematics classrooms. We discuss this in more detail in the Discussion.

Structural invariance

Having established partial metric and scalar invariance across the groups meant that the variance and mean of the factors representing recognition of each obligation are comparable between the groups. The structural invariance analysis tests whether the two groups differ in the variances and means of the obligation recognition scores.

With the PROSE-disciplinary instrument, the result of the test comparing between the previous partial scalar model (in which the factor variance in the college instructors had been freely estimated) and the constrained model (in which the factor variance is constrained to be equal to one across the groups) suggested that the constrained model is not significantly worse than the partial scalar invariance model (DIFFTEST χ^2 (1) = 0.550, $p = 0.458$). Similarly, the model constraining the factor means to be equal to zero for both groups was not significantly worse than the partial scalar invariance model as well as the model constraining the factor variances to be equal to one across the groups (DIFFTEST χ^2 (1) = 1.631, $p = 0.202$). These results suggest that the two groups are not significantly different in mean and variance in the degree to which participants recognize the disciplinary obligation.

In the tests for the PROSE-institutional instrument, the result of testing the equality of the factor variance between the groups suggested a significant decrease in model fit compared to the previous partial scalar invariance model (DIFFTEST χ^2 (1) = 6.024, $p = 0.014$). Specifically, the factor variance of college instructors' recognition of the institutional obligation was significantly lower than that of high school teachers (fixed 1 for high school teachers; 0.652 for college instructors). Also, the factor mean of college instructors' recognition of the institutional obligation was significantly higher than that of high school teachers (fixed 0 for high school; 0.902 for college instructors, SE = 0.101, $p < 0.001$). In other words, there was less variability among the college instructors regarding the degree of recognition of the institutional obligation than among high school teachers, whereas college instructors more strongly recognized the institutional obligation than the high school teachers.

We compared the variance and mean of the factor representing the interpersonal obligation using the model that had achieved partial invariant properties regarding the structure, factor loadings, and thresholds with the PROSE-interpersonal instrument. The factor invariance model suggested no significant difference in the factor variance between the groups (DIFFTEST χ^2 (1) = 0.177, $p = 0.674$). In contrast, there was a significant difference in factor means, meaning that college instructors recognize the interpersonal obligation significantly less than high school teachers (fixed 0 for high school; −0.25 for college instructors, SE = 0.104, $p = 0.018$). The result suggested no significant differences in factor variances and significantly less recognition score for college instructors.

Table 2 Summary of the comparisons of factor mean and variance between high school teachers and college instructors

Instrument	Factor mean	Factor variance
Disciplinary	No sig. difference	No sig. difference
Institutional	HS teachers < College instructors	HS teachers > College instructors
Interpersonal	HS teachers > College instructors	No sig. difference
Individual	HS teachers > College instructors	HS teachers < College instructors

With the PROSE-individual instrument, the comparison test showed that there is a significant difference in factor variance between the groups (DIFFTEST χ^2 (1) = 6.281, $p = 0.012$). Specifically, college instructors were more varied in the degree of recognition of the individual obligation than high school teachers (fixed 1 for high school teachers, 1.514 college instructors, $p < 0.001$), meaning that high school teachers are more similarly disposed to depart from planned instructional actions to attend to the individual obligation than college instructors. The factor mean between the groups was also shown to be significantly different, indicating that college instructors had significantly lower scores in the degree of recognition of the individual obligation than high school teachers (fixed 0 for high school; −0.994 for college instructors, SE = 0.139, $p < 0.001$). Table 2 presents the summary of comparisons in the means and variances of the four factors between high school teachers and college instructors.

In sum, the comparison analyses suggested that college instructors recognized the institutional obligation more than high school teachers, whereas they recognized the obligation toward individual students or interpersonal relationships among students to a smaller degree than high school teachers. In the next section, we discuss the findings based on the differences between the contexts of high school and college mathematics instruction.

Relationships between different obligation scores

In the previous sections, we confirmed the unidimensionality of each instrument designed to measure recognition of each obligation and identified some differences among the four obligations in the patterns of difference regarding the factor mean and variance between the groups. The similarities and differences in these patterns across different obligations motivated us to examine the correlations among the four obligation scores. Each obligation score was estimated from each of the four unidimensional models that had been shown to be invariant across groups in the previous sections. The correlations between the exported obligation scores for each of the samples (the correlations for high school teachers and college instructors are below and above the diagonal, respectively) are presented in Table 3.

