DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

The Mathematical Knowledge for Teaching Community College Algebra Instrument

Vilma Mesa[1] – Inah Ko[1] – Irene Duranczyk[2] – the VMQI Research Team[3]

[1] University of Michigan

[2] University of Minnesota

[3] April Ström, Chandler-Gilbert Community College; Patrick Kimani and Laura Watkins, Glendale Community College; Mary Beisiegel and Camden Collins, Oregon State University; Claire Boeck, Rabia Kirikçilar, Inah Ko, Vilma Mesa, and Melissa Valerie, University of Michigan; Bismark Akoto, Irene Duranczyk, and Dexter Lim, University of Minnesota.

ORCID: 0000-0003-0460-1850

Address: 610 East University SEB 3119, Ann Arbor MI 48109-1259

email: vmesa@umich.edu

Abstract

We describe the theorization and validation of an instrument to measure mathematical knowledge for teaching college algebra at community colleges, through five different studies that use different data sources: (1) expert review of the items, (2) cognitive interviews with community college instructors, analysis of psychometric properties of data from (3) community college instructors and from (4) mathematics majors with no teaching experience, and (5) associations of the scores with teaching experience and frequency of engaging in specific tasks of teaching. This work contributes to efforts in developing instruments that measure mathematical knowledge for teaching at community colleges, and to validate the interpretation of what these instruments measure. Future studies can assess the extent to which high and low performance on the instrument is related to other measures of quality instruction and with student performance, and to assess changes in knowledge for teaching mathematics after participation in instructional development programs with this population.

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

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

The Mathematical Knowledge for Teaching Community College Algebra Instrument

Research on teacher knowledge has fundamentally changed how we perceive the complexity of teaching mathematics and amplified the close connection between such knowledge and the specific mathematical ideas and meanings that are negotiated in the classroom. Simultaneously, the field of mathematics education has been developing expertise in measurement and psychometrics which has made it possible to develop instruments that can reliably measure intended abstract constructs such as learning of algebra (Wladis et al., 2018) and mathematical knowledge for teaching in elementary grades (Phelps et al., 2014) and in geometry (Ko & Herbst, 2020). Researchers have been working to corroborate empirically that such knowledge is multidimensional, because the differentiation could support focused professional development programs for teachers. Some research groups (e.g., LMT, COACTIV, TEDS-M, DTAMS) have tried to distinguish dimensions of mathematical knowledge for teaching—either content or pedagogical—with mixed success (Ball et al., 2008; König et al., 2011; Krauss et al., 2008; Saderholm et al., 2010). Even if it is not possible to clearly distinguish all distinct components of such knowledge, a reliable measure that can be used to make valid inferences about the amount of mathematical knowledge for teaching an instructor may possess can provide support for assessing the effectiveness of a professional development program, especially if the program addresses the theory supporting the construction of the instrument.

To identify areas of need for professional development for community college faculty and to subsequently establish impacts of such programs, we have developed an instrument that measures the mathematical knowledge for teaching college algebra courses at community colleges, MKT-CCA, which is complementary to another instrument we developed that assesses the quality of instruction using videos of mathematics classrooms, the *Evaluating the Quality of*

Instruction in Postsecondary Mathematics instrument (EQIPM, Mesa et al., 2020). In this paper, we summarize the development of the MKT-CCA instrument and put forward an argument that the data gathered from this instrument can be used to measure community college mathematics instructors' knowledge for teaching college algebra.

Approach to Validation

Progress in quantifying mathematical knowledge for teaching has been enhanced by the research in validation methods. Current work indicates that validation and validity are not believed to be properties of an instrument but of the interpretation and use of the scores obtained with the instrument. The AERA's *Standards for Educational and Psychological Testing* (2014) defined instrument validation as "a process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use" (p. 11). Specifically, Kane (2006) proposed an argument-based validation framework which requires: (1) "specifying the proposed interpretations and uses of the scores, and (2) evaluating the overall plausibility of those interpretations and uses" (Kane, 2012, p. 4). The Standards document (AERA, 2014) described the process of validation as accumulating evidence to evaluate the soundness of this intended interpretation of test scores for the proposed use. The Standards document lists five types of evidence as the key sources that illustrate different aspects of validity:

1. *Evidence Based on Test Content* includes "logical or empirical analyses of the adequacy with which the test content represents the content domain and of the relevance of the content domain to the proposed interpretation of test scores" (p. 14).

2. *Evidence Based Internal Structure* includes analyses that indicate "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (p. 16). These analyses will shed light on the conceptualized dimensions or components of the construct.
3. *Evidence Based on Response Processes* includes descriptions of the "fit between the construct and the detailed nature of the performance or response actually engaged in by test takers" (p. 15).
4. *Evidence Based on Relations to Other Variables* includes analyses of the degree to which "the construct should be related to some other variables" and how those relationships are consistent with the construct underlying the proposed test score interpretation (p. 16).
5. *Evidence Based on Consequences of Testing* includes assessment of "the soundness of the proposed interpretations for their intended uses" (p. 19).

The standards note that not every type of evidence is required in all contexts, but that evidence should be collected to support each proposition underlying the intended interpretation and use of test scores.

In this paper, we present the evidence that the instrument that we developed, the MKT-CCA, measures mathematical knowledge for teaching college algebra at community colleges, by presenting findings from four interconnected studies, each providing the needed evidence for our argument. We present an abridged literature review on work done on mathematical knowledge for teaching, and this is followed by two sections that provide evidence based on the content of the instrument: its theoretical underpinnings and the context chosen to investigate mathematical knowledge for teaching as a multidimensional construct. This is followed by five studies that address evidence based on the internal structure of the instrument, in response processes, and in

relations to other variables. Currently, we do not have evidence regarding consequences of testing, but we do anticipate problematic uses of the instrument. We conclude by providing information about the elements of the Interpretation and Use Statements framework proposed by Carney et al. (2022).

Investigating *Mathematical Knowledge for Teaching* as a Multidimensional Construct in Community Colleges and College Algebra

The work on mathematical knowledge for teaching (MKT) beyond K-12 education has evolved significantly since 2000. Initially, the focus of studying mathematical knowledge for teaching was primarily within K-12 education, with research and professional development efforts aimed at improving the mathematical content knowledge and pedagogical skills of elementary and secondary school teachers for the improvement of student performance (Ball *et al.*, 2008; Hill et al., 2007; Hill et al., 2005; Hill et al., 2004). However, the recognition of the MKT construct as sensitive to the educational environment, student population and specific mathematics content knowledge lead to recent studies for specialized MKT at the college level (Speer et al., 2015). Speer *et al.* (2015) contributed significantly to this effort by examining how MKT constructs could be applied to both secondary and college mathematics teachers. Their work provided a foundation for understanding the specific knowledge and skills required for teaching mathematics at the college level. Nabb and Murawska (2019) highlight this shift, emphasizing unique challenges faced by community college instructors who teach developmental courses. Community college instructors require knowledge of mathematical concepts and teaching strategies that must be tailored to adult learners with diverse educational backgrounds and learning needs.

While scholars have assumed that mathematical knowledge for teaching is a multidimensional construct, finding clear empirical evidence for the hypothesized distinction between dimensions has been difficult (Charalambous et al., 2020). A central assumption in the design of instruments that assess mathematical teacher knowledge is that such knowledge can be elicited through the mathematical work embedded in activities that teachers do when they teach (Ball & Forzani, 2007), but the conceptualizations of such work has varied. Ball and colleagues' conceptualization (Ball *et al.*, 2008; Hill, Ball, et al., 2008; Hill, Blunk, et al., 2008) builds upon Shulman's (1986) distinction between subject matter and pedagogical content knowledge. Ball and colleagues hypothesized that the knowledge needed to teach mathematics could be further parsed out into two major components, subject matter knowledge and pedagogical content knowledge. They proposed that the subject matter knowledge is composed of two types of knowledge, the one that people would use in a variety of settings in work and life, called *common content knowledge*, and the knowledge that would be needed for teaching, called *specialized content knowledge*. On the other hand, pedagogical content knowledge was described as including three different types of knowledge, (a) knowledge that teachers would need about the content to teach as related to the students they teach, for example, knowledge of misconceptions or of learning trajectories, called *knowledge of content and students*; (b) knowledge of examples that can be used to illustrate an idea or that would be important to motivate a topic or to surface misconceptions and that would be necessary for teaching, called *knowledge of content and teaching*; and (c) knowledge of how the topics to teach are organized into a curricular sequence, in which grades or courses would students need what knowledge, etc., called *knowledge of curriculum*. To date, empirically showing that mathematical knowledge for teaching is indeed composed of all these dimensions has proven very difficult; the instruments that assess mathematical knowledge for teaching in

elementary school have found some evidence of multidimensionality (Hill *et al.*, 2004; Schilling, 2007). However, they typically established a unidimensional score representing overall mathematical knowledge for teaching in subsequent studies (Hill, 2010) rather than multiple sub-scores presenting different dimensions of the knowledge as the items did not clearly load onto hypothesized factors.

A different conceptualization of the knowledge needed in practice emerged from the work of decision making by Herbst and colleagues (Herbst & Chazan, 2012; Ko & Herbst, 2020). To understand teacher work from the teachers' perspective, they defined a teacher's role in instruction as managing instructional exchanges in classrooms. Here, "instructional exchanges" refer to how knowledge is transacted between students' work and the mathematical ideas at stake. In other words, to manage these exchanges, teachers are expected to provide students with activities or problems that support the development of the targeted mathematical ideas; students then produce work based on these activities, and teachers interpret this work to determine whether the students have mastered the intended knowledge (e.g., Herbst & Chazan, 2011). Carrying out these two interconnected activities—bridging the gap between the knowledge at stake and students' work—requires teachers to use their own mathematical knowledge to make appropriate instructional decisions. By focusing specifically on these instructional exchanges, Herbst and colleagues offer a different approach to interpret key tasks of teaching that any teacher, regardless of context or subject matter, can engage in. Ko and Herbst (2020) provided evidence that, at least in the context of teaching of high school geometry, it is possible to differentiate two distinct dimensions of the knowledge, one related to knowledge needed to choose givens for a geometry problem and another related to understanding students' work. They suggest that unclear boundaries among the dimensions of knowledge—based on a

knowledge-type framework (e.g., common and specialized content knowledge)—tend to create challenges in identifying distinguishable dimensions of teachers' mathematical knowledge, because teachers must use multiple dimensions simultaneously in the moment of teaching.

Community colleges are an ideal context to explore the multidimensionality of mathematical knowledge for teaching. While community colleges have multiple functions within the U.S. educational system in higher education (Mesa, 2017), they typically offer curriculum that would be equivalent to the first two-years of a college degree in other institutions. More importantly, they are institutions primarily devoted to teaching. Many instructors choose to be at a community college because they value teaching over research, appreciate the small class sizes, and can reach many students who otherwise would not be able to get a college education. Moreover, because of the different organization of the colleges, instructors tend to be called on to provide non-instructional student support not necessarily tied to classroom instruction (Aguilar-Smith & Gonzales, 2021; Finnegan, 2019; Kisker et al., 2023). Relative to other higher education institutions in the U.S., community college instructors teach more courses, on average, about five courses per semester, reaching many more students and accumulating substantial teaching experiences with diverse student populations. These experiences are related to the tasks of teaching we aim to investigate.

