MATHEMATICS TEACHERS' RECOGNITION OF AN OBLIGATION TO THE DISCIPLINE AND ITS ROLE IN THE JUSTIFICATION OF INSTRUCTIONAL ACTIONS

Patricio Herbst[1], Justin Dimmel[1], Ander Erickson[1], Inah Ko[1], Karl W. Kosko[2]

[1]University of Michigan[1], [2]Kent State University

We describe the conceptualization, development, and piloting of two instruments--a survey and a scenario-based assessment--designed to assess, teachers' recognition of an obligation to the discipline of mathematics and the extent to which teachers justify actions that deviate from what's normative on account of this obligation. We show how we have used classical test theory and item response theory to select items for the instruments and we provide information on their reliability, using a sample of 88 high school mathematics teachers.

FOCUS: THE DISCIPLINARY OBLIGATION

This paper reports on efforts to conceptualize and measure teachers' recognition of an obligation to the discipline of mathematics and contributes to an agenda for research that attempts to identify sources of justification for actions in mathematics teaching. This agenda is predicated on the need to have robust ways of predicting how efforts at instructional improvement might fare as they are implemented. The general problem is, given an instructional system located in an institutional context, where, by force of custom, teacher and students are expected to act in ways that are normative, what sources of justification are available for practitioners to use so as to justify, for themselves and colleagues, actions that might depart from the norm?

In their practical rationality framework, Herbst and Chazan (2012) proposed the notion of *professional obligations* as a set of those sources of justification. They identify four obligations --to the discipline of mathematics, to students as individuals, to the class as a social group, and to the institutions where instruction is located (e.g., school, district). This report elaborates on the disciplinary obligation and presents results of our attempts to develop two instruments designed to study it empirically.

THEORETICAL FRAMEWORK

The problem of why teachers do what they do in classrooms has often been studied using perspectives that consider instructional action as dependent on factors ascribed to the individual teacher (e.g., beliefs, goals, knowledge) and disconnected from considerations of environment (cultural, historical, or institutional; Cooney, 1984;

Schoenfeld, 2010). Often this work has revealed that teachers' perceptions of environmental conditions account for mismatches between what individual teachers might profess to want to do and what they might acknowledge to be able to do in practice (Skott, 2009). Understanding these environmental conditions in which mathematics teachers work is an important terrain for our field still to cover.

Important progress has been made in the last 20 years to conceptualize and study mathematics instruction as an interaction among teacher, content, and students in environments (Cohen et al., 2003). Our field shows plenty of examples of how teacher and students collaboratively shape meanings as they undertake the work of teaching and learning mathematical ideas (Arzarello et al., 2009). Analyses of classrooms as activity systems have helped document the notion that classroom interaction often relies on tacit norms that regulate how teacher and students customarily exchange knowledge (Bauersfeld, 1980; Herbst, 2006).

International studies of mathematics teaching have added attention to the situatedness of instruction in larger systems, in particular national cultures (Stigler & Hiebert, 1999) but one could just as well say historical periods and societal institutions. This scholarship suggest the need to examine in more detail how demands of the environment might affect mathematics instruction, with the hope that this understanding might help explain teachers' instructional actions and decisions.

The discipline of mathematics is an important element in the environment of mathematics instruction in all countries, but it is plausible that it might affect instruction in different ways. In their account of the practical rationality of mathematics teaching, Herbst and Chazan (2012) identify an obligation to the discipline as a source of justification for decisions and actions. They define this obligation in general by saying that "the mathematical knowledge teachers teach needs to be a valid representation of the mathematical knowledge, practices, and applications of the discipline of mathematics" (p. 610). This obligation to the discipline is a reasonable hypothesis that can be traced back to Schwab's (1978) writings on the curriculum or to the heavy investment of mathematicians in the reforms of the 50s and 60s (Kilpatrick, 2012). Research also documents how teachers' views on instructional action, what they consider appropriate or inappropriate to do, are often grounded on disciplinary considerations (Ball, 1993; Lampert, 1990). We could accept as a hypothesis that this obligation affects all teachers of mathematics and still expect this obligation to affect teachers differently. In this paper we offer a conceptualization of those possible differences and we share details of the development of two instruments designed to study those differences.

TEACHERS AND THE OBLIGATION TO THE DISCIPLINE

The discipline of mathematics exercises its role as stakeholder of instruction in various ways. Quite often policy considerations of the state of mathematics education incorporate the views of mathematicians (Becker & Jacob, 2000). Mathematicians are involved in the professional development of teachers and in decisions over the

curriculum for teacher education (see also Wilson, 2003; Ball et al., 2005). Questions can be asked about this influence.

