Measuring Mathematical Knowledge for Teaching College Algebra at Community Colleges

Vilma Mesa
University of Michigan

Irene Duranczyk
University of Minnesota

Kimberly M. Wingert
RMC Corporation

The VMQI Research Team[1]

We present preliminary findings regarding the dimensionality of a 34-item instrument designed to measure the mathematical knowledge used in teaching college algebra at community colleges and the performance of items within the instrument. The instrument assumed a six-dimension model with two organizers of knowledge, one related to knowledge needed to perform two specific tasks of teaching, choosing problems *and* understanding student work, *and the other related to knowledge of three function types,* linear, exponential, and rational functions*. Multidimensional item response theory models were applied to a sample of 416 community college mathematics instructors. A three-dimension model structured by function types better fitted the data than a unidimensional model. Two- and six-dimensional models, structured by the tasks of teaching or the combination of function types and tasks of teaching, did not converge so they are not discussed. We discuss implications and work to further validate the instrument.*

Keywords: mathematical knowledge for teaching, college algebra, community colleges

Our project (Mesa et al. 2020) sought to develop and validate an instrument to measure community college faculty knowledge of teaching college algebra (MKT-CCA). Theoretical work on teacher knowledge has hypothesized distinct components of such knowledge (e.g., pedagogical, curricular, and content, Shulman, 1986) but empirical support has been elusive (e.g., Copur-Genturk et al., 2019). While there is a consensus that such knowledge should be multidimensional, and multiple efforts that further specify content or pedagogical knowledge according to cognitive level, knowledge type, or mathematical topic have flourished (Ball et al., 2008 [MKT]; König et al., 2011 [COACTIV]; Krauss et al., 2008 [TEDS-M]; Saderholm et al., 2010 [DTAMS]), the question about multidimensionality remains open (Schilling, 2007).

Using an approach that focuses on content-specific teacher decision making (e.g., in high school algebra or geometry), Herbst and colleagues (e.g., Herbst & Chazan, 2012; Ko & Herbst, 2020; Herbst & Ko, 2019) have empirically shown that teachers' mathematical knowledge used in two tasks of teaching are distinguishable via two distinct (albeit correlated) measures for knowledge needed to choose givens for a problem and understanding students' work in geometry. They surmise that the main challenge previous studies had in developing multiple measures is related to their knowledge-type-based framework given that there is no clear boundary among the types (e.g., common and specialized content knowledge) and that teachers need to use multiple dimensions of knowledge simultaneously in the moments of teaching. Grounding their work on the notion of instructional exchanges, they seek to understand, from the teacher's perspective, how knowledge is "transacted" in classrooms. Herbst (2006) notes that in their role as teachers, they are expected to provide students with activities or assignments that

[1] The VMQI Team: Vilma Mesa, Claire Boeck, Inah Ko, University of Michigan; Patrick Kimani, April Ström, Laura Watkins, Maricopa Community Colleges; Irene Duranczyk, Bismark Akoto, Siyad Gedi, Dexter Lim, University of Minnesota; Mary Beisiegel, Oregon State University

would support the development of mathematical ideas and that as students produce work in response to those activities and assignments, teachers must interpret the work to ascertain whether the student has learned or not. These two activities, choosing problems and understanding students' work, require teachers to sort through their own mathematical knowledge to make decisions that are guided by the goal of the instructional situation at stake.

We took this perspective in the design of an instrument that would assess community college instructors' knowledge of college algebra topics for teaching (Mesa et al., 2023). The core tasks of teaching used in this study are choosing problems and understanding student work. We defined *choosing problems* as a task of teaching that requires instructors to select activities or assignments that represent mathematical ideas accurately and give students opportunity to work on a mathematical idea at stake. Further, we defined *understanding student work* as a task of teaching that requires instructors to make sense of the ideas that students produce, either orally or in writing, and evaluate their mathematical correctness. Each task demands knowledge specific to teaching mathematics. Choosing problems requires knowledge about mathematical ideas presented in the problem, and about how to assess if ideas in the problem align with an instructional goal. For understanding student work, an instructor needs to have knowledge about interpreting students' work, typical errors students make, various strategies for solving a problem and how they could be represented, and the mathematically correct answer. We grounded these two tasks of teaching in three different types of functions, linear, rational, and exponential functions, as they are foundational to college algebra and for further work in mathematics. In this paper, we document the process of validating the structure of the instrument, seeking to identify whether there are indeed six distinct dimensions of mathematical knowledge for teaching, through the hypothesized organization along two tasks of teaching and the three function types (see Figure 1).