College Algebra is one of the first courses for which students earn college credit and is usually required to enter a calculus sequence. The course has a large subscription: In the year 2021, an estimated 388,993 students took a college algebra course at a two-year college (Blair & Kirkman, 2025, p. 66). Instructors who teach this course encounter a wide range of students' learning needs and therefore have multiple opportunities to think about how to teach material that tends to be considered basic and to interpret students' work in the process; thus, these instructors may call on

a particular expertise to be successful in teaching college algebra. The course typically includes topics that contribute to the development of notions of functions and change. The notion of function in college algebra typically builds upon rules and strategies related to symbolic manipulation of equations leading to the development of variables and the notions of domain, range, and change. Change primarily drives the calculus-based curriculum that is typically taught in the first two-years of many college degrees, including health, sciences, social sciences, mathematics, and engineering and technology related fields. Understanding linear growth and how it is used to approximate nonlinear growth of polynomials and other functions is fundamental to derivatives in calculus; differentiating this growth from exponential growth is quite fundamental to modeling situations in real-life context. Change can also be studied by looking at behavior of functions for very large values of inputs, and close to points of discontinuity. Change in the context of rational functions is ideal to investigate functional behavior of objects, requiring the manipulation of individual components (polynomials in the numerator and denominator) to make inferences about a new object and to investigate change close to points of discontinuity and at infinity. Measuring knowledge needed to teach linear, exponential, and rational functions, is advantageous to identify variations in instructors' knowledge because they can be explored in multiple mathematical registers, including verbal, numerical, algebraic, and graphical. But more importantly, these notions can be quite difficult for students to learn (Jurdak & El Mouhayar, 2014; Mayes, 1993; Nair, 2010; Sproesser et al., 2022; Thompson & Carlson, 2017; Wilkie & Ayalon, 2018), which demands instructors' knowledge, not only of the mathematical content involved but also of typical errors students make and of examples and problems that can support students in developing understanding of these ideas. Moreover, there is evidence that suggests that as instructors teach advanced mathematics courses,

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

they become more knowledgeable about core notions that must be fully addressed in earlier courses (Ko et al., 2024).

Theorization of the MKT-CCA

To develop our instrument, we proposed that mathematical knowledge for teaching could be organized along two dimensions: tasks of teaching and function types (Mesa et al., 2023). “Tasks of teaching” refer to all the separable components related to the work of teaching—for example, identifying content to teach, selecting examples to illustrate a principle, assigning activities for students to practice or demonstrate their understanding, and reading or listening to what students say to determine what they understand and figure out how to respond. Among these tasks, two that manage instructional exchanges are particularly representative: the task of choosing problems and the task of understanding student work. Choosing problems is “a task of selecting activities or problems that represent mathematical ideas accurately and give students the opportunity to work on a mathematical idea at stake” (p. 3846). Understanding student work involves making sense of the ideas that students produce—either orally or in writing—and evaluating their mathematical correctness (p. 3847). We claim that each of these teaching tasks demands knowledge specific to teaching mathematics and that these demands are distinct enough to differentiate the knowledge used in each task. For example, choosing problems requires an understanding of the mathematical ideas involved in the problem and the ability to assess whether those ideas align with an instructional goal; on the other hand, understanding student work requires knowledge about typical errors students make, strategies they may use to solve problems, the ways mathematical work is represented, in addition to the criteria for determining mathematical correctness of that work.

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

For the second dimension, which sought to encompass the mathematical content typically taught in college algebra, we chose three mathematical functions: linear, rational, and exponential. Not only are these functions core topics in typical college algebra course,[1] there exists a substantial body of research about student thinking, at all levels of schooling, which was foundational to inform the item development process.

Within linear functions, slope and rate of change are two of the most commonly documented topics that are difficult to understand in school algebra; in the context of Cartesian representations and in real world situations, students have major difficulties interpreting horizontal and vertical intercepts and the connection between a graph of a line and its symbolic representation (e.g., Ellis, 2007; Lobato & Siebert, 2002; Moschkovich et al., 1993; Postelnicu, 2011). Students also exhibit difficulties in understanding and interpreting proportional reasoning in this context (Van Dooren et al., 2003; Van Dooren et al., 2018). In general, connecting the rate of change of a function to the slope of the line representing that function is a major hurdle in the learning of linear functions (Postelnicu, 2011; Stump, 2001). Research on rational functions highlights the significant role of knowledge of asymptotic behavior for calculus notions related to function behavior at infinity or at points of discontinuity; knowledge of rational functions involves algebraic manipulation of expressions and an understanding of limits. College algebra content builds skills related to the algebraic operations on polynomials to identify key features of a rational function. Research on students' thinking about rational functions suggests that students

[1] A feature of the college algebra curriculum as taught at community colleges in the United States is that there is no standardized, or mandated set of topics that should be covered across states, districts, or colleges within a state--although there might be some exceptions. The mathematical content we selected to design items for the instrument encompass not just foundational notions for advanced mathematical work, but content that is covered in college algebra courses across the United States. Our analysis of student learning objectives documents from eight states (Arkansas, Colorado, Florida, Kansas, Missouri, New Mexico, Ohio, and the Lone Star System in Texas) resulted in 24 different topics, with only one, graphs, being common across all of them. The three topics chosen—linear, rational, and exponential functions—were mentioned in the student learning objectives in at least five of the documents noted.

believe that all rational functions are discontinuous somewhere, that asymptotes are invisible lines that are never touched by the function (Nair, 2010; Yerushalmi, 1997), that vertical asymptotes occur at points where a function is undefined, or that every function that has a variable in the denominator has a vertical asymptote. Some students assume that a rational function is undefined at the zeros of the numerator and are "unsure about the conditions under which a hole exists" (Nair, 2010, pp. 103-104). Research on exponential growth and decay points to the difficult nature of understanding the behavior of exponential functions, concretely, rates of change and the multiplicative unit needed to define the function (Confrey & Smith, 1994). Research with in-service teachers shows that finding the midpoint values of a half-life problem and understanding how such values are connected to prior or future values of the function is very difficult. Part of the difficulty lies in understanding the covariational nature of the quantities involved in describing growth and decay (Ström, 2006) and in reasoning multiplicatively in the context of exponential behavior (Ström, 2008). A recent study using GeoGebra also noted difficulties in algebraic manipulations of the expressions, difficulties with representation in a coordinate system, including difficulties in interpreting intercepts and meaning of axis (Tuda & Rexhepi, 2023). Given these documented difficulties, we expected that the selected topics would effectively reveal differences in instructors' knowledge of the content for teaching.

Combining these two organizers, tasks of teaching and function types, we proposed a blueprint with six dimensions (Figure 1) that guided the development of the instrument. While these six dimensions cannot represent all the knowledge that any instructor who teaches community college algebra uses in teaching such a course, they provide a good sampling of the knowledge that is foundational and thus necessary to teach the course.

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

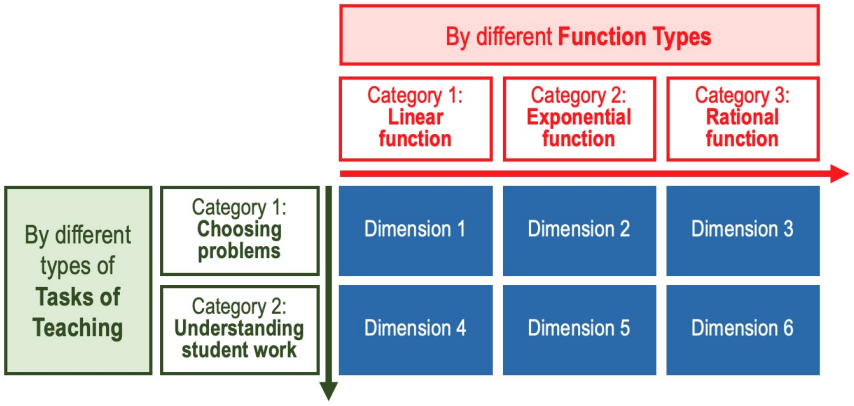


Figure 1: Blueprint for the development of the MKT-CCA instrument.

This theorization provides the logical underpinnings of the adequacy with which the content for the instrument we developed represents content domain, and of its relevance for the proposed interpretation of scores.

Development of the MKT-CCA

The 55-item MKT-CCA was developed in four phases over a two-year period. In the **construction phase**, we generated items through a two-week long writing camp with 23 people, including community college instructors, college algebra instructors at four-year institutions, mathematicians, graduate students, and researchers. The item camp was conducted virtually because of the pandemic. The teams were constituted by the project's PIs and two or three additional collaborators. Each person in a team was assigned to write at least 10 items on a particular function and a particular task of teaching. A first draft of the item was then shared among people within the team, for revisions and comments. A revised item was then passed to other teams, who reviewed the item and made additional suggestions. At the end of the camp, we had generated close to 250 different multiple-choice items (161 for understanding student work, and 89 for choosing problems). We created an item database in which we kept track of each item, the author, the content, and the document where it was located. Once the camp was over, the five

PIs, a postdoctoral fellow, and three graduate students reviewed a subset of items by answering the items and identifying whether each was targeting a unique mathematical idea and a unique task of teaching, and whether it had a unique and unambiguous way to be answered with the distractors given. Over half of the items (*n*=159) were discarded because they could be interpreted in more than one way, or the 'correct' answer was based on teaching preferences, or it was not possible to identify a correct or unique response within the distractors provided.

In the second phase, we conducted **cognitive interviews** with the 36 of these items (6 per dimension) with 12 instructors. Instructors received two items in advance and were presented with four additional items during the interview. We asked: How did you work on the items? How did you choose a response? What was your reasoning to arrive at a response? And, how familiar are you with these items? The interviews helped us identify problems with language, item organization, and content; we also attended to whether the participants, without prompting, noted the specific task of teaching and the function content. We used this information to revise 170 items (evenly distributed across function types and tasks of teaching) using a tiered review system that involved several teams. Nearly half of the revised items were deemed unusable through this process. As there were some dimensions that had a small number of items (e.g., rational functions) we sent a survey with 13 open-ended questions to 30 community college instructors requesting ideas of misconceptions that students would have with particular topics or for problems that could be used in a teaching context related to linear, rational, or exponential functions that would address those misconceptions. In addition, we asked this group to provide input on 12 items. The analysis of these data generated 16 new items, for a total of 72 items, that supplemented items in the dimensions that had a low number of items.

In the third phase, we engaged in two rounds of **expert review** for all available items. In the first round, a set of 50 items was sent to 10 people (5 were members of our advisory board, and included researchers with experience developing instruments and who were also faculty with experience teaching at community colleges or at universities; five were community college instructors who had an average of 22 years of teaching experience and were familiar with the purpose of the instrument). We used their feedback to refine the language in the items and to clarify the instructional goals of the items. The second round occurred prior to sending the instrument for piloting; we sent a revised set of 36 items to two mathematicians who had extensive experience in instrument development. They were asked to comment on the mathematical content, on whether the options were either unambiguously correct or incorrect, and whether the items were addressing the blueprint (Figure 1). Their review revealed that the items were targeting the blueprint, but that a handful of multiple-choice items could have more than one correct answer, or had mathematical errors, or were double barreled items (they could be addressing the two tasks of teaching, which would make inferences about one of the dimensions difficult). All this information was used to revise the items.

In the fourth phase, we conducted a **pilot test** of all the available items, 60. These were distributed in three forms. Half of the items were answered by 20 different people, and half by 40. This distribution was chosen to generate as much data as possible to examine basic psychometric properties of the items but maintaining a reasonable amount of time for taking the test. An examination of the responses, including discrimination, difficulty, and distractor parameters helped us identify items that needed further revision (e.g., those with low reliability or with options chosen by fewer than 5 participants). We engaged in further revision of items until we had what we believed was a solid set of items, a total of 34 items, that included 27

multiple-choice and seven testlets (each with four True/False options, resulting in 55 different scores), that almost reached the goal of having six items per dimension. The Appendix shows three of such items, one from each type of function.