As shown in Table 3, three obligation scores (disciplinary, interpersonal, and individual) were positively related to each other and the size of these correlations ranged from moderate to large in both groups. This result of significant associations among the scores is aligned with our hypothesis that the scores commonly reflect the extent to which teachers recognize their professional obligations as sources of justification for actions deviating from what is customary in instruction. Interestingly, however, the score of institutional obligation was observed to be negatively related to other obligation scores in both groups. This result and the previous result showing contrasted patterns of the institutional

Table 3 Correlations between instructors' obligation scores

	Disciplinary	Institutional	Interpersonal	Individual
Disciplinary	–	−0.06	0.58***	0.37***
Institutional	−0.13*	–	−0.01	−0.21**
Interpersonal	0.60***	−0.08	–	0.37***
Individual	0.48***	−0.11*	0.44***	–

*$p<0.05$, **$p<0.01$, ***$p<0.001$

The correlations for high school teachers and college instructors are below and above the diagonal, respectively

obligation compared to other obligations may imply that the nature of the institutional obligation is somewhat different from that of other obligations. We discuss this in the next section.

Discussion

In this study, we conducted a series of multiple-group item factor analysis to ensure the comparability of the latent factor means and variances for recognition of each obligation between high school teachers and college instructors. After establishing partial measurement invariance between the groups, we compared the extent to which the two groups of instructors recognize each of the four obligations in terms of the degree to which instructors agreed that a teacher was justified to depart from their lesson plan on account of a given obligation. The comparisons of factor means and factor variances (i.e., structural invariance tests) conducted for each of the four obligations suggested several differences in the mean level and variance of the recognition of the obligations between the two groups.

First, high school teachers showed significantly higher levels of recognition of the interpersonal and individual obligations than college instructors. A possible explanation for this might be that college instructors are disposed to see students as adults who can care for themselves as individuals or their own interpersonal relations, thus they may feel less obligated to manage students' individual or interpersonal issues than high school teachers do. Another possible explanation for this is the amount of class time college instructors typically have with individual students in comparison with the time high school teachers spend with their students, which is especially important when one considers that our college instructors responded to items contextualized in lower division undergraduate classes, which are usually large. As described in Crisp et al. (2009), in large college classes it might be unrealistic to expect instructors to care about individual students or interpersonal relationships among students. It is therefore likely that the instructional environment in college leads college instructors to recognize their obligation in managing students' interpersonal or individual issues less than high school teachers do.

Second, in the comparison of factor variance between the groups, there was a greater variance in the level of college instructors' recognition of individual obligation than high school teachers. Motivated by Lande and Mesa's (2016) finding of an association between the instructional decisions made by community college mathematics faculty and their faculty status in their institutions, we further examined whether the variance in college instructors' recognition could be explained by their professional position in their institutions. Specifically, we predicted the level of college instructors' individual obligation by a binary variable indicating whether they are graduate student instructors ($N=105$) or non-graduate student instructors including postdoctoral researchers, part-time lecturers, non-tenure track faculty, or tenure-track professors ($N=128$). The regression analysis showed that graduate student instructors' recognition of the individual obligation is significantly higher than that of non-graduate student instructors ($\beta=0.167$. SE $=0.078$, $p=0.033$). A possible explanation for this might be that graduate student instructors have more chances to interact with individual students and may feel closer to being students themselves than to being members of the community of mathematics faculty instructors (Shultz et al., 2019). As graduate students, they may identify more with the experiences of students as individuals as they were recently, or still see themselves, in the same position. Another reason for the greater variance in the recognition of the individual obligation for college instructors may have something to do with the lack of common pedagogical training or professional development required for college instructors. In other words, college instructors are more likely to vary in their knowledge about students and of the different needs and goals of students compared to high school teachers who are required to complete education courses and professional development where they learn about cognitive, emotional, and social needs of students.