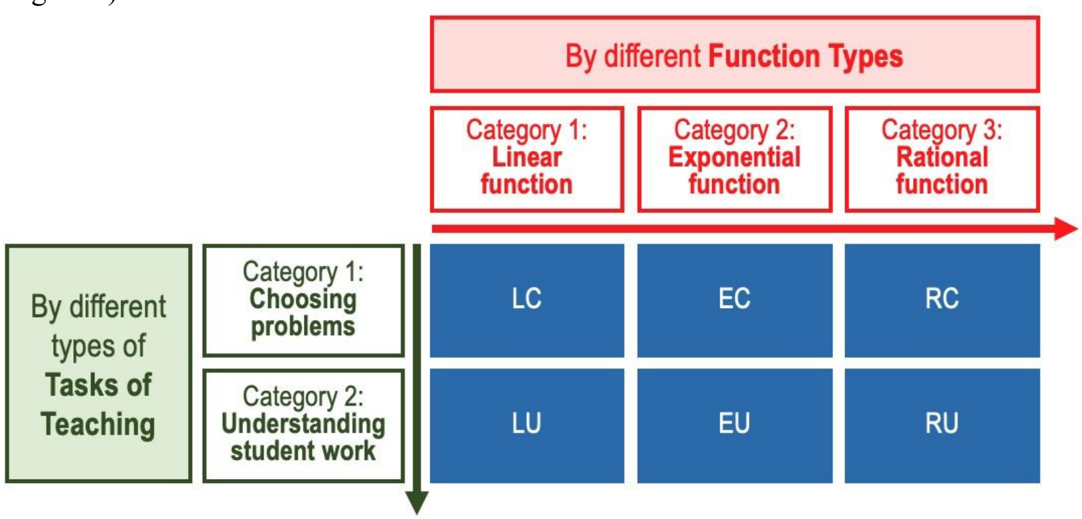

Figure 1. Blueprint for the development of the MKT-CCA instrument. (Mesa et al., 2023)

Methods

The instrument was developed over a two-year period that included an item development camp (13 practitioners, 11 researchers, 240 items), cognitive interviews to improve the items (12 instructors, 36 items), solicitation of situations and student work (15 instructors, 12 items), expert review of the mathematical content (2 mathematicians, 36 items), piloting (120 instructors, 60 items), and further revision of the items (see Figure 2).

Mr. Trevena is creating a lesson about applications of exponential functions. Which of the following news headlines would be the most likely to provoke a discussion about exponential growth patterns? **Choose one option only.**

A. Census data show that the population of the city tripled from 1970 to 1980 then doubled from 1980 to 2000.

B. Prices have increased 7% over the past 5 years, while usage of public transportation has decreased 5% over the same period.

C. Twice as many single people ages 20-25 are living with their parents today as they were 20 years ago and economists predict this trend will continue.

D. The water level of the local river has increased by 13% in the past decade and 35% during the past 30 years.

(a)

Ms. Ucci's class has just finished discussing the notation of $f(t)=a(b)^{rt}$ and the meaning of parameters and variables for exponential functions. To assess her students' understanding she gave them this task:

A radioactive substance decays according to the function $A(t)=100(1/2)^{(⅕)t}$ where t is measured in years and $A(t)$ denotes the amount of the unstable portion of the substance in micrograms at t. What is the meaning of ⅕ in the function?

Which of the following students' responses demonstrates a correct understanding for the meaning of ⅕? **Choose one option only.**

A. ⅕ means that the substance needs to decrease four more times to get to half of its original amount.

B. ⅕ means that ⅕ of the substance *A* is left after half of a year.

C. ⅕ indicates that it will take five years for the substance *A* to reach its half life.

D. ⅕ means that ⅕ of the substance A has decayed in half of a year.

(b)

Figure 2. Sample items in exponential function addressing (a) choosing problems and (b) understanding student work.

The final instrument included 34 items (27 multiple-choice items (MC) and seven testlets (T), resulting in 55 different questions) that were balanced across the six hypothesized dimensions (see Figure 3).