We used this instrument to establish scores that measure community college instructors' mathematical knowledge for teaching college algebra, specifically the knowledge needed to perform two types of teaching tasks involving three types of functions. To evaluate the validity of our proposed interpretation and use of the scores, we explored the following research questions (RQs, Table 1), each addressing different sources of validity evidence needed to evaluate the validity of our proposed interpretation and use of test scores. We provide the methods used for each study followed by their specific results. A discussion section will then draw the validity arguments together to provide a holistic interpretation of the validity findings for the MKT-CCA.

Table 1: Research questions and evidence

Research Question		*Study, Evidence*
RQ 1.	Does the instrument capture the hypothesized distinction between the knowledge dimensions?	1: Evidence Based Internal Structure
RQ 2.	Does the process the respondents use when answering the items reflect the characteristics of the construct we intend to measure?	2: Evidence Based on Response Processes
RQ 3.	What is the relationship between instructor performance in the MKT-CCA test and their background (teaching experience) and demographic characteristics (gender, status, etc.)?	3, 4: Evidence Based on Relations to Other Variables
RQ 4.	What is the relationship between instructor performance in the MKT-CCA test and the reported frequency of use of the two tasks of teaching?	
RQ 5.	Can the MKT-CCA be used to provide a comparable measure of MKT among math majors? In particular,	5: Evidence Based Internal Structure, Evidence Based on Relations to Other Variables

Research Question		*Study, Evidence*
RQ 5.1.	Does the unidimensional scale of MKT work for math majors, and	
RQ 5.2.	How do the responses from math majors without teaching experience differ from the responses of the community college instructor sample?	

All the studies rely on the same community college instructor sample; Study 2 includes a smaller sample of community college instructors who responded to the MKT-CCA instrument. Study 5 includes a sample of math majors. As all studies use the same MKT-CCA instrument in the next section we describe the sampling for the community college instructors. The presentation of each study includes specific sampling, instrumentation, analyses, and results.

Community College instructors Recruitment and Sampling

We projected that we would need at least 500 active community college mathematics faculty with recent experience teaching college algebra that would be representative of the population of community college instructors in the United States and its territories. We began by selecting institutions, using a two-stage sampling process targeting community colleges and public institutions granting associate degrees. Searching the Integrated Postsecondary Education Data Systems (IPEDS) database we compiled a list of 1,374 institutions that fit these two criteria. Next, we developed a stratified sample of 199 institutions, from which we assumed we could generate a database of at least 800 instructors that would then provide proportional representation for the institutions. The stratification included four categories, region of the U.S., campus setting, size, and diversity of the student body, which together allow us to select institutions across a wide range of characteristics ensuring representativity of the institutions in the sample and that

have been associated with variation in student achievement at community colleges (Calcagno et al., 2008). See Table 2.

Table 2: Stratification categories for institutional sampling.

Stratification	*% of Institutions*
Region of the country	
Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont	14%
Midwest: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin	25%
West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming	22%
South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia	37%
Territories: Guam, Puerto Rico, Republic of Palau	2%
Campus Setting	
Urban (city)	34%
Suburban	21%
Small city (town)	25%
Rural	20%
Campus Size	
Small (Less than 1,000 FTE students)	15%
Small/Medium (1,000-4,999 FTE students)	44%
Medium/Large (5,000-19,000 FTE students)	35%
Large (20,000 or more FTE students)	6%
Campus Diversity	
Greater than or equal to 50% white	64%
Less than 50% white	36%

Note. These are national percentages representing community colleges and public institutions granting associate degrees. Campus Diversity was determined using the NCES data student and faculty data from https://nces.ed.gov/ipeds/datacenter/SelectVariables.aspx

Upon inspection of the 199 institutions, we decided to add all 122 the minority serving institutions (Tribal College, Historical Black College, Asian and Hispanic serving institutions) to oversample their mathematics faculty to ensure a better representation of institutions serving these students. Thus, our initial institutional sample included 321 institutions. We used institutions' websites to build a database of all the instructors who taught mathematics at these institutions; this generated an instructor sample of 1386 instructors from non-minority institutions and 2,172 from minority serving institutions. These instructors received an invitation to participate in the project which required filling a screening form that asked them to check interest and availability to participate in different aspects of the project (i.e., taking tests, participating in cognitive interviews, and giving feedback on items). This list of interested instructors was fundamental for randomly selecting instructors to participate in cognitive interviews for item development (12 out of 37 interested), give feedback to items (15 out of 144), and pilot the instrument (121 out of 268). To recruit for the testing of the final version of the instrument, we reached out to all faculty in all the institutions in our list (~7,700 instructors) to reach at least 600 instructors responding to the test. This number was needed to run the dimensional analysis for the instrument, given that it contained 55 items (27 multiple-choice items and 7 testlets with four True/False options).

We turn now to Study 1, which seeks to gather evidence to support the claim that the instrument captures the hypothesized distinction between the knowledge dimensions.

Study 1: Does the instrument capture the hypothesized distinction between the knowledge dimensions?

The MKT-CCA was administered via an online Qualtrics survey; all the items were presented in the same order, in groups of five items per page. Instructors were asked not to use technology or other aids in answering the items, and to allocate at least two hours to complete the test. Instructors could take breaks. The test was not proctored.

To assess dimensionality of the construct measured by the MKT-CCA instrument, we first tested whether all items cohesively measure an overall knowledge construct. Using the scored item responses, we first conducted a two parameter item response theory (2PL IRT) model, which accounts for differences in both item difficulty and item discrimination, with 55 items on data from 416 teachers who completed all 55 items and who spent less than 1200 seconds in the survey (suggesting that respondents would have spent less than a minute per item). Further evaluation of item discrimination estimates would lead us to remove 17 items that provided little information and show low ability to differentiate among teachers (item discrimination under 0.61). As our goal was to develop items measuring mathematical knowledge for teaching algebra in community colleges, we expected for all items to correlate well to represent overall MKT. We reviewed the content of these items (6 multiple-choice items and 11 True/False items) and decided to remove them because these items did not capture any additional essential aspects of the construct, and their content was being covered in other better-performing items. With the remaining 38 items (23 multiple-choice items and 15 items from 5 testlets with 2-4 True/False sub-items), we conducted a six-dimensional multidimensional IRT (MIRT) model, where each dimension is estimated by the items designed to measure one task of teaching and one function type in alignment with the instrument blueprint (Figure 1).

After conducting a six-dimensional model structured by three types of function across two types of teaching simultaneously, we also conducted a three-dimensional model structured by three types of functions, with each dimension combining two tasks of teaching, as well as a two-dimensional model structured by two tasks of teaching, with each dimension combining three function types, separately. Differentiation by different tasks of teaching or function types was also tested—either within the same function type (for testing distinctions between different tasks of teaching) or within the same task of teaching (for testing distinction between different function types) —while controlling for other factors. For our analysis, we scored the responses to items as binary, with incorrect responses marked as 0 and correct responses marked as 1. An Exploratory Factor Analysis (EFA) was also conducted to examine whether there was any meaningful structure among the items suggested from data, without imposing any assumptions on the model. All analyses were conducted in Mplus (Muthén & Muthen, 1998-2023). Latent constructs were identified by setting factor means to 0 and factor variances to 1. Each test was conducted using both Maximum Likelihood (ML) and Weighted Least Squares Mean and Variance adjusted (WLSMV) estimators, with sum scores for testlet items and with individual item scores for testlet items, respectively to examine whether different scoring approaches yield any inconsistencies in results.

Findings

The six-dimensional model with 38 items distinguishing factors by three different types of functions across two tasks of teaching failed to converge using both the maximum likelihood (ML) and weighted least square mean and variance adjusted (WLSMV) estimators. This suggests that our hypothesized model, which differentiates all six dimensions, doesn't align with our data and implies that the knowledge dimensions measured by our items do not distinctively represent

the hypothesized knowledge dimensions. We also tested the same six-dimensional model with sum scores for five testlet items, but the results were consistently non-positive definite, indicating that the model could not be estimated with the given structure.

Alternatively, we conducted a three-dimensional model, with each dimension estimated by a different function type item set: linear, exponential, and rational. We did this to examine if the knowledge construct is differentiated by the three function types. The model using WLSMV suggested that the three-dimensional model fit the data well (RMSEA=0.025, CFI=0.944, TLI=0.941). However, the three factors were highly correlated (0.88 to 0.96), indicating that factors may not be significantly distinguishable from each other. To test whether the three factors were indeed distinguishable by function type, we conducted a model comparison test using DIFFTEST (Muthén & Muthen, 1998-2023), and compared the three-dimensional model to three two-dimensional models, each combining a pair of function types (linear + rational, exponential+rational, linear+exponential). Each of these two-dimensional models fit the data well, but the model combining the linear and rational functions was not significantly worse than the three-dimensional model, (DIFFTEST χ^2=9.031, df=2, p=0.0109), indicating that linear and rational function factors are not statistically distinguishable (whereas linear and exponential, exponential and rational are). An examination of the tetrachoric correlation matrix for each pair of 38 items, also did not show a grouping according to function type; in fact, the correlation between some linear and rational items was higher than 0.5.

Next, we conducted a two-dimensional model to assess whether the knowledge dimensions were distinguishable by two different tasks of teaching. In this model, each factor was estimated by the items designed to measure the knowledge in understanding student work (USW) and choosing problems (CP). The two-dimensional model using WLSMV estimator, where all the USW items

loaded on one factor and all CP items on another, resulted in a non-positive definite problem. This indicates that the model could not be estimated. The same result occurred with the ML estimator and when using sum scores for testlets. We also examined whether the dimensions by different tasks of teaching were distinguishable *within* each function type. However, due to non-positive definite issues within linear and rational function type items, the model could not be estimated. The finding implies that the two factors, USW and CP factors are not significantly distinguishable. In conclusion, we cannot confirm that MKT-CCA items are structured by either function type or task of teaching.

As the knowledge construct measured by our items did not seem distinguishable by either task of teaching or function type, we examined whether there was any meaningful structure among the items suggested by an exploratory factor analysis (EFA), without imposing any assumptions on the model. The EFA was conducted under multiple conditions, including using the full dataset with missing data, using only complete cases, and using randomly selected subset of the data. The results suggest one main factor, and other factors where a small number of items are loaded when the analysis is forced to distribute the items into two to four factors. Items commonly suggested to be loaded on different factors include four exponential items and two linear items. We examined the content of these items but could not find any characteristics that would justify loading them on different factors. Accordingly, we excluded the six items that did not load on the main factor then fit the unidimensional model with the remaining 32 item scores (19 multiple-choice and 13 True/False items from 7 testlets). The fit statistics suggest that the model fits well (RMSEA=0.016, CFI=0.981, TLI=0.980), when treating testlet subitems as individual items. Item factor loadings for all the items were greater than 0.3 in the standardized version. This unidimensional CFA was also conducted under multiple other conditions, including

analyses on a randomly selected subset of the data, and the results remained consistent across these conditions.

Specifically, the CFA model that accounted for dependence among items within the same testlet by using summed testlet scores showed improved model fit (RMSEA=0.008, CFI=0.997, TLI=0.997). Based on this, we used the summed scores for sub-testlet items (total 26 item scores) in subsequent analyses. The unidimensional model using only the 19 multiple choice items (without testlets) was also tested and showed good fit (RMSEA=0.011, CFI=0.995, TLI=0.994).

Based on all the analyses, we conclude that our instrument measures a unidimensional construct involving two tasks of teaching and three function types. Even though the structure does not align with our hypothesized structure, all MKT items coherently measure an MKT construct. Thus, we decided to establish a unidimensional scale for the MKT and report evidence supporting the validity of this unidimensional score.

Study 2: Does the process that the respondents use when answering the items reflect the characteristics of the construct that we intend to measure?