Third, one unanticipated finding was that college instructors had higher levels of recognition of the institutional obligation than high school teachers. This result may be explained by the fact that the college instructors who participated in this study were mostly from large research institutions where multiple sessions are offered for the same lower division undergraduate courses used as context for the items. In such college courses, there are often requirements from the mathematics department around details like the course timeline, textbook used, and exam questions. Those requirements may lead college instructors to feel obliged to make instructional decisions aligned with the larger institutional needs of the course. This explanation is consistent with that of Hora and Ferrare (2013) who conducted interviews with faculty teaching in mathematics and science at research universities. Some faculty members in the study expressed that they felt obliged to follow existing course syllabi and did not have the right to change the course structure or reading lists (p.235). As alluded to by Hora and Ferrare (2013), the institutional environment that may constrain (or encourage) the range of possible practices available to instructors needs to be considered to understand the decisions instructors make. We could say that, in our study, the potential effect of institutions like research universities on instructors' decisions in a classroom was observed through instructors' recognition of their obligation to institutions.

In contrast to the other three obligations, no significant difference was identified in the mean recognition of the disciplinary obligation between the two groups of instructors. A possible explanation for this result might be related to the fact that both groups of instructors have had extensive training in mathematics as a discipline. Thus, the obligation to the discipline of mathematics might be easily recognizable for both groups of instructors as basic responsibilities of mathematics instructors, so a difference in the recognition between the groups was not significant. It is also possible that our items do not sufficiently discriminate among individuals with high recognition of the obligation. To detect a difference in the level of disciplinary obligation between the groups, more difficult items (i.e., items that present scenarios in which the instructor's action is not easily justifiable on account of the obligation to the discipline of mathematics) may need to be included for a future study. An item similar to A114x that instantiates the disciplinary obligation in the case of elaborating on the mathematical theory could be an example of such items. Herbst and Chazan (2020) also show examples of how work on tasks chosen for a particular instructional goal may enable the instructor to observe students' engagement in valuable aspects of mathematical practice unrelated to the instructional goal that might justify momentary departures from the goal; these events could also be used as context for more discriminatory items.

Lastly, the institutional obligation showed negative relationships with other obligations. This result corroborates qualitative findings that instructors tend to talk about their obligation to the institution differently from the other three. They initially hesitate to acknowledge that the institution plays any role in their teaching (Shultz, 2018), and do not feel like it is an obligation because they position themselves as part of the institution (Shultz et al., 2019). However, given more probing in the same dataset, Shultz (2018) found that the instructors also revealed that the institution guides many of their decisions, such as those surrounding scheduling and topic coverage. For future studies, it would be valuable to complement the responses elicited with the PROSE instrument with responses to another instrument in which they report what they actually do when confronted with institutional demands in their own classroom.

The combination of findings provides support for the conceptual premise that the four hypothesized domains of obligations are different and they are recognized differently by mathematics instructors depending on the institutions where their work of teaching takes place. In turn, this helps suggest that the work of teaching mathematics is different across these institutional contexts, inasmuch as instructors of mathematics in those contexts recognize some of the same obligations but in significantly different ways.

Limitations and research implications

In this study, measurement invariance tests were conducted for the unidimensional models established for each of the four obligation instruments. In other words, the partial measurement invariance retained in this study is limited to each single, separated obligation domain (or stakeholder) and it does not account for relationships among the four obligations in the tests of measurement invariance. Further work will require a larger sample of high school and college instructors to establish the viability of comparing four obligations simultaneously using a four-dimensional model. Although we had to test measurement invariance for each obligation domain separately due to the sample size constraint, our alternative analysis examining the correlations among the four obligation scores within each group showed that the correlations were very similar between the two groups. Specifically, the correlations among the different factors were distinct enough from each other. Therefore, we expect that the results would not differ significantly from the results in a multidimensional model.

Multidimensional models also need to be considered in a future study for each obligation which might be better representable with more than one dimension. More investigation on multiple aspects of each obligation would help us to establish a greater degree of validity on the obligation measures applicable across different teacher populations.

Given the possibility that unequal sample size (471 high school teachers and 239 college instructors) can reduce the power of the tests (Kaplan & George, 1995), we may need to be able to replicate the result of no significant difference in the mean of disciplinary obligation between the two groups in a future study with a larger sample of college instructors to solidify the result. Although there is no exact criterion for the ratio between the groups, the difference in sample size in our study is not considered to be a significant concern, based on the study that examined the effects of group size differences on the results of factorial invariance tests (Yoon & Lai, 2018).