		Function type (#MC, #T)			Total
		Linear	Rational	Exponential	
Task of teaching	Choosing Problems	3, 2	5, 1	3, 3	11, 6
	Understanding Student Work	5, 1	5, 0	6, 0	16, 1
Total		8, 3	10, 1	9, 3	27, 7

Figure 3. Total number of multiple-choice items (MC) and testlets (T) per dimension in the final instrument.

We invited a sample of 4,350 instructors teaching from 752 different institutions to complete the instrument. The sample was constructed to ensure representation of the diversity of the colleges in terms of size, location, and diversity of the student population. The data were collected between October 2022 and March 2023. We obtained responses from 954 participants. Data were excluded if (a) the participant's position as an instructor at a community college could not be verified (n=185), (b) it was a duplicate response (n=85), (c) the participant entered their assigned ID but did not complete any portion of the survey (n=91), (d) the participant indicated that they had never taught college algebra (n=17), or (e) the participant responded to the 34 items in less than 1,200 seconds (n=20). There were 556 teachers who responded to some portion of the instrument after these data exclusions were applied. Final analyses focused on the 416 teachers who responded to all 55 questions on the MKT-CCA instrument. Accuracy on each question was submitted to a 2PL multiple item response theory (MIRT) model using maximum likelihood estimation in Mplus (Muthén & Muthén, 1998-2023). The dimensionality of the instrument was tested using likelihood ratio tests comparing alternative models. The six-dimensional model, where each factor represents one task of teaching and one function type, and two-dimensional models, where each factor represents one task of teaching, did not converge. The three-dimensional model included correlated factors, with each factor represented by a set of function-type items (linear, exponential, and rational). In the one-dimensional model all items were loaded onto one factor. Items were considered for deletion when the item did not provide much information and had a low discrimination estimate (less than 0.61, indicating low ability to differentiate teachers).

Results

The sample included faculty from 260 different community colleges in 43 states with 50% enrolling a majority of non-White students; 47% of the colleges were located in a city, about a quarter (26%) in a suburb, and the rest in rural (12%) or small towns (14%); most of the colleges were in the West (37%) or South (36%), with 19% located in the Midwest, and 9% in the Northeast region of the United States. Forty-eight percent of the participants identified as male, and 46% as female; in terms of race, 76% identified as White, 10% as Asian, 4% as Black, 2% as mixed, and 4% chose Other. Seventy-eight percent of the participants held full-time positions; 32% had a temporary appointment, and 9% were on tenure track. The average number of years of teaching experience was close to 17 years (mean=16.76, SD = 9 years; range: 1.5 to 47 years). The majority (63%) held a master's degree in mathematics, mathematics education or another mathematics related field, and 12% held PhDs (about 5% were in Mathematics Education.)

Discrimination and difficulty within the item response theory (IRT) paradigm are indices of item performance. Discrimination gives a measure of the differential capacity of an item, that is the ability of the item to differentiate participants by their ability level. The discrimination is the slope of the item response function graph (i.e., the item characteristic curve); the higher values suggest that the item has a high ability to differentiate examinees. Values of discrimination are reported as positive numbers (i.e., a negative value would be concerning because the probability of getting the answer correct should not decrease as ability increases). Difficulty is defined as the ability level at which we would expect examinees to have a probability of 0.50 (assuming no guessing) of selecting the correct response to the item. The higher the difficulty is, the higher the ability required from an examinee to have a 50% chance of answering an item correctly. Values of item difficulty typically range from –3 to +3; items with difficulty close to –3 are very easy items; items with difficulty close to +3 are very difficult for the pool of examinees. The discrimination values from the 3-dimensional model across the whole set of 55 items ranged from 0.183 to 2.555, whereas the level of difficulty of the full set of items ranged from –4.914 (very easy) to 2.140 (difficult).