Following the literature, we planned to interview 30 participants (~7% of the 416 respondents with complete data) to corroborate, via cognitive interviews, that the participants had used the anticipated knowledge in answering the items. These 30 participants were randomly selected using a stratification based on performance on the full MKT-CCA instrument (maximum score of 55, separated into quintiles). We sought to include participants from different regions of the United States and different employment status. We invited 100 participants (20 per MKT-CCA quintile) and interviewed 26 participants with a relatively balanced distribution across the five

quintiles (4 in the 2nd quintile, 5 each in the 1st and 3rd quintile, and 6 participants each in the 4th and 5th quintile). Following recommendations from our advisory board, we randomly selected six items (5 multiple choice and one testlet) for the interviews and used a 5-question interview protocol (1. Why did you choose that option? 2. What specific mathematical ideas do you see addressed explicitly in this item? 3. What common misconceptions might students have about the mathematical ideas in this item? 4. Does the scenario presented remind you of something you've experienced in your teaching? If so, how? 5. Any other thoughts about this item?). After piloting the process with three PhD graduate students (pure mathematics, applied mathematics, and mathematics education), we removed the testlet item, due to time concerns, and conducted the interviews with five items, two on rational functions, two on exponential functions, and one on linear functions.

To analyze these data, we followed the guidelines set by the Educational Testing Service (2015), the *Standards for Educational and Psychological Testing* (AERA, 2014), and Bostic et al. (2021) in conceptualizing and planning the cognitive interviews. Cognitive interviews for validation seek to establish whether the knowledge we hypothesized participants would use in answering the items correctly and incorrectly was indeed used. Our item templates were designed with this in mind. For each multiple-choice option three statements were included, whether the option was correct or incorrect, what knowledge should be used to answer the item, and what mathematical reasoning would lead to a decision in choosing the option. The distinction between knowledge and reasoning was made because knowledge can lead to different types of reasoning in solving a mathematical problem. We sought to be comprehensive in the possible knowledge and reasoning used. To analyze the cognitive interviews, we distributed the items across the researchers, and each person examined the recordings and notes taken to produce summaries related to the

process participants used to answer the item, and for the option chosen, whether the participant had used the anticipated reasoning and knowledge. Each researcher also listed all the mathematical notions and misconceptions that the participants mentioned were related to the item. Next, pairs of researchers developed summaries that considered all the responses to a given item. In a meeting with all the researchers, we used the summaries to answer the question: does the evidence support the notion that for this particular item participants are using mathematical knowledge for teaching ideas associated with the mathematical content of the item (linear/rational/exponential functions) and are they making comments related to teaching? We used these two questions to read the summaries collectively. Next, we aggregated data across participants and across items, regarding the proportion of participants who answered the item correctly, the proportion of participants who used the expected knowledge, and the difference between participants' responses in the MKT-CCA and in the cognitive interview for each item. These were used to provide evidence that the items captured the anticipated knowledge, and by extension that the instrument can be used to assess this knowledge.

Findings

We analyzed 130 responses (from 26 participants commenting on 5 items) to identify whether the respondent had used the anticipated knowledge when choosing an answer. For the most part, participants chose the correct option (94 responses out of 130, 72%) and of these responses, 87 (92% of correct responses) used the anticipated knowledge. Of the 36 incorrect responses, 26 (72%) show that the participants had engaged in incorrect reasoning that we had anticipated as illustrative of not having the targeted knowledge. Out of the 130 responses, 16 (12%) were obtained with knowledge or reasoning we did not anticipate. Thus overall, in 87% of the responses, the participants' responses indicated that the items were targeting the anticipated

knowledge. When asked about the mathematical notions that the items addressed, all participants identified topics directly relevant to each item (e.g., intercepts, slopes, exponential growth and decay, asymptotes, holes, zeros, contextualized problems, graphing, proportions, solving equations, exponents, polynomials, etc.). The analysis of the potential misconceptions that instructors said could be elicited with an item suggested specific errors and misconceptions that were relevant to the notions at stake (e.g., not being able to differentiate between growth factor or growth rate, having difficulties factoring or identifying zeros of functions, or not recognizing proportions of differences). Across the board, participants indicated that the topics of the items were familiar (in only two responses, the instructors indicated that they had not seen the content for rational functions in their college algebra course), although for 25% of the responses, the participants said that the format or the way in which the questions were posed was unfamiliar. The analysis of the 16 responses of knowledge we did not anticipate, reflected knowledge or reasoning other than the one we intended; some used arithmetic rather than algebraic heuristics, elimination of options, or guesses; but some showed reasoning that followed the same misconception as the one proposed in the items or an incorrect application of the pertinent knowledge. We surmise then, that overall, their responses suggest that the instructors were not sufficiently familiar with the knowledge targeted by the item.

Thus, the information gathered through the cognitive interviews supports the claim that the items elicit mathematical knowledge for teaching for foundational topics in college algebra related to linear functions, rational functions, and exponential functions. We affirm this based on the participants' approaches when answering the items, the varied range of mathematical notions they mentioned, the different student misconceptions that they experienced students having, and the level of familiarity with the content and the context of the items. While the items are only a

subset of the items in the instrument (5 out of 32), given that the rest of the items are targeting comparable material and are correlated, we believe that it can be expected that the instrument assesses the mathematical knowledge for teaching it was designed to assess. In addition, and across all the items, the instructors said that the situations had occurred to them, explicitly mentioned misconceptions, and recognized the errors in the items or in the options as errors students would make. They also mentioned problems that they would use either to present the material or to address the misconceptions seen. For the items that had a real-world situation (e.g., radioactive material) participants noted difficulties with modeling or being mindful of selecting authentic and relatable problems. Thus, during the cognitive interviews, the instructors mentioned the items as eliciting knowledge that related to understanding student work and choosing problems.

Having established the dimensionality of the instrument and gathered evidence that it assesses the expected knowledge, we turn into investigating the relationship between the instrument and other variables in Study 3 (participants' background and demographics) and Study 4 (reported frequency of use of two tasks of teaching).

Study 3: What is the relationship between instructor performance in the MKT-CCA test and background (educational and teaching experience) and demographic characteristics (gender, status, etc.)?

To investigate the relationship between instructor performance in the MKT-CCA test and background and demographic characteristics, we distributed a background and demographics survey containing 19 questions, organized into three sections. The first section included nine questions about work and preparation (e.g., their situation as a community college employee, participation in professional development, degrees, courses taught, and number of years teaching

mathematics). The second section included six questions that describe six instructional practices (e.g., cooperative learning, lecture) and asks for the frequency that instructors use them in their classes using a 5-point scale (*Never*, *At most ½ of class sessions*, *More than ½ of class sessions*, *Every class session*, *Multiple times in class sessions*). The third section included four demographic questions (gender, ethnicity, race, age). The instrument was administered after participants completed the MKT-CCA test.

We used a structural equation model (SEM) to examine the extent to which these background variables predicted the latent construct MKT. Specifically, we used variables that had been hypothesized to have an impact on MKT, specifically: experience in teaching post-college algebra courses (Ko *et al.*, 2024), years of full time mathematics teaching (Hill *et al.*, 2004; Hill et al., 2012), educational background (e.g., highest degree and field, Desimone et al., 2016; Holm & Kajander, 2020), extent of professional development (Carney et al., 2016; Koellner & Jacobs, 2015; Österholm et al., 2024), frequency of instructional strategies used (e.g., inquiry- or not, (Charalambous & Hill, 2012; Chin et al., 2016; Marshall et al., 2009), and demographics (e.g., gender and race, Dee, 2005, 2007; Metzler & Woessmann, 2012). These analyses were done with the various sets of scores: 32 (from all retained items, 19 multiple-choice items and 13 true/false items from 7 testlets), 26 (19 multiple-choice items and sum scores from 7 testlets), and 19 scores (from only the multiple-choice items). We report on the results with 26 scores because all the results were consistent across the three sets and the measurement model with 26 scores that accounted for correlations among testlet subitems showed the best model fit in the dimensional analysis.

Findings

The majority of the 416 instructors (57%) were middle-aged, between 36 and 55 years old, White (76%), with 95% indicating they were non-Hispanic. Most participants (70%) reported having a master's degree, while 24% indicated having an EdD, EdE, or a PhD degree. The highest degree held by most participants was a master's in Mathematics (43%), followed by a master's in Mathematics Education (15%) and a PhD in Mathematics (12%). In total, 59% of participants have a Mathematics-related degree as their highest degree. Nearly half of the participants (46%) had taught post-college algebra courses (such as Calculus 1, 2, 3, Business Calculus, Differential Equations, Linear Algebra, etc.) 10 or more times, whereas 16% of the participants have never taught mathematics courses beyond College Algebra Courses (see Table 3). On average, instructors (*N*=407) reported having nearly 17 years of full-time equivalent teaching experience (mean=16.8 years, SD=9 years, range: 1.5 to 47). Of the 416 participants, 76% were full-time instructors hired to teach math (N=310), 2% were full-time instructors hired for another subject but also teaching math (N=10), 9% were part-time instructors who desired to teach math part-time (N=37), 8% were part-time instructors who desired full-time math positions (N=35), and less than <1% were part-time instructors desiring full-time positions in other subjects (N=2).

Table 3: Instructor Sample Characteristics (*N*=416).

Characteristics		#	%
Gender			
	Man	201	48
	Woman	192	46
	No response	23	6
Age			
	26-35	46	11
	36-45	133	32

46-55	108	26
56-65	73	18
Over 65 years old	31	7
Prefer not to answer	19	5
No response	6	1
Hispanic or Latino (excludes 11 blank (3%) responses)	22	5
Race		
American Indian or Alaska Native	3	<1
Asian	47	11
Black or African American	16	4
Native Hawaiian or Other Pacific Islander	2	<1
White	317	76
Other	18	4
No response	13	3
Highest Degree and Field		
Master's	304	73
Mathematics	179	43
Mathematics Education	62	15
Other	63	15
Doctoral	106	26
Mathematics (PhD)	51	13
Mathematics Education (EdD, PhD)	25	6
Other (EdD, EdE, PhD)	30	7
No response	4	1
Other	2	<1
Current Role as a Math Instructor		
I am a full-time instructor, hired to teach math	310	75
I am a full-time instructor, hired to teach another subject but I also teach math	10	2
I am a part-time instructor who desires to teach only part-time in math	37	9
I am a part-time instructor who wishes to teach full-time in math	35	8
I am a part-time instructor who wishes to teach full-time in some other subject	2	<1
Other	19	5
No response	3	<1

Math-specific, off-campus Professional Development sponsored by a professional organization Participation		
0~2 times a year	355	85
At least 3 times a year[a]	53	13
No response	8	2
General off-campus Professional Development Participation		
0~2 times a year	344	83
At least 3 times a year	60	14
No response	12	3%
Using IBL in the classroom portion of College Algebra		
At least some portion of class sessions[b]	187	45
Never	220	53
No response	9	2
Experience Teaching Math Courses Post College Algebra		
Never Taught	68	16
Taught less than 5 times	79	19
Taught at least 5 times but less than 10	74	18
Taught 10 or more times	190	46
No response	5	1

Notes: a. 3-4 times a year, 5 or more times a year. b. Includes At most 1/2 of class sessions, More than 1/2 of class sessions, Every class session, and Multiple times in class sessions.

The SEM model was tested using 26 item scores using predictors we hypothesized would be associated with MKT-CCA based on literature and previous studies. Specifically, the estimated MKT-CCA was predicted by experience in teaching post-college algebra courses, number of years teaching mathematics, indicators of highest degree (PhD in mathematics, a Master's degree in mathematics, PhD or EdD in mathematics education, a Master's degree in mathematics), frequency of general off campus professional development participation, frequency of mathematics-specific off campus professional development participation, indicator of using IBL in the class, indicator of full-time math instructor position, and indicators of Men and White. We report estimated unstandardized (B) and standardized coefficients (β) in Table 4. The model fits

the data well (RMSEA=0.015, CFI=0.969, TLI=0.967). The predictors of post-college algebra teaching experience, holding a PhD in mathematics, having a full-time math instructor position, and participating in math-specific off-campus professional development, being White showed significant positive effects on MKT-CCA scores whereas the indicator of being Hispanic or Latino showed a significant negative effect on MKT-CCA scores. For example, participants holding a highest degree PhD in mathematics have MKT-CCA scores that are 0.82 higher (SD=1, Mean=0) compared to instructors without a PhD in mathematics.