While the factor mean comparisons were based on partial invariance models, the direction of the latent construct comparisons (which group has higher or lower degree of obligation recognition) were consistent regardless of whether the threshold constraints of non-invariant items are released or not. Also, the results were consistent with different criteria used in evaluating model fit (using change in CFI or chi-square DIFFTEST). However, for the individual obligation, we had to release more than a half of the total item constraints, so the factor mean comparison was conducted only with two invariant items. There is, therefore, a definite need to have more items probing instructors' recognition of the individual obligation and examine whether it is possible to establish at least partial measurement invariance with enough number of items to reliably measure the construct.

Conclusion

The findings of this study contribute to the understanding of the professional position of mathematics teachers according to the schooling level where they teach. While mathematics instruction at different education levels has some common characteristics, our data shows that there are differences across levels. This understanding contributes to explain instructional actions as more than an expression of individual resources: Indeed, the institutional position of instructors comes along with obligations to stakeholders to which instructors need to adapt. The fact that differences are observed in individual recognition scores across institutional environments suggests that more than individual differences account for these scores. While the items required participants to indicate their agreement with decisions made by a virtual instructor, these findings may also help us conjecture how the environment impacts teacher decision-making. Shultz (2020, 2022) has investigated, in particular, how the recognition of obligations mattered in instructors' decisions to use inquiry-based learning in lower division college mathematics courses.

This study also contributes in several ways to the use of scenario-based surveys for studies measuring constructs enacted in classroom contexts (Herbst & Chazan, 2015). By way of using scenario-based surveys, we ask questions about specific scenarios with attention to instructional context and, at the same time, we provide the scalability of a targeted construct, which is possible through the use of surveys. The scenarios depicted as storyboards with nondescript cartoon characters allow adjusting the instructional contexts to be aligned with high school or college mathematics classrooms while maintaining the commonality in the characteristics of the target constructs. Our results of partial measurement invariance between the groups support the plausibility of the use of scenario-based instruments in comparing the degree of recognition of obligations by two different groups of instructors whose teaching is taking place in different institutional contexts. Further, analogous work investigating differences in recognition of obligations across high, middle, and elementary school mathematics can help further document how mathematics teaching varies across levels of instruction.

Appendix A

See Tables 4, 5, 6 and 7.

Table 4 Disciplinary obligation: Mean and estimated standardized factor loading of each item

Items	Statement	Mean		Factor loading (SE)	
		HS	C	HS	C
A101L	The teacher should assign students a problem similar to what they have been doing in class, rather than have the students continue working on the real-world problem	5.08	4.89	0.57(0.04)	0.56(0.05)
A102x	The teacher should keep to what is in the textbook, rather than require students to use information that is different from what is in the textbook	4.85	4.75	0.58(0.04)	0.68(0.04)
A104x	The teacher should thank the student for contributing then ask the class how to do the problem originally posed, rather than build on a non-standard method	4.22	4.13	0.59(0.03)	0.53(0.05)
A105x	The teacher should move to the next topic, rather than elaborate on details of a topic that are outside the scope of the course	4.39	4.34	0.6(0.03)	0.56(0.05)
A105L	The teacher should keep the definition as originally written on the board, rather than adjust the definition in order to make it more general	3.78	3.78	0.38(0.04)	0.49(0.05)
A106x	The teacher should emphasize the method they are learning in class, rather than encourage the use of an alternative method	4.36	4.38	0.49(0.04)	0.51(0.05)
A107L	The teacher should confirm that the student's method is appropriate, rather than ask the student to consider if the method would work in all cases	4.28	4.22	0.54(0.03)	0.47(0.06)
A108x	The teacher should help check the work that was finished, rather than pay attention to the idea from a student who had not yet finished	4.66	4.74	0.53(0.04)	0.57(0.05)
A109x	The teacher should conform to the terms that are in the textbook, rather than promote the use of a different term	4.55	4.32	0.46(0.04)	0.63(0.05)
A111x	The teacher should stick to the mathematics at hand, rather than take class time to make connections to other mathematical ideas	4.01	4.55	0.47(0.04)	0.41(0.06)
A112x	The teacher should give students problems with specific numbers to practice, rather than ask students to produce a generalization	4.98	4.94	0.69(0.03)	0.68(0.04)
A114x	The teacher should give students additional practice problem, rather than elaborate on mathematical theory	4.04	4.62	0.71(0.03)	0.72(0.04)

Table 5 Institutional obligation: Mean and estimated standardized factor loading of each item