The three-dimensional model including all 55 MKT-CCA items fit the data significantly better than the one-dimensional model, $\chi^2(3) = 12.37$, $p < .01$. The Akaike's Information Criterion (AIC) and sample size adjusted Bayesian Information Criterion (BIC) also favored the three-dimensional model. Four items contributing to the linear latent factor, seven items contributing to the exponential latent factor, and six items contributing to the rational latent factor were identified for deletion. The discrimination for all items considered for deletion was less than .61. After removing these 17 items, indicators representing the accuracy on the remaining 38 items were submitted to a final 2PL multidimensional IRT(MIRT) model using maximum likelihood estimation. As before, the three-dimensional model fit the data significantly better than the one-dimensional model, $\chi^2(3) = 9.93$, $p = .02$. The AIC and sample size adjusted BIC also favored the three-dimensional model. All item discrimination parameters were greater than .61 in this final model on the reduced set of items. Examination of the total information curve for each dimension (see Figure 4) revealed that the MKT-CCA instrument provided good coverage of a wide range of latent ability levels, but the test provided the most information about teachers with lower mathematical teaching knowledge. A total of 15 out of 16 of the linear items had negative difficulty parameters (below average MKT-CCA is needed to have a 0.5 probability of getting the item correct), 12 out of 14 of the exponential items had negative difficulty parameters, and all the rational items had negative difficulty parameters. The discrimination across the retained 38 items ranged from 0.612 to 2.472, whereas difficulty level ranged from -3.465 (very easy) to 2.181 (difficult).

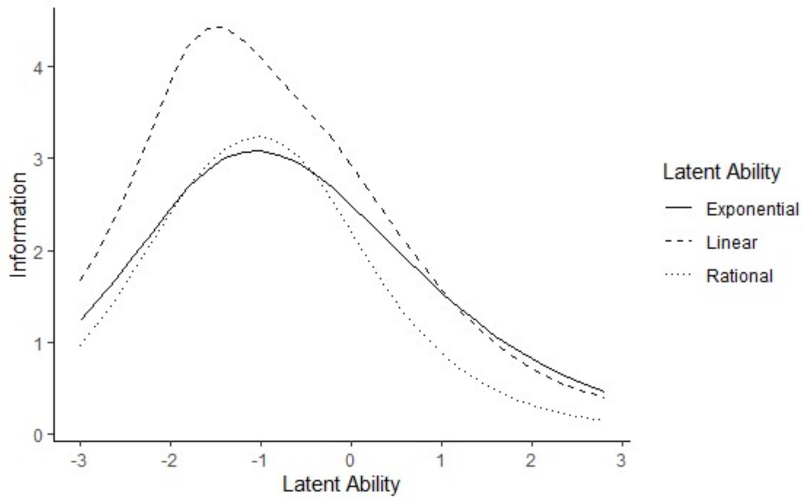

Figure 4: Total information curves for the three dimensions of the final model.

Discussion

The six-dimensional model did not converge, potentially because the sample size was insufficient. While the three-dimensional model fit the data better, we still need to further examine whether the three dimensions do really correspond to distinct function types, given that our hypothesized two-dimensional model was by tasks of teaching, and not about whether the types of functions were distinct. In other words, we need to rule out that the items that are purportedly designed to assess a specific function type, assess that function type and not others.

A two-dimensional model by tasks of teaching did not converge; in other words, it was not possible to differentiate knowledge along the two dimensions associated with choosing problems or understanding student work. We believe that this might be due to how we operationalized the choosing problems dimension, relative to the work by Ko and Herbst (2020) who focused exclusively on choosing givens (set of givens that had enough information for students to solve the problem) for problems in geometry. Such specificity might have contributed to make a clear distinction possible in Ko and Herbst's context. In our study, choosing problems is a task of teaching that requires instructors to select activities or assignments that include much more than specific givens, and thus our broader definition of a task may cause multidimensionality within the choosing problems dimension: that is in crafting items about choosing problems, we may have triggered knowledge related to understanding students' work.

As is apparent in Figure 4, we found that, overall, the instrument was not very difficult even though the instrument allows us to discriminate between high and low scorers. Because of resource limitations, and the need to reach faculty in multiple institutions in the United States, we administered the instrument online (instead of in person), which might have contributed to the test not being as difficult, as we believe that participants might have used resources such as textbooks or online applications to respond to the items, even though this was discouraged. Despite this, we believe that the instrument could be a promising tool to examine community college faculty knowledge for teaching these types of functions and design content-focused faculty professional development.

Next Steps

Our next steps in the validation process include conducting cognitive interviews with participants and testing the model with a population without teaching experience. With the cognitive interviews, we plan to validate our hypotheses about how instructors with or without the hypothesized knowledge would answer a randomly selected set of items. In doing so, we plan to corroborate that instructors who answer the item correctly are doing so because they have the needed knowledge, and that when an item is answered incorrectly, is because the instructor does not possess such knowledge. In randomly selecting the items, we will be able to establish the generalizability of the instrument. Moreover, cognitive interviews will allow us to shed light regarding possible ways in which the two tasks of teaching we hypothesized were intertwined in ways that instructors could not call on one or the other task of teaching independently of each other.