Table 4 Summary of Regression Analysis for Background Variables Predicting Instructors' MKT-CCA (N = 402[a]) estimated by 26 scores.

	MKT-CCA		
Variable	B	SE B	β
Experience in teaching post-college algebra	0.19**	0.06	0.19**
Number of years teaching mathematics	-0.01	0.01	-0.05
Highest degree PhD in math	0.82**	0.23	0.22**
Highest degree MA in math	0.31	0.17	0.13
Highest degree in PhD or EdD in math education	0.07	0.28	0.02
Highest degree MA in math education[b]	0.22	0.18	0.07
Full-time Math instructor position	0.47**	0.14	0.17***
General Professional Development (PD)	-0.27	0.18	-0.08
Participation of Math specific, off-campus PD provided by professional organization	0.39*	0.18	0.11*
Use IBL in the class	-0.04	0.12	-0.02
Gender (Men=1)	0.11	0.13	0.05
Race (White=1)	0.74***	0.15	0.25***
Hispanic or Latino (1=Hispanic or Latino)	-0.80**	0.30	-0.14**
R^2	0.26		

Note: a. Included here is the sample of faculty for whom there was complete data for the variables shown. b. Ten participants whose highest degree spans multiple disciplines (e.g., both mathematics and mathematics education) have been excluded.

Study 4: What is the relationship between instructor performance in the MKT-CCA test and the reported frequency of use of the two tasks of teaching?

To investigate the relationship between instructor performance in the MKT-CCA test and reported frequency of using the two tasks of teaching, we administered a subset of items from the Task of Teaching Inventory developed by Herbst and Colleagues [H&C]. The original survey consists of 33 items asking how frequently teachers engage in generic tasks applicable to any mathematics course. The survey and some items from this survey were described and analyzed in Ko and Herbst (2020). Our study used seven items total: three were taken from Ko and Herbst [KH] (2020), two from the original survey, and two were developed by the research team. Three items are intended to capture the frequency that participants engage in specific activities related to the task of teaching of choosing problems (C1: *Assist a student in a small way, such as by giving a mathematical hint or asking a question, without solving the problem for them* [H&C]; C2: *Evaluate how well a problem meets your instructional goal;* C3: *Modify problems from a textbook or from colleagues* [KH]; and four to capture the frequency that participants engage in specific activities related to the task of teaching of understanding students' work (U1: *Notice that a student's mathematical approach to a problem is valid even though it is not standard* [H&C]; U2: *Give detailed comments on a student's written work*; U3: *Read students' work to figure out their thinking process* [KH]; U4:*Try to understand how a student came up with their answer to a problem* [KH]). These were assessed on a 5-point scale (1- Never, Almost Never, Sometimes, Often, 5-Very Often). This instrument was distributed as an optional instrument after the MKT-CCA and the Background and Demographics instruments.

Figure 2 shows the reported frequency of using each of seven activities, in the Task of Teaching instrument.

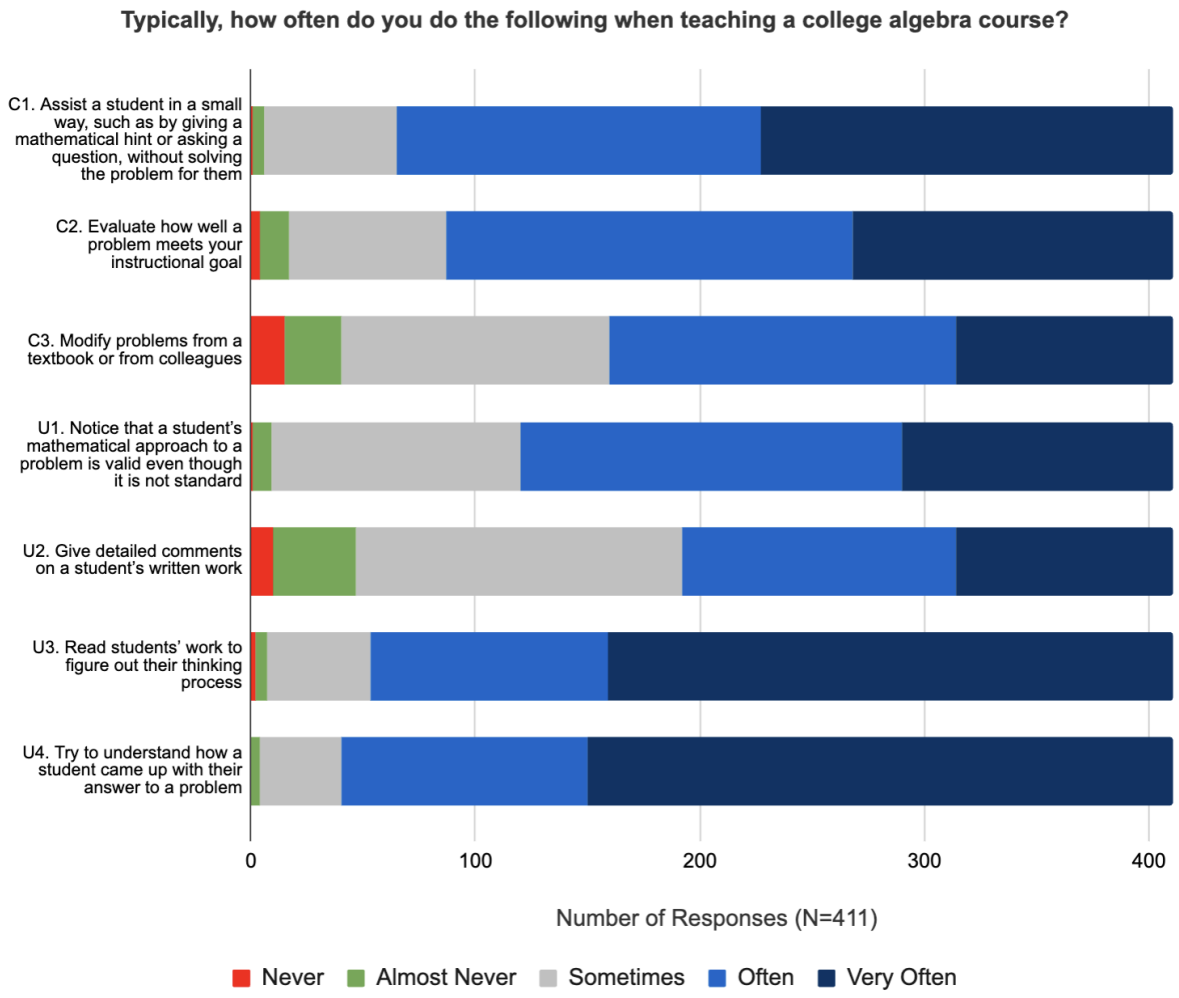


Figure 2: Reported frequency of using each of the eight activities.

Findings

As shown in Figure 2, most instructors frequently engage in all seven activities. Specifically, more than 50% of instructors often perform two tasks related to understanding student work, namely "Try to understand how a student came up with their answer to a problem" and "Read students' work to figure out their thinking process." Compared to other activities, instructors indicate that they are less likely to give detailed comments on a student's written work. Additionally, a noticeable number of instructors (n=15, 4%) reported they have never modified

problems from a textbook or from colleagues. A two-factor confirmatory factor analysis was conducted with the seven items loaded onto either factor representing understanding student work or choosing problems. The results suggested that the model did not fit well, but modification indices suggested an error correlation between two items (U1, C1) that involved recognizing a student's unique approach to problem-solving in mathematics rather than directing the student into standard methods. Given the interpretable reason for the error correlation, we fit the same model with the error correlation between the items. The result of the model suggests that the model fits the data well (RMSEA=0.073 with its confidence interval includes 0.05, CFI=0.988, TLI=0.979), indicating that four items (U4: *Trying to understand how students came up with their answer to a problem*, U3: *Reading students' work to figure out their thinking process* and three items, U2: *Give detailed comments on a student's written work*, U1: *Noticing that a student's mathematical approach to a problem is valid even though it is not standard in a college algebra course)* measure frequency of using understanding student work and three items (C3: *Modifying problems from a textbook or from colleagues*, *C2: Evaluating how well a problem meets your instructional goal*, C1: *Assisting a student in a small way, such as by giving a mathematical hint or asking a question, without solving the problem for them*) measure frequency of choosing problems.

Next, the two factors representing the frequency of each teaching task were used as predictors of MKT-CCA in a SEM model. The results suggest the model fits the data well (RMSEA=0.012, CFI=0.993, TLI=0.992). Specifically, the frequency of choosing problems significantly affects MKT (β=0.377, SE=0.117, p=0.001), whereas the frequency of understanding student work is not a significant predictor of MKT estimated by 26 item scores. The diagram below shows the

estimated relationship between the two teaching tasks and MKT, as well as MKT and the items (see Figure 3).

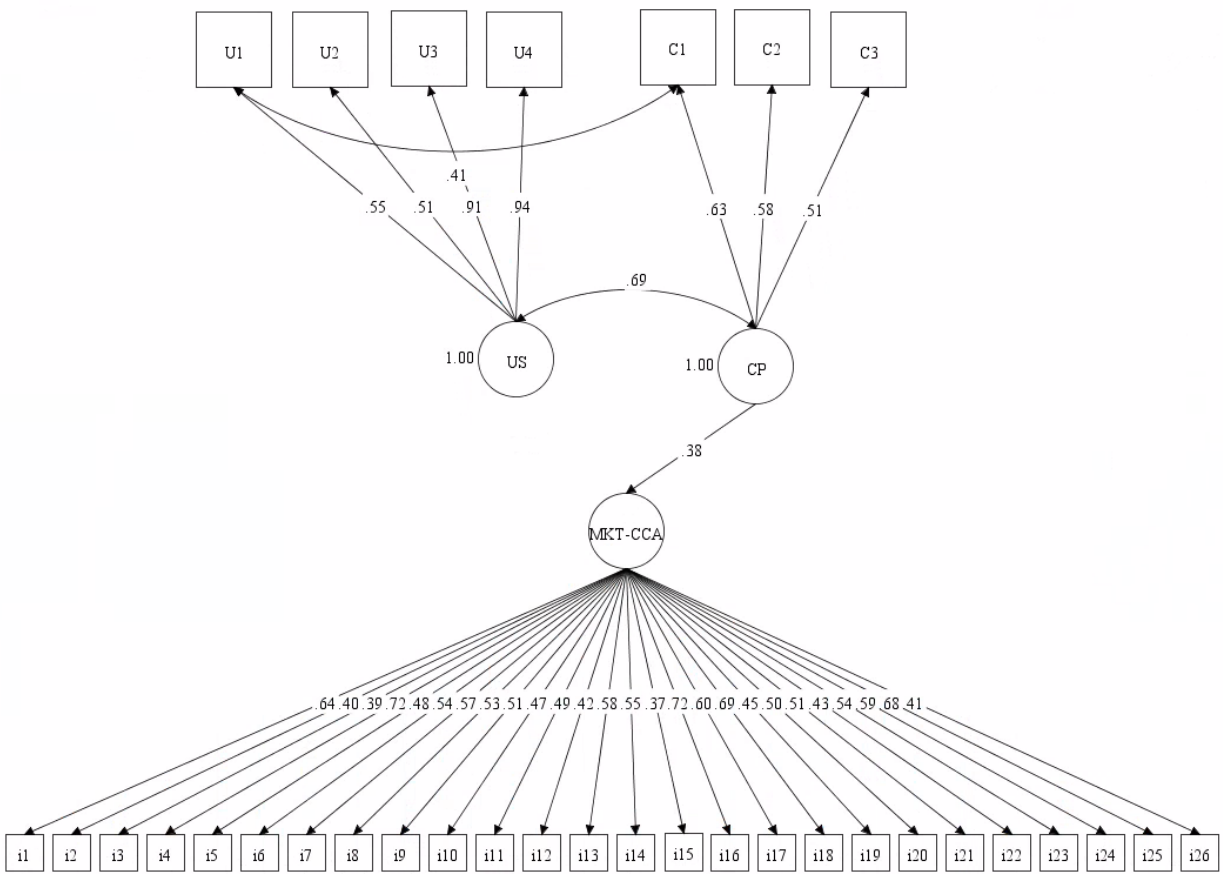


Figure 3: Structural model diagram showing only significant ($p < .05$) standardized effects.