How much and in what ways do teachers recognize an obligation to the discipline? In our earlier analyses of teacher discussions prompted by representations of practice we inspected the rationale that teachers gave for endorsing or opposing actions that deviated from an instructional norm. Among those rationales, participants would make various kinds of references to the discipline: They would draw on the need to show how mathematicians really work, on the need to avoid making unwarranted assumptions, or on the value of writing an elegant proof. The discipline was a salient source of justification, though not the only one (Nachlieli & Herbst, 2009). We undertook a two-pronged approach for the development of research instruments that could help us eventually understand teachers' relationship to the disciplinary obligation. On the one hand we set out to develop a survey that would allow us to gauge the extent to which an individual teacher recognizes an obligation to the discipline. On the other hand we set out to develop a scenario-based questionnaire that would allow us to gauge the extent to which an individual teacher would justify deviating from actions that are normative in instruction on account of an obligation to the discipline. With both instruments our goal was to be able to eventually implement them at scale, so we aimed for final products that could be answered by individuals working on a computer alone and for less than an hour.

MEASURING RECOGNITION OF THE DISCIPLINARY OBLIGATION: THE PR-OB-MATH QUESTIONNAIRE

We have laid out the first steps in investigating recognition of mathematics teachers obligation to the discipline by developing a questionnaire that asks participants to consider statements about mathematics teaching (e.g., "Mathematics teachers do their best to get students to appreciate mathematical elegance") and then asks them to "rate the degree to which mathematics teachers are expected, as professional educators, to act in the manner this statement describes" using a 4-point Likert-type of scale that ranges from (1 = Teachers are always expected to act in this manner to 4 = Teachers are never expected to act in this manner). This instrument, unlike our scenario-based instrument described below, is meant to be used with teachers of mathematics at different levels and nonteachers alike, all of them being asked to indicate their stances toward statements that say what a teacher of mathematics is purportedly expected to do. We developed the survey through several iterations that included brainstorming, item writing, internal and external vetting, piloting with teachers, and examining the collected pilot data using classical test theory (CTT; Crocker & Algina, 1986) and item response theory (IRT; Bond & Fox, 2007).

We started the design process with two versions of the questionnaire, one (ETD, "expected to do") roughly similar to the final one described above and another one (ATS, "appropriateness to say") that included the target statement (e.g., "Mathematics teachers should do their best to get students to appreciate mathematical elegance") in

quotation marks and asked the respondent to rate, using a 6-point Likert-type scale ranging from (1=Very Inappropriate to 6=Very Appropriate), the appropriateness of making such a statement to a fellow mathematics teacher in a teachers' lounge. Following internal review, we vetted our initial set of items through a process of cognitive interviews with secondary mathematics teachers (Karabenick et al., 2007). Initial interviews suggested that teachers did not always interpret the ATS statements as we had intended. Some statements were perceived as inappropriate to say to a colleague but not because of being objectionable actions, but rather because they were too obvious and saying them would insult a colleague's intelligence. We then introduced other contexts where those statements could be made and vetted both the contexts and the statements with additional teachers. These interviews revealed a need to adjust the social context where such disciplinary statements were made (to that of a mentor teacher speaking with a student teacher), assess the validity of items, and revise or discard items. This resulted in a set of 10 items for each ATS and ETD list. We piloted those items with mathematics teachers from a Midwestern U.S. state (n = 44) and found them to have low internal consistency (α = .49). Efforts to improve reliability via item analysis were not fully successful. In particular we decided to discontinue the ATS items and write more ETD items making sure to anchor statements to emblems of mathematical work that were familiar to teachers. This yielded a list of 26 items.

We piloted the 26 items with a sample of 42 high school mathematics teachers from the Midwest during the Summer of 2013. All statements were rated on a 4-point Likert-type scale in increasing degree of obligation (from 1=Never, to 4=Always). During the piloting of these items we discovered a few difficulties with items, including some modulated (should) statements mixed with descriptive statements, and statements that, along with the rating scale, might yield readings that included double and even triple negatives (e.g., "When introducing a new concept to students, mathematics teachers should not give descriptions that are mathematically imprecise" was not only modulated by "should" but also would become a double negative if participants responded "never"). We rewrote the statements so that they would all be descriptive and that their readings would yield at most one negative (e.g., the statement above became "When teaching students a new property, mathematics teachers ensure that it is described precisely"). This last version of the 26 items was piloted with 46 high school teachers from the Midwest, during the Fall of 2013. Table 1 shows descriptive statistics of the 26 items for both samples.

For the analysis, first, we conducted classical item analysis (looking at the item-total correlations and the changes in alpha coefficient after removal of an item) to remove problematic items among the 26 original items within each of the samples (Summer and Fall 2013). While alpha values (0.756 for Summer and 0.757 for Fall) were acceptable, some items had negative or very low positive item-total correlations. We eliminated 8 items that did not meet a .3 threshold of item-total correlation and as a

result reduced the item set to 18 items. This increased the alpha score of the remaining items to 0.804 and 0.799 for Summer and Fall samples respectively.