To determine whether there is something special about the instrument as related to community college mathematics faculty, we administered the test to a sample of 100 undergraduate students, who would be knowledgeable of the mathematics, but who did not have any teaching or tutoring experience. We are in the process of analyzing the data to identify whether the samples behave similarly or not. We anticipate major differences in the modeling, which would further confirm that the instrument is useful to assess knowledge for teaching these types of functions at community colleges.

As part of the data collection process, we gathered background information about faculty and their teaching experiences, their history of courses taught, their educational background, engagement in professional development, and their beliefs about mathematics, its teaching, and its learning. Further analysis of these data will shed light about how these characteristics relate to performance on the instrument by different groups of instructors, which would corroborate prior findings relating associations between courses community college faculty teach and performance on instruments assessing teacher knowledge (Ko et al., in press). Such analysis will provide further validation for the dimensionality, difficulty, and discrimination of the instrument.

Acknowledgments

Funding provided by the National Science Foundation EHR Core Awards #2000602 (Michigan), #2000644 (Maricopa), #2000527 (Minnesota), and #2000566 (Oregon). Opinions, findings, and conclusions or recommendations expressed do not necessarily reflect the views of the Foundation. Special thanks to Siyad Gedi, the Research on Teaching Mathematics in Undergraduate Settings research lab, and to the Undergraduate Research Opportunity Program at the University of Michigan for their support to this work.

References

Ball, D. L., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389-407.

Copur-Gencturk, Y., Tolar, T., Jacobson, E., & Fan, W. (2019). An empirical study of the dimensionality of the Mathematical Knowledge for Teaching construct. *Journal of Teacher Education, 70*(5), 485-497. https://doi.org/10.1177/0022487118761860

Herbst, P. (2006). Engaging students in proving: A double bind on the teacher. *Journal for Research in Mathematics Education, 33*, 176-203.

Herbst, P., & Chazan, D. (2012). On the instructional triangle and the sources of justifications for the actions of the mathematics teacher. *ZDM Mathematics Education*, *44*, 601-612.

Herbst, P., & Ko, I. (2019, April). *Reconsidering the organization of mathematical knowledge for teaching*. Paper presented at the Annual conference of the American Educational Research Association, Toronto, Canada.

Ko, I., & Herbst, P. (2020). Subject matter knowledge of geometry needed in tasks of teaching: Relationship to prior geometry teaching experience. *Journal for Research in Mathematics Education*, *51*(5), 600-630.

Ko, I., Mesa, V., Duranczyk, I., Herbst, P., Kohli, N., Ström, A., & Watkins, L. (in press). Understanding the characteristics of mathematical content knowledge for teaching algebra in high schools and community colleges. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739X.2021.2006348

König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F.-J. (2011). General pedagogical knowledge of future middle school teachers: On the complex ecology of teacher education in the United States, Germany, and Taiwan. *Journal of Teacher Education, 62*(2), 188-201. https://doi.org/10.1177/0022487110388664

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM*, 40(5), 873-892.

Mesa, V., Ko, I., Boeck, C., Duranczyk, I., Akoto, B., Lim, D., Kimani, P., Watkins, L., Ström, A., & Beisiegel, M. (2023). *A blueprint for measuring mathematical knowledge for teaching college algebra at community colleges*. 13th Congress of the European Society for Research in Mathematics Education, Budapest, Hungary.

Mesa, V., Watkins, L., Duranczyk, I., Beisiegel, M., Kimani, P., & Ström, A. (2020). *Algebra Instruction at Community Colleges: Validating Measures of Quality Instruction (AI@CC 2.0)* [Grant]. University of Michigan, Maricopa Community Colleges, University of Minnesota, Oregon State University.

Muthén, L. K., & Muthén, B. (1998-2023). *Mplus user's guide: Statistical analysis with latent variables, user's guide*. Muthén & Muthén.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) instrument. *School Science and Mathematics, 110*(4), 180-192. https://doi.org/10.1111/j.1949-8594.2010.00021.x

Schilling, S. G. (2007). The role of psychometric modeling in test validation: An application of multidimensional item response theory. *Measurement, 5*(2-3), 93-106.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.