Our final study compares the results of the MKT-CCA from math majors without teaching experience to those from the community college instructor sample, to provide further evidence that (1) the unidimensional MKT scale is applicable to math majors, and (2) how do the responses differ between the two samples. The first analysis provides evidence based on the internal structure of the instrument, whereas the second provides evidence based on relations to other variables.

Study 5: How do the responses from math majors without teaching experience differ from the responses of the community college instructor sample?[2]

To recruit students we identified large, public, four-year colleges across all states that offered a mathematics degree and that had a high percentage of admitted students (40%-80%). This list resulted in 445 institutions that varied in location and size. From this group, we randomly selected about 10 percent of institutions (n=47) and identified a faculty member from the mathematics department who could distribute an invitation to their mathematics majors. The invitation described the project and asked students to fill a screening questionnaire if they were interested in participating in the project. The screening questionnaire asked for their major; their experiences tutoring in elementary, middle, high school or college; and the courses they had tutored. The questionnaire also asked for their teaching experience and asked for an email that could be used to distribute the MKT-CCA link to eligible students: those without tutoring experience beyond middle school or who had no teaching experience. Once they completed the questionnaire, they received a gift card for $25.00. A second round of invitations was sent to an additional set of 32 institutions, as many students were interested (n=446) but a large percentage (47%) were ineligible. The survey was open for approximately four weeks. Students who had not completed the survey were sent several reminders to complete it. Of the 191 students who received a link to complete the questionnaires, 149 started the questionnaire, and 110 completed it. The students came from 24 different institutions, located in 19 states. As a group, the institutions in which the students were enrolled were diverse; they tended to be large institutions

[2] Portions of this section are adapted from our earlier work presented at the 27th Annual Conference on Research in Undergraduate Mathematics Education, Alexandria, VA

(average enrollment: 27,946, SD=13,550), non-selective (average of percent admitted students was 66%; SD=12%), and relatively affordable (average tuition cost: $16,390, SD=$4,578). The institutions were in large, midsize, and small cities; in large and midsize suburbs; and in fringe and remote towns.

The MKT-CCA questionnaire sent to the students was exactly like the one sent to the instructors. Students were asked questions about the courses they have taken (e.g., pre-college algebra courses, courses in the calculus for STEM sequence, non-math major courses, math major courses) and questions asking whether they had teaching or tutoring experiences or plans for teaching. We included six questions asking how frequently they experienced six instructional practices (e.g., cooperative learning, lecture) in their classes using a 5-point scale (*Never*, *At most ½ of class sessions*, *More than ½ of class sessions*, *Every class session*, *Multiple times in class sessions*), and six questions asking how frequently they engaged in activities related to choosing problems or understanding student work (*Solve problems from different textbooks*, *Evaluate if my solution is the same as the one from my peers*, *Try to understand how my peers came up with their answer to a problem*, *Ask the teacher to show me the solution to a problem*, *Ask an online tutor for hints on problems*, *Explain my solutions to peers*); these were assessed on a 5-point scale (Never, Almost Never, Sometimes, Often, Very Often). The questionnaire ended with four demographics questions (gender, ethnicity, race, and age).

The analysis of the student data proceeded in three phases. We first compared 32 MKT-CCA binary item scores (coded as incorrect and correct) between instructors and students. We then conducted an independent sample *t*-test to compare the mean proportion of correct responses on each of the MKT-CCA items between the two groups. Next, we conducted the unidimensional CFA of the MKT-CCA in the student sample using the 19 multiple-choice items, and then we

compared the estimated MKT-CCA scores between instructor and student groups using the MIMIC (Multiple Indicator Multiple Cause) model. In the latent model, we decided to use only the 19 multiple-choice items instead of all 32 item scores due to the small student sample size, aiming to maintain the minimum sample-to-item ratio of 5-to-1 suggested in studies (Suhr, 2006). We used the MIMIC model to examine differences between instructor and student groups and to identify items that may function differently across groups in measuring MKT-CCA. Difference in MKT-CCA score should reflect actual differences in the underlying construct being measured, not from item bias favoring one group. Items with significantly different item intercepts, even after adjusting for latent mean differences, are considered differential item functioning (DIF) (Kaplan, 2008).

Findings

Among the 110 students who participated in the survey, 25 were excluded from the analysis because they had tutoring experience (n=11) or had completed the MKT instrument in less than 1200 seconds (the same criteria applied to the instructor sample). The exclusion criteria resulted in 85 completed student responses, and it constitutes the analytical sample. Table 5 presents characteristics of this sample.

Table 5: Student sample characteristics, N= 85.

Characteristics	#	%
Age		
Under 18 years old	0	0
18-19 years old	30	35
20 years old	20	24
21-22 years old	21	25
22-23 years old	5	6
Over 23 years old	9	11
Gender		

Woman	36	42
Man	43	51
Transgender	1	1
Non-Binary, Agender, Questioning	5	5
Hispanic or Latino	11	13
Race*		
American Indian or Alaska Native	2	2
Asian	21	26
Black or African American	4	5
Native Hawaiian or Other Pacific Islander	1	1
White	62	76
Tutoring Experience in Middle or High School		
No	85	100
Math courses taken in college		
Pre-college algebra courses	65	76
College Algebra level courses	68	80
Calculus courses	85	100
Math courses for majors	50	59
Non-math major college courses	51	60

Note: *82 students provided race information. Five among the 82 students chose more than one option, thus the total percentage is over 100.

The majority (57%) of the participating students were between 18 and 22 years old, identified as men (51%) and as non-Hispanic (87%) and White (76%). All the participating students indicated that they have completed calculus courses and most also took other mathematics courses, including pre-college algebra, college algebra level courses, math courses for majors, and non-math major college courses. As noted, none of the students chosen indicated that they had tutoring experience. A SEM model that examined the effect of individual course variables (indicating whether a student had taken a specific course) on MKT-CCA score suggested that there was no statistically significant effect of any individual course on students' MKT scores.

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

Tests comparing the mean proportions of correct responses between the two samples for each of the MKT-CCA items showed statistically significant differences in 22 out of 32 items all distributed across the six dimensions, (3 Exponential-Choosing problems, 3 Exponential-Understanding student work, 5 Linear-Choosing problems, 3 Linear-Understanding student work, 4 Rational-Choosing problems, and 4 Rational-Understanding student work). Since there was no clear pattern across different function types or tasks of teaching, we surmised that the type of function or task of teaching was not the main factor causing the difference. All the differences, defined as the mean proportion of instructors' correct responses minus that of students', were positive, meaning that instructors answered a significantly higher proportion of MKT-CCA Items correctly than students. This supports our hypothesis that instructors should have higher item scores across all items, providing expected evidence for relations with other variables, here, group membership (students or instructor).

Next, we evaluated the unidimensional model of MKT-CCA in the student sample using the 19 multiple choice items. A two-parameter item factor analysis was conducted with the 19 items using the WLSMV estimator. The model fit the data well (RMSEA=0.011, CFI=0.996, TLI= 0.995). All item loadings were significant and greater than 0.3. This provides evidence for the internal structure of MKT-CCA within the student sample.

The MIMIC (Multiple Indicator Multiple Cause) model regresses MKT-CCA on the group variable that indicates whether a sample is from a group of instructors (0) or students (1), while establishing the unidimensional MKT-CCA simultaneously, fits the data well (RMSEA=0.021, CFI=0.979, TLI=0.977). The result suggests that being in the student group has a significantly negative effect on MKT-CCA. Specifically, on average, the student group scored 0.40 standardized deviations lower on the MKT-CCA score than the instructor group. This effect size

is medium according to (Cohen, 2013) criterion (small, medium, and large for values of 0.20, 0.50, 0.80). To determine whether the difference was not driven by a few items favoring one group, we tested whether the two groups responded similarly to each item while controlling for their overall MKT-CCA level. Items that show differences after adjusting for latent mean differences are considered to reflect differential item functioning (DIF)(Kaplan, 2008). There were two DIF items favoring the instructors (linear-choosing and rational-understanding). For example, one item on choosing problems in linear functions presented a professor's intended learning goal and asked respondents to choose the form of a linear equation that would best help students achieve that goal. The correct answer is a form of a linear equation that is less commonly introduced in textbooks compared to other forms like the slope-intercept or point-slope form. The form is more commonly found in advanced mathematics courses or specialized texts on analytic geometry. As a result, students might be less familiar with it than instructors, who have broader experience teaching various courses related to algebra. This familiarity disadvantage may have contributed to students' score lower than instructors at the same level of MKT-CCA.

An adjusted MIMIC model that included additional paths from the MKT-CCA to the two identified DIF items showed improved model fit (RMSEA=0.012, CFI=0.993, TLI=0.993). The estimated difference in MKT-CCA between instructors and students remained significant (standardized β =-0.35), indicating that instructors tend to have higher MKT-CCA than students, even after controlling for DIF items. These analyses suggest that when administered to a population that presumably has the mathematical knowledge but no teaching experience (here assumed for the students to be in the form of tutoring), their scores are significantly lower than

those of the intended population, suggesting that they possess lower amounts of the knowledge assessed by the instrument.

The result confirmed our hypothesis that students would score lower on the MKT-CCA, as the measured knowledge reflects the mathematical knowledge used in teaching, and instructors with experience teaching the content would be expected to perform better. However, it remains unclear whether the instructors' higher scores stem from teaching experience or from taking more advanced mathematics courses compared to the student sample. Future analysis could examine item-level response patterns among instructors to identify qualitative features of the items influenced more by teaching experience versus those affected by advanced mathematics coursework. Evidence of good model fit across both instructor and student samples supports the use of the instrument to establish a unidimensional MKT-CCA score for both populations, including graduate students interested in teaching.

Discussion

We set out to investigate whether the instrument we developed measures mathematical knowledge for teaching college algebra at community colleges. We answer the question by providing multiple sources of evidence to support our intended interpretation and use of the MKT-CCA scores. We answer the question affirmatively.

First, evidence regarding the internal structure of the MKT-CCA (Studies 1 and 5) suggests that it measures an unidimensional construct. Despite our efforts to identify knowledge specifically related to the tasks of teaching, our modeling did not support the distinction. Moreover, while some pairs of types of functions were distinguishable, items from two pairs of functions, rational and linear were not. Further examination of pairs of items that are highly correlated across types

of function may be necessary to identify the underlying connections across items (e.g., in some exponential function items the options referred to linear growth). For our purposes, and given the content covered by the items, as corroborated by the cognitive interviews, and the finding that teaching advanced courses that build on the content in college algebra, we believe that the instrument does assess mathematical knowledge of college algebra as taught at community colleges.

While we built on the work by Ko and Herbst (2020) to propose a distinction between the knowledge needed for each task of teaching, a major difference is that our operational definition of choosing problems was more general than theirs (choosing givens for a geometric proof problem); this generality, together with the wider range of mathematical topics (linear, rational, and exponential functions, as opposed to proofs in geometry) might have resulted in the difficulty in differentiating these tasks of teaching with our instrument.