We inspected the data set with the goal of running 1-parameter IRT model with the pooled Summer and Fall 2013 samples. Since there had been slight variations in the statement of the items, we inspected first whether the items were functionally equivalent using a DIF analysis on the remaining 18 items. To meet assumptions of DIF and Rasch analysis, we recoded responses from the 4-point scale to dichotomous, using responses 1-2 as 0 and 3-4 as 1. This recoding appeared legitimate given than none of the values of the scale expressed a neutral stance. The DIF analysis showed that 3 of the 18 items functioned very differently in both samples, so we excluded them from the Rasch analysis. (Dorans et al, 1992).

	All 26 items	Selected 13 items
Summer (n=42)	M = 2.96 (.30) α = 0.76	M= 3.00 (.40) α = 0.76
Fall (n=46)	M = 3.02 (.26) α = .75	M= 3.02 (.38) α = 0.76

Table 1: Descriptive statistics for items of the PR-OB-MATH instrument

We fit a Rasch model to the pooled samples data for the remaining 15 items and inspected the fit statistics from the Rasch model analyses, excluding 2 more items that, according to Bond and Fox (2007), had poor fit. Thus, the original 26 items could be reduced to 13 items after removing problematic items from iterative item analyses. The selected 13 items were also examined using a Rasch model. The Rasch model with the final selected 13 items shows sufficient item reliability (0.95), but low person reliability (0.52), lower than 0.80 considered acceptable. This means that our items distinguish easier and difficult items well but our items may not be sensitive enough to distinguish between high and low scorers. Table 1 below lists descriptive statistics of our samples with the initial 26 items and the final 13 items.

THE DISCIPLINARY OBLIGATION'S ROLE IN JUSTIFYING ACTIONS

To investigate the extent to which teachers' recognition of the disciplinary obligation matters in the justification of instructional actions, we developed a scenario-based questionnaire. While the PR-OB-MATH instrument provides a way of assessing the extent to which a mathematics teacher recognizes an obligation to the discipline of mathematics, the role this recognition plays in practical action and decision-making is not apparent: While somebody might recognize an obligation to some extent, the impact of such recognition on action might also depend on practical circumstances such as what they might be expected to do. A *situational judgment test* (Cabrera & Nguyen, 2001) or a scenario-based assessment would give us a chance to explore that question. These assessments have a long history in human resources management, where they are presented as a written vignette or a video. Video-based tests of situational judgment are widely used by personnel departments under the presumption

that a more realistic scenario will result in responses that reflect what candidates will actually do (Weekley & Jones, 2006). Scenarios have also been used to explore teacher decision-making and attitudes for many of the same reasons (Bishop & Whitfield, 1972; Shavelson et al., 1977). Carter et al. (1988) actually presented both novice and expert teachers with slides with visual images from classrooms and used this to compare how experience influenced their descriptions.

To assess the possible impact of recognition of the disciplinary obligation in action, we created items in which the participants view a teaching scenario, represented as a storyboard, and are asked to choose between two courses of action, one considered normative and another that deviates from the former in response to the disciplinary obligation. The introduction to each item would say "In the following slideshow we invite you to consider a scenario in which a high school teacher deviates from a lesson in order to address an issue of mathematical importance. We are interested in the extent to which you think the teacher's action is justifiable." After considering the scenario, participants are asked to indicate "how much you agree or disagree with the following statement:" and given a statement of the form "The teacher should [do what was hypothesized as normative], rather than [do what the teacher had done in the scenario]." To rate their agreement participants are given a 6-point Likert scale ranging from 1 = Strongly Disagree to 6 = Strongly Agree.

We specified 15 such items including scenarios such as providing a definition different than the one given in the textbook, letting a student pursue the consequences of a faulty assumption, modifying the usual format of a task to engage students in a mathematical practice, etc. Because participants had to respond to scenarios that they could relate to, we specified each of the 15 items in general but designed scenarios that adapted that general specification to particularities of instruction in Early elementary (grades K-2), upper elementary (3-5), middle school (6-8), or high school (9-12). (This paper reports high school teachers' data only.) As a rule these scenarios were realized using a set of cartoon characters and the *Depict* software tool that allows us to create storyboards using cartoon characters and speech bubbles. The scenarios were then embedded in a questionnaire created and administered in the LessonSketch platform (www.lessonsketch.org).