Items	Statement	Mean		Factor loading (SE)	
		HS	C	HS	C
A201L	The teacher should go over the topics in the set of review materials, rather than ask the students to read them on their own	3.70	4.08	0.35(0.04)	0.33(0.07)
A202x	HS: The teacher should answer all of the student questions, rather than stick to those in the curriculum	3.07	3.63	0.35(0.04)	0.42(0.06)
	College: The teacher should answer all of the student questions, rather than stick to those in the syllabus				
A202L	The teacher should take time to answer the students' questions, rather than move faster through the material so that the class catches up with the other sections	1.94	2.61	0.74(0.04)	0.53(0.06)
A204L	The teacher should use the remaining days to provide detailed explanations for as much of the new materials as time permits, rather than briefly summarize the remaining topics	2.82	3.30	0.54(0.04)	0.42(0.07)
A205L	The teacher should continue to discuss the problem as long as students have questions about it, rather than dismiss those questions because the problem will not appear on the test	2.50	2.73	0.59(0.04)	0.49(0.06)
A211x	The teacher should answer the students' questions, rather than ignore it to keep pace with the other classes	2.27	3.07	0.68(0.03)	0.67(0.07)
A217x	HS: The teacher should wait to test students because they haven't spent enough time on the topic rather than give them a test in order to assign them grades by the school's deadline	2.24	3.12	0.41(0.04)	0.31(0.06)
	College: The teacher should wait to test students until they have spent enough time on the topic rather than give them a test in order to give the exam at the same time as the other sections				

Table 6 Interpersonal obligation: Mean and estimated standardized factor loading of each item

Items	Statement	Mean		Factor loading (SE)	
		HS	C	HS	C
A301x	The teacher should have stopped the interruptions earlier, rather than let the argument go on for so long	3.80	3.98	0.37(0.05)	0.44(0.06)
A303x	The teacher should answer the student's question, rather than have him share it with the whole class	4.10	3.87	0.65(0.04)	0.83(0.04)
A303L	The teacher should move on to another problem or activity, rather than insist that students share their work with a partner	4.54	4.24	0.47(0.04)	0.48(0.06)
A304L	The teacher should allow the students that are done with their classwork to move on to the homework for the evening, rather than encourage them to check their answers with each other	4.20	4.14	0.53(0.04)	0.44(0.06)
A305x	The teacher should review the other problems on the board, rather than get more students to talk about Kappa's solution	3.99	3.66	0.57(0.04)	0.7(0.04)
A305L	The teacher should ask a student who used a traditional method to explain the solution, rather than ask the student at the board to repeat the explanation for the class to understand the alternative method	4.31	4.29	0.63(0.04)	0.56(0.05)
A311x	The teacher should call on someone who has their hand in the air, rather than call on someone who has not participated yet	4.35	3.58	0.57(0.04)	0.49(0.06)

Table 7 Individual obligation: Mean and estimated standardized factor loading of each item

Items	Statement	Mean		Standardized estimated	
		HS	C	HS	C
A402L	The teacher should help the student understand the context of the original problem, rather than change the context of the problem for that student	3.60	3.36	0.34(0.05)	0.49(0.06)
A403L	The teacher should grade all quizzes in the same way, rather than offer to grade one student's quiz on the basis of what the student was able to finish	2.69	1.75	0.45(0.05)	0.56(0.07)
A405L	The teacher should ask the student to focus on the work at hand, rather than adjust lessons to engage a particular student	3.68	3.00	0.53(0.05)	0.62(0.06)
A409x	The teacher should acknowledge the student and continue with the lesson rather than allow the student to come to the front and demonstrate his method for the rest of the class	3.66	2.79	0.64(0.04)	0.64(0.06)
A415x	The teacher should stick with the original topic rather than follow up on the aside brought up by the student	4.11	3.31	0.63(0.04)	0.59(0.06)

Appendix B

See Table 8.