We demonstrated that a unidimensional measure of MKT-CCA is applicable to the student sample, and that the estimated knowledge scale is comparable across the combined sample. In addition, we were able to show that a different sample of non-teachers, students who have experienced the content included in the MKT-CCA but who declare having no teaching or tutoring experience, perform below our target sample of community college instructors. This signals that the instrument is assessing some component of knowledge that the instructors, but not the students, possess that may be related to their teaching experiences. The sampling we chose does not totally rule out the possibility that the higher performance of instructors relative to the students is that the instructors may simply have more mathematical knowledge than the students. The sampling also does not differentiate between the students and instructors when answering choosing problems or understanding student work questions. As the students did not

have teaching experience it would be their mathematics knowledge and their approach as a learner that would inform their decision making about choosing problems or understanding student work. This could have added to the MKT-CCA gap between the instructors and the students. A future study will seek to rule out this possibility, by recruiting math majors with different levels of teaching expertise. Nevertheless, we believe that these results could enhance training and curriculum development for students, who might eventually be interested in teaching mathematics at the community college or possibly at other post-secondary institutions.

Evidence based on process (Study 2), suggests that instructors used the mathematical knowledge that was hypothesized during the item design process and that they recognized the two specific tasks of teaching underlying the items. Therefore, the items, individually and collectively, draw on knowledge needed for teaching these college algebra topics, as situated in activities related to understanding students' work and choosing problems.

Evidence related to connections to other variables (Studies 3, 4, and 5) is also positive. We confirmed that the kind of teaching experience that instructors have is related to mathematical knowledge for teaching. Teaching many courses alone is not related to gains in MKT, but teaching courses that are beyond the content included in the test does result in gains in MKT. This corroborates findings with the community college population (Ko *et al.*, 2024). As instructors experience teaching advanced courses, student difficulties with the content addressed by the instrument become more relevant when they emerge, which may provide additional ways of understanding that might not be foreseen just by teaching college algebra alone. As they teach the more advanced courses, they engage with student work that signals different ways of thinking about the content and might be pressed to select problems that will address those different ways of thinking.

We found that, after controlling for teaching advanced mathematics courses, instructors who had PhD in mathematics as their highest degree performed better than instructors without this degree. This suggests that our instrument measures a mix of knowledge, not just for teaching college algebra but possibly another broader mathematical knowledge that would be acquired by going through a doctoral program.

It is notable that experience teaching post-college algebra courses and participation in mathematics-specific professional development showed significant effects on MKT-CCA, while general teaching experience and general professional development participation did not. This result implies that the measured MKT-CCA captures knowledge specific to the mathematical content. Interestingly, when instructors took mathematics professional development more than three times a year, it showed a significant effect on the MKT-CCA, relative to instructors who did less than that. Research on professional development has documented that sustained engagement with professional development can improve quality of instruction (Garet et al., 2001; Goos et al., 2018; Wilson & Berne, 1999), our findings indicate that sustained mathematics specific professional development will be beneficial for this population of instructors.

After controlling for teaching more advanced courses, and other variables in the model (Table 4), we found that full-time status for instructors showed a significant effect on the MKT-CCA. Full-time instructors are hired to teach between 12 and 15 credits of mathematics per semester, giving them continued access to many students every semester, students who will bring different needs that will be evident in their written work and in their classroom interaction. Part-time instructors may not have the same opportunities to teach as many courses or students as full-time instructors and may be supplementing other professional activities with a position at a local community college.

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

The unexpected effect of being White or Hispanic/Latino may suggest potential underlying structural factors that would need further investigation. However, given this unbalanced sample (76% of the sample identified as White and only 13% as Hispanic), these effect estimations may not be reliable enough to capture the full variability of the smaller group, so the result should be interpreted cautiously.

We found that the frequency factor of choosing problems (CP) (Study 4) was significantly associated with MKT-CCA scores, whereas the understanding student work frequency factor (USW) was not, even though these two were highly correlated (Figure 3). That choosing problems predicts MKT-CCA corroborates our theoretical stance regarding the importance of these tasks; having experience choosing problems gives instructors an advantage beyond just knowing the mathematics, suggesting that the knowledge that we are assessing is grounded in their practice. One possible explanation is that understanding student work is a more common task of teaching, whereas in choosing problems, there is more variation for instructors: Some instructors may use already given problems in their curriculum whereas others may create their own problems (Figure 2). If understanding student work is a more common task of teaching there will be less variability in the USW factor, which may lead to having insufficient statistical power to detect significance. In a future study we may need more items addressing USW and a larger sample of instructors that would add heterogeneity and variability in the frequency of engaging in this task. The high correlation between the two frequency factors states that the more instructors engage in understanding student work the more they engage in choosing problems and vice versa; therefore, the two tasks inform each other; this is reasonable and expected. But they are also distinguishable as factors with only one of them having a significant effect on MKT-CCA, so there is evidence that instructors do perceive these as different activities.

Interpretation and Use Statement

Carney *et al.* (2022) proposed specifying ten elements of the Interpretation and Use Statements (IUS) to provide information that helps end users—particularly within the mathematics education research community—determine whether the score interpretation from an instrument aligns with their intended use (p. 336). These elements include questions related to construct articulation, operationalization and administration, and score interpretation and use.

Following the recommendation, we describe our developed MKT-CCA instrument according to each IUS element:

> The MKT-CCA instrument measures the amount of mathematical knowledge specific to teaching college algebra at community colleges. This knowledge is related to specific tasks of teaching, to teaching mathematics content associated with three function types, and to experience in teaching courses beyond college algebra. (E1: What construct is being measured?). Measuring this construct is important because the instrument is tailored to the community college population and is grounded in the content of a course that is required for students who seek to complete a STEM major. Ascertaining that instructors who teach the course have the knowledge needed to teach it is crucial for professional development and teacher preparation (E2: Why is the construct important to measure?). The MKT-CCA consists of 34 items, including 27 multiple-choice questions and seven testlets (each containing four True/False items, resulting in 55 scored responses) (E3: How is the construct measured?). The target population of this instrument are community college instructors who teach college algebra (E4: Who is the target population?). The MKT-CCA is administered through online Qualtrics. Participants are

instructed not to use external aids or technology and to allocate at least two hours for completion, with the option to take breaks. The test is not proctored (E5: What is the context for administration?). There are no costs associated with administering the instrument. However, currently, it is only available for research purposes (E6: What costs are associated with the instrument?). Each item is scored dichotomously (0=incorrect, 1=correct), and sub-items in a testlet are scored as the sum of the number correct. A total knowledge score is estimated by a unidimensional IRT model, standardized to have a mean of 0 and standard deviation 1 (E7: How are scores calculated?). The score distribution is centered on 0, as the average; negative scores indicate below-average knowledge and positive scores indicate above-average knowledge (E8: How should scores be interpreted?). The instrument is proposed to be used to assess changes in knowledge after engaging in focused professional development (E9: How should scores be used?). The instrument should not be used for evaluation of teaching, nor for making hiring decisions, as such uses have not been tested (E10: What cautions or warnings should be considered?).

Conclusion

The collected evidence from the analyses presented allows us to claim that, related to our interpretation and use, the instrument measures community college instructors' mathematical knowledge for teaching college algebra. Specifically, the knowledge is used in doing specific tasks of teaching, teaching mathematics content associated with three function types, and is related to experience in teaching courses beyond college algebra. The instrument is proposed to be used to assess changes in knowledge after engaging in focused professional development. We

provided evidence based on the content of the instrument: coverage, expert assessment (for item design and map to the blueprint), on response processes, internal structure (Studies 1 and 5), and evidence related to other variables (e.g., teaching advanced mathematics courses, Study 3; the frequency of use of tasks of teaching, Study 4; and higher score for instructors than students, Study 5).

We have also provided strong rationale for the use of the variables collected, based on our theorization of the knowledge construct. Moreover, as presented in the above sections, we made all efforts possible to ensure representativeness of the samples of test takers and a similar testing environment. Likewise, we have fully reported all the corrections and adjustments needed in the statistical procedures. Because of the nature of the items, we do not have concerns about the scoring and as at this time the instrument is intended solely for research purposes, training or scoring principles are well known by the research team. Much work however is needed to augment the usability of the instrument. For example, it would be important to establish that the scores of this instrument converge with measures of the same construct (e.g., the Algebra MKT instrument designed for school teachers (e.g., the Algebra MKT instrument designed for school teachers, Ko *et al.*, 2024) and that it diverges from measures of different constructs (e.g., teachers' beliefs). At this stage we still envision the instrument to be applied to assess the extent to which knowledge changes over time, and as such, we believe that it will be a key tool to assess the extent to which professional development that targets knowledge of teaching college algebra courses that hones into attending to specific content areas and tasks of teaching in the content of those content areas, could be useful. The feedback that we received from instructors who were involved in the development process, from writing items, providing feedback on items,

answering the items, and answering cognitive interviews suggest that the nature of the items reflected their work and targeted important and foundational content in college algebra.

Acknowledgements

Funding for this work was provided by the National Science Foundation, EHR Core Awards #2000602 (Michigan), #2000644 (Maricopa), 2000527 (Minnesota), and #2000566 (Oregon). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Foundation. Special thanks to Claire Boeck, Siyad Gedi, and Kim Wingert. Thanks to the participants; without them this work would not have been possible.

References

Aguilar-Smith, S., & Gonzales, L. D. (2021). A study of community college faculty work expectations: Generous educators and their managed generosity. *Community College Journal of Research and Practice*, *45*(3), 184–204. https://doi.org/10.1080/10668926.2019.1666062

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. American Educational Research Association.

Ball, D. L., & Forzani, F. M. (2007). What makes education research "educational"? *Educational Researcher*, *36*(9), 529–540.

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

Blair, R., & Kirkman, E. E. (2025). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States. Fall 2021 CBMS Survey*. American Mathematical Society. https://www.ams.org/learning-careers/data/cbms-survey/cbms2021

Bostic, J. D., Sondergeld, T. A., Matney, G., Stone, G., & Hicks, T. (2021). Gathering response process data for a problem-solving measure through whole-class think alouds. *Applied Measurement in Education*, *34*(1), 46–60.

Calcagno, J. C., Bailey, T., Jenkins, D., Kienzl, G., & Leinbach, T. (2008). Community college student success: What institutional characteristics make a difference? *Economics of Education Review*, *27*(6), 632–645. https://doi.org/https://doi.org/10.1016/j.econedurev.2007.07.003

Carney, M. B., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics education. *Journal for Research in Mathematics Education*, *53*(4), 334–340. https://doi.org/10.5951/jresematheduc-2020-0087

Carney, M. B., Brendefur, J. L., Thiede, K., Hughes, G., & Sutton, J. (2016). Statewide mathematics professional development: Teacher knowledge, self-efficacy, and beliefs. *Educational Policy*, *30*(4), 539–572. https://doi.org/10.1177/0895904814550075

Charalambous, C. Y., & Hill, H. C. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship. *Journal of Curriculum Studies*, *44*(4), 443–466. https://doi.org/10.1080/00220272.2011.650215

Charalambous, C. Y., Hill, H. C., Chin, M. J., & McGinn, D. (2020). Mathematical content knowledge and knowledge for teaching: Exploring their distinguishability and contribution to student learning. *Journal of Mathematics Teacher Education*, *23*(6), 579–613.

Chin, E.-T., Lin, Y.-C., & Tuan, H.-L. (2016). Analyzing changes in four teachers' knowledge and practice of inquiry-based mathematics teaching. *The Asia-Pacific Education Researcher*, *25*(5), 845–862. https://doi.org/10.1007/s40299-016-0304-3

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Confrey, J., & Smith, E. (1994). Exponential functions, rates of change, and the multiplicative unit. *Educational Studies in Mathematics*, *26*, 135–164.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, *95*(2), 158–165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human resources*, *42*(3), 528–554.