After internal review and edition, we convened a focus group including experienced teachers and individuals with strong mathematics background to check whether our hypothesized normative actions were seen as normative by members of the profession and whether the deviations from those normative actions were seen as attending to an obligation to the discipline. After incorporating the group's feedback, the items were piloted with the same groups of high school mathematics teachers described above in Summer 2013 (n=42) and Fall 2013 (n=46). Since items were exactly the same and participants come from the same geographic pool we pooled the samples. In order to fit a 1-parameter IRT model to this data, we recoded responses from a 6-point Likert scale to a dichotomous scale, using responses 1-3 as 0 and 4-6 as 1. The IRT analysis showed good item reliability (.935) and a good range of possible theta scores for participants

(-4.71 to 4.73), indicating that the items, as a set, discriminate between participants that have more or less of the latent trait being measured. In this case, that latent trait is recognition that obligation to the discipline of mathematics can justify actions in a mathematics classroom.

As noted above, while the two instruments, the PR-OB-MATH and the Justifications of Actions scenario-based assessment examine teachers' relationship to the disciplinary obligation, they operationalize different conceptualizations of it and they involve the participants in different activities. It does make sense nevertheless to ask whether and how scores in one instrument are related to scores in the other. We found however no significant correlation between these scores and no significant correlation between either of those scores and years of mathematics teaching experience.

SIGNIFICANCE AND CONCLUSION

We have made significant progress toward validating two instruments that can help operationalize the notion of professional obligation, which contributes to understanding the rationality behind the work of mathematics teaching. The importance and usefulness of this work goes beyond increasing capacity to describe, explain, and predict instruction; it can also contribute to the development of a professional discourse for mathematics teaching. Indeed, mathematics teachers are professionals but the discourse on which they can justify their actions sits uncomfortably between the individual knowledge and preferences of practitioners and the general discourses of academic disciplines such as mathematics or psychology: The teaching profession can use the development of a shared professional discourse that can better support their practical work. Better understanding how professional obligations impact what teachers deem appropriate to do can help in the long run develop a shared professional discourse of justification.

References

Arzarello, F., Paola, D., Robutti, O., & Sabena, C. (2009). Gestures as semiotic resources in the mathematics classroom. *Educational Studies in Mathematics*, *70*(2), 97-109.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, *93*(4), 373-397.

Ball, D., Ferrini-Mundy, J., Kilpatrick, J., Milgram, R. J., Schmid, W., & Schaar, R. (2005). Reaching for common ground in K-12 math education. *AMS Notices*, *52*(9), 1055-1058.

Bauersfeld, H. (1980). Hidden dimensions in the so-called reality of a mathematics classroom. *Educational Studies in Mathematics*, *11*(1), 23-41.

Becker, J. P., & Jacob, B. (2000). The politics of California school mathematics: The anti-reform of 1997-99. *Phi Delta Kappan*, *81*(7), 529-537.

Bishop, A. J., & Whitfield, R. C. (1972). *Situations in teaching*. New York: McGraw-Hill.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model*. London: Psychology Press.

Cabrera, M., & Nguyen, N. (2001). Situational judgment tests: A review of practice and constructs assessed. *Intl. Journal of Selection and Assessment*, *9*(1-2), 103-113.

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, *39*(3), 25-31.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119-142.

Cooney, T. J. (1985). A beginning teacher's view of problem solving. *Journal for Research in Mathematics Education*, *16*(5), 324-336.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. NY: Holt.

Dorans, N. J., Schmitt, A., & Bleistein, C. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, *29*, 309-319.

Herbst, P. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education*, *37*, 313-347.

Herbst, P., & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM*, *44*(5), 601-612.

Karabenick, S. A., Woolley, M., Friedel, J., Ammon, B., Blazeski, J., Bonney, C., … Kelly, K. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139-151.

Kilpatrick, J. (2012). The new math as an international phenomenon. *ZDM*, *44*(4), 563-571.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *Amer. Educational Research Journal*, *27*(1), 29-63.

Nachlieli, T., Herbst, P., & González, G. (2009). Seeing a colleague encourage a student to make an assumption while proving. *Journal for Research in Mathematics Education*, *40*(4), 427-459.

Schoenfeld, A. H. (2010). *How we think: A theory of goal-oriented decision making and its educational applications*. New York: Taylor & Francis.

Schwab, J. J. (1978). Education and the structure of the disciplines. In I. Westbury & N. Wilkof (Eds.), *Science, curriculum, and liberal education* (pp. 229-272). Chicago: University of Chicago Press.

Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. *Amer. Ed. Research Journal, 14*(2), 83-97.

Skott, J. (2009). Contextualising the notion of 'belief enactment'. *Journal of Mathematics Teacher Education*, *12*(1), 27-46.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap*. New York: Free Press.

Weekley, J. A., & Jones, C. (2006). Video-based situational testing. *Personnel Psychology*, *50*(1), 25-49.

Weiss, M., Herbst, P., & Chen, C. (2009). Teachers' perspectives on "authentic mathematics" and the two-column proof form. *Educational Studies in Mathematics*, *70*(3), 275-293.

Wilson, S. M. (2003). *California dreaming: Reforming mathematics education*. New Haven, CT: Yale.