Table 8 Internal consistency of each instrument

Obligation	Number of items retained	High school		College	
		Cronbach's Alpha	Avg. inter-item correlation	Cronbach's Alpha	Avg. inter-item correlation
Disciplinary	12	0.80	0.25	0.82	0.27
Institutional	7	0.66	0.22	0.60	0.18
Interpersonal	7	0.71	0.26	0.74	0.29
Individual	5	0.61	0.24	0.66	0.28

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511. https://doi.org/10.1177/0022487109348479

Bieda, K. N., Sela, H., & Chazan, D. (2015). You Are Learning Well My Dear: Shifts in novice teachers' talk about teaching during their internship. *Journal of Teacher Education, 66*(2), 150–169. https://doi.org/10.1177/0022487114560645

Blair, R., Kirkman, E. E., & Maxwell, J. W. (2018). Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2015 CBMS survey. *American Mathematical Society*. https://doi.org/10.1090/cbmssurvey/2015

Bourdieu, P. (1990). *The logic of practice*. Stanford University Press.

Briggs, A. R. J., Clark, J., & Hall, I. (2012). Building bridges: Understanding student transition to university. *Quality in Higher Education, 18*(1), 3–21. https://doi.org/10.1080/13538322.2011.614468

Brousseau, G. (1997). *Theory of didactical situations in mathematics: Didactique des Mathematiques 1970–1990* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Trans.). Dordrecht, The Netherlands: Kluwer.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456. https://doi.org/10.1037/0033-2909.105.3.456

Chazan, D., Herbst, P., & Clark, L. (2016). Research on the teaching of mathematics: A call to theorize the role of society and schooling in mathematics. In D. Gitomer & C. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1039–1097). AERA.

Cheng, M., Barnes, G.P., Edwards, C., Valyrakis, M., Corduneanu, R., & Koukou, M. (2015). *Transition skills and strategies: Transition models and how students experience change*. Glasgow, Scotland: Quality Assurance Agency for Higher Education Retrieved from https://www.enhancementthemes.ac.uk/docs/ethemes/student-transitions/transition-models-and-how-students-experience-change.pdf

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119–142. https://doi.org/10.3102/01623737025002119

Crisp, G., Palmer, E., Turnbull, D., Nettelbeck, T., Ward, L., LeCouteur, A., Sarris, A., Strelan, P., & Schneider, L. (2009). First year student expectations: Results from a university-wide student survey. *Journal of University Teaching and Learning Practice, 6*(1), 13–26. Retrieved from http://ro.uow.edu.au/jutlp/vol6/iss1/3

Golbeck, A. L., Barr, T. H., & Rose, C. A. (2018). Fall 2016 departmental profile report: Annual survey of the mathematical sciences in the US. *Notices of the AMS, 65*(8), 952–962.

Herbst, P., Dimmel, J., Erickson, A., Ko, I., & Kosko, K. (2014). Mathematics teachers' recognition of an obligation to the discipline and its role in the justification of instructional actions. In C. Nicol, P. Liljedahl, S. Oesterle, & D. Allen (Eds.), *Proceedings of the 2014 annual meeting of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 273–280). Vancouver, Canada: Simon Fraser University.

Herbst, P., & Ko, I. (2018). *Recognition of professional obligations of mathematics teaching and their role in justifying instructional actions*. Paper presented at the Annual Meeting of AERA, New York, NY.

Herbst, P., & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *Zdm–mathematics Education, 44*(5), 601–612. https://doi.org/10.1007/s11858-012-0438-6

Herbst, P., & Chazan, D. (2015). Studying professional knowledge use in practice using multimedia scenarios delivered online. *International Journal of Research and Method in Education, 38*(3), 272–287. https://doi.org/10.1080/1743727X.2015.1025742

Herbst, P., & Chazan, D. (2020). Mathematics teaching has its own imperatives: Mathematical practice and the work of mathematics instruction. *Zdm–mathematics Education, 52*, 1149–1162. https://doi.org/10.1007/s11858-020-01157-7

Herbst, P., Chazan, D., Kosko, K., Dimmel, J., & Erickson, A. (2016). Using multimedia questionnaires to study influences on the decisions mathematics teachers make in instructional situations. *Zdm–mathematics Education, 48*, 167–183. https://doi.org/10.1007/s11858-015-0727-y

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K., Hollingsworth, H., Jacobs, J., & Stigler, J. (2003). *Teaching mathematics in seven countries: TIMSS 1999 video study.* Washington, DC: National Center for Educational Statistics. Retrieved from http://timssvideo.com/timss-video-study

Hiebert, J., & Stigler, J. W. (2017). Teaching versus teachers as a lever for change: Comparing a Japanese and a U.S. perspective on improving instruction. *Educational Researcher, 46*(4), 169–176. https://doi.org/10.3102/0013189x17711899

Hora, M. T., & Ferrare, J. J. (2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. *Journal of the Learning Sciences, 22*(2), 212–257. https://doi.org/10.1080/10508406.2012.729767

Kaplan, D., & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 2*(2), 101–118. https://doi.org/10.1080/10705519509539999

Kosko, K. W., & Gao, Y. (2016). The tale of two teachers' use of prompts in mathematical discussions. *Issues and Ideas in Education, 4*(2), 111–130. https://doi.org/10.15415/iie.2016.42009

Lande, E., & Mesa, V. (2016). Instructional decision making and agency of community college mathematics faculty. *Zdm–mathematics Education, 48*(1–2), 199–212. https://doi.org/10.1007/s11858-015-0736-x

Leatham, K. R. (2006). Viewing mathematics teachers' beliefs as sensible systems. *Journal of Mathematics Teacher Education, 9*(1), 91–102. https://doi.org/10.1007/s10857-006-9006-8

Loewenthal, K. M. (2001). *An introduction to psychological tests and scales* (2nd ed.). Psychology Press.

Marston, S. H. (2010). Why do they teach? A comparison of elementary, high school, and college teachers. *Education, 131*(2), 437–454.

Mesa, V. (2014). Using community college students' understanding of a trigonometric statement to study their instructors' practical rationality in teaching. *Journal of Mathematics Education, 7*(2), 95–107.

Muthén, L. K., & Muthén, B. O. (1998−2015). Mplus user's guide.

Perry, C., & Allard, A. (2003). Making the connections: Transition experiences for first-year education students. The *Journal of Educational Enquiry, 4* (2).

Rasmussen, C., & Ellis, J. (2015). Calculus coordination at PhD-granting universities: More than just using the same syllabus, textbook, and final exam. In D. Bressoud, V. Mesa, & C. Rasmussen (Eds.), *Insights and recommendations from the MAA National Study of College Calculus* (pp. 107–115). Mathematical Association of America.

Rasmussen, C., & Kwon, O. N. (2007). An inquiry-oriented approach to undergraduate mathematics. *The Journal of Mathematical Behavior, 26*(3), 189–194. https://doi.org/10.1016/j.jmathb.2007.10.001

Shultz, M. (2018). *Stories of agency: Do graduate students perceive themselves as part of the mathematical community?* Unpublished manuscript. Ann Arbor: University of Michigan.

Shultz, M., & Herbst, P. (2021). *How do the social resources for instructional decision-making differ for graduate students and faculty in mathematics?* Unpublished manuscript.

Shultz, M. (2020). *The rationality of college mathematics instructors: The choice to use inquiry-oriented instruction (Doctoral dissertation)*. University of Michigan.

Shultz, M. C. (2022). The rationality of undergraduate mathematics instructors: The choice to use inquiry-oriented instructional practices. *Journal for Research in Mathematics Education, 53*(3), 227–246.

Shultz, M., Herbst, P., & Schleppegrell, M. (2019). The expression of agency by graduate teaching assistants and professors in relation to their professional obligations. *Linguistics and Education, 52*, 33–43. https://doi.org/10.1016/j.linged.2019.05.006

Sofroniou, A., & Poutos, K. (2016). Investigating the effectiveness of group work in mathematics. *Education Sciences, 6*(4), 30. https://doi.org/10.3390/educsci6030030

Sonnert, G., Sadler, P. M., Sadler, S. M., & Bressoud, D. M. (2015). The impact of instructor pedagogy on college calculus students' attitude toward mathematics. *International Journal of Mathematical Education in Science and Technology, 46*(3), 370–387. https://doi.org/10.1080/0020739X.2014.979898

Speer, N. M., Smith, J. P., & Horvath, A. (2010). Collegiate mathematics teaching: An unexamined practice. *The Journal of Mathematical Behavior, 29*(2), 99–114. https://doi.org/10.1016/j.jmathb.2010.02.001

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. Free Press.

Upcraft, M. L., & Kramer, G. L. (1995). *First-year academic advising: patterns in the present, pathways to the future*. Monograph Series Number 18.

Webel, C., & Platt, D. (2015). The role of professional obligations in working to change one's teaching practices. *Teaching and Teacher Education, 47*, 204–217. https://doi.org/10.1016/j.tate.2015.01.007

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859