Desimone, L. M., Hochberg, E. D., & Mcmaken, J. (2016). Teacher knowledge and instructional quality of beginning teachers: growth and linkages. *Teachers College Record*, *118*(5), 1–54. https://doi.org/10.1177/016146811611800507

Educational Testing Service. (2015). *2014 ETS standards for quality and fairness*. Educational Testing Service.

Ellis, A. B. (2007). Connections between generalizing and justifying: Students' reasoning with linear relationships. *Journal for Research in Mathematics Education*, *38*(3), 194–229.

Finnegan, C. L. (2019). The multiple roles of faculty in supporting community college students. *New Directions for Community Colleges*, *2019*(187), 63–72. https://doi.org/https://doi.org/10.1002/cc.20370

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Goos, M., Bennison, A., & Profitt-White, R. (2018). Sustaining and scaling up research-based professional learning for mathematics teachers *Mathematics Teacher Education and Development*, *20*(2), 133–150. https://mted.merga.net.au/index.php/mted/article/view/430

Herbst, P., & Chazan, D. (2011). Research on practical rationality: Studying the justifications of actions in mathematics teaching. *The Mathematics Enthusiast*, *8*(3), 405–462. https://doi.org/10.54870/1551-3440.1225

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. https://doi.org/https://doi.org/10.1007/s11858-012-0438-6

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *41*(5), 513–545.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*(4), 372–400.

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

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2-3), 81–92. https://doi.org/10.1080/15366360701486999

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30.

Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education*, *118*(4), 489–519.

Holm, J., & Kajander, A. (2020). Seeking intersections: Math degrees, beliefs, and elementary teacher knowledge. *Canadian Journal of Science, Mathematics and Technology Education*, *20*(1), 27–41. https://doi.org/10.1007/s42330-019-00069-3

Jurdak, M. E., & El Mouhayar, R. R. (2014). Trends in the development of student level of reasoning in pattern generalization tasks across grade level. *Educational Studies in Mathematics*, *85*, 75–92.

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (pp. 17–64). American Council on Education and Praeger.

Kane, M. (2012). Validating score interpretations and uses. *Language testing*, *29*(1), 3–17.

Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (Vol. 10). SAGE publications.

Kisker, C. B., Cohen, A. M., & Brawer, F. B. (2023). *The American community college*. John Wiley & Sons.

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. https://doi.org/https://doi.org/10.5951/jresematheduc-2020-0163

Ko, I., Mesa, V., Duranczyk, I., Herbst, P., Kohli, N., Ström, A., & Watkins, L. (2024). 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*, *55*, 590–614. https://doi.org/10.1080/0020739X.2021.2006348

Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, *66*(1), 51–67. https://doi.org/10.1177/0022487114549599

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 – Mathematics Education*, *40*(5), 873–892. https://doi.org/https://doi.org/10.1007/s11858-008-0141-9

Lobato, J., & Siebert, D. (2002). Quantitative reasoning in a reconceived view of transfer. *The Journal of Mathematical Behavior*, *21*(1), 87–116. https://doi.org/10.1016/S0732-3123(02)00105-0

Marshall, J. C., Horton, R., Igo, B. L., & Switzer, D. M. (2009). K-12 science and mathematics teachers' beliefs about and use of inquiry in the classroom. *International Journal of Science and Mathematics Education*, *7*(3), 575–596. https://doi.org/10.1007/s10763-007-9122-7

Mayes, R. (1993). Asymptotes and other limiting functions. *The Mathematics Teacher*, *86*(5), 411-414.

Mesa, V. (2017). Mathematics education at public two-year colleges. In J. Cai (Ed.), *First compendium for research in mathematics education* (pp. 949–967). National Council of Teachers of Mathematics.

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. In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)* (pp. 3843–3849). Alfréd Rényi Institute of Mathematics and ERME.

Mesa, V., Lamm, R., Watkins, L., Duranczyk, I., Ström, A., & Kohli, N. (2020). A confirmatory factor analysis of EQIPM, a video coding protocol to assess the quality of community college algebra instruction. 23rd Annual Conference on Research on Undergraduate Mathematics Education, Boston, MA.

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics*, *99*(2), 486–496. https://doi.org/https://doi.org/10.1016/j.jdeveco.2012.06.002

Moschkovich, J., Schoenfeld, A., & Arcavi, A. (1993). Aspects of understanding: On multiple perspectives and representations of linear relations and connections among them. In T. A. Romberg, T. Carpenter, & E. Fennema (Eds.), *Integrating research on the graphical representation of functions* (pp. 69–100). Erlbaum.

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

Nabb, K., & Murawska, J. (2019). Mathematical knowledge for teaching developmental courses at the community college: An unexplored terrain. *Community College Journal of Research and Practice*, *44*(8), 563–583. https://doi.org/ https://doi.org/10.1080/10668926.2019.1626300

Nair, G. S. (2010). *College students' concept images of asymptotes, limits, and continuity of rational functions* (Publication Number 3435663) [Ph.D., The Ohio State University]. Dissertations & Theses @ Big Ten Academic Alliance; ProQuest Dissertations & Theses Global; ProQuest Dissertations & Theses Global Closed Collection. United States, Ohio.

Österholm, M., Bergqvist, T., Liljekvist, Y., & van Bommel, J. (2024). Professional development at national scale: Effects on teacher knowledge and practice. *Studies in Educational Evaluation*, *83*, 101381. https://doi.org/https://doi.org/10.1016/j.stueduc.2024.101381

Phelps, G., Weren, B., Croft, A., & Gitomer, D. (2014). Developing content knowledge for teaching assessments for the Measures of Effective Teaching study. *ETS Research Report Series*, *2014*(2), 1–92.

Postelnicu, V. (2011). *Student difficulties with linearity and linear functions and teachers' understanding of student difficulties* (Unpublished doctoral dissertation ed.). Arizona State University.

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/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: Interdisciplinary Research and Perspectives*, *5*(2-3), 93–106. https://doi.org/10.1080/15366360701487021

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

Speer, N., King, K. D., & Howell, H. (2015). Definitions of mathematical knowledge for teaching: Using these constructs in research on secondary and college mathematics teachers. *Journal of Mathematics Teacher Education*, *18*(2), 105–122.

Sproesser, U., Vogel, M., Dörfler, T., & Eichler, A. (2022). Changing between representations of elementary functions: students' competencies and differences with a specific perspective

on school track and gender. *International Journal of STEM Education*, *9*(1), 33. https://doi.org/10.1186/s40594-022-00350-2

Ström, A. D. (2006). The role of covariational reasoning in learning and understanding exponential functions. *PME-NA 2006 Proceedings*, *2*, 624–630.

Ström, A. D. (2008). *A case study of a secondary mathematics teacher's understanding of exponential function: An emerging theoretical framework* (Unpublished doctoral dissertation ed.) [Ph.D.]. Arizona State University.

Stump, S. L. (2001). Developing preservice teachers' pedagogical content knowledge of slope. *The Journal of Mathematical Behavior*, *20*(2), 207–227. https://doi.org/https://doi.org/10.1016/S0732-3123(01)00071-2

Suhr, D. D. (2006). *Exploratory or confirmatory factor analysis*. SAS Institute.

Thompson, P., & Carlson, M. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 421–456). National Council of Teachers of Mathematics.

Tuda, S., & Rexhepi, S. (2023). Exploring exponential functions using GeoGebra. *Brillo Journal*, *3*(1), 43–58. https://doi.org/10.56773/bj.v3i1.45

Van Dooren, W., De Bock, D., Verschaffel, L., & Janssens, D. (2003). Improper applications of proportional reasoning. *Mathematics Teaching in the Middle School*, *9*(4), 204–209.

Van Dooren, W., Vamvakoussi, X., & Verschaffel, L. (2018). *Proportional reasoning* (Vol. 30). UNESCO, International Academy of Education, & International Bureau of Education.

Wilkie, K. J., & Ayalon, M. (2018). Investigating years 7 to 12 students' knowledge of linear relationships through different contexts and representations. *Mathematics Education Research Journal*, *30*, 499–523.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: an examination of research on contemporary professional development. *Review of Research in Education*, *24*(1), 173–209.

Wladis, C., Offenholley, K., Licwinko, S., Dawes, D., & Lee, J. (2018). Development of the elementary algebra concept inventory for the college context. Research in Undergraduate Mathematics Education (RUME) Conference, , San Diego, CA.

Yerushalmi, M. (1997). Emergence of new schemes for solving algebra word problems: The impact of technology and the function approach. In E. Pehkonen (Ed.), *Proceedings of the 21st conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 165–178). University of Helsinki.

Appendix

We present three items, one from each type of function and on distinct tasks of teaching as illustrative of the items in the instrument. We include the difficulty for the instructor and the student sample.

Choosing problems multiple-choice item on rational functions. Instructor difficulty 0.69; Student difficulty: 0.45

Instructor Jenkins wants to use the following set of rational functions to show that attending to the structure of the symbolic representation of a rational function can reveal properties of the function.

$$f(x) = \frac{x^2+2x-3}{x-1} \qquad g(x) = \frac{x^2+2x-3}{x^2-1} \qquad h(x) = \frac{x^2}{x^2-1}$$

Of the options below, select the option that is an **INACCURATE** description of what this set of functions affords.

- ○ The set illustrates the difference between functions that have vertical asymptotes and functions that have holes.
- ○ The set illustrates rational functions with different ranges.
- ○ The set illustrates that when the degree of the denominator is the same as the degree of the numerator then the function will have a hole.
- ○ The set illustrates that some rational functions can be simplified to a function that looks like a linear function.

DO NOT CITE WITHOUT PERMISSION FROM AUTHORS

Choosing problem multiple-choice item on linear functions. Instructor difficulty: 0.51; Student difficulty: 0.31

Instructor Owens gave students the following problem:

> *Green Lawn Care Company charges $30 per month with a $15 initial fee for mowing their lawn. The total expense in dollars for mowing their lawn after x months can be modeled by $f(x) = 30x + 15$. Find the inverse function and interpret the meaning of both $f(x)$ and its inverse.*

Lucinda, a student, found the inverse of $f(x)$ correctly and said: "For $f(x)$, the total expense increases each month. At one month, the service costs $45. For the inverse, at that month, the cost is... wait, how can it be negative? That does not make any sense!"

Several students agreed with Lucinda. Instructor Owens wants to assign the class a mathematical task to work on in pairs to address the concern raised by Lucinda. Indicate which of the following tasks is the **LEAST** likely to help resolve this issue.

- ○ Graph $f(x)$ and its inverse on separate sets of axes. For each graph, consider what labels the axes should have.
- ○ For $f(x)$, begin by looking at the x-value on its graph and then find the corresponding y-value to construct your interpretation. For the inverse of f, look at the y-value on the graph of $f(x)$ then find the corresponding x-value.
- ○ Graph both functions in the same Cartesian plane and describe the relationship between the slope of $f(x)$ and the slope of its inverse.
- ○ Find $f(1)$, $f(2)$, $f(45)$ and $f(75)$ and $f^{-1}(1)$, $f^{-1}(2)$, $f^{-1}(45)$ and $f^{-1}(75)$ and explain what the input represents in each function.

Understanding student work multiple choice item in the context of exponential functions.

Instructor difficulty 0.43; Student difficulty 0.20.

To gauge their progress in finding multiplicative rates of change for exponential functions, Professor Monoz gave students the following table of values illustrating exponential population growth of bacteria as a function of time. He asked them to find the multiplicative rate at which the function is changing between $x = 3$ and $x = 4$ and explain how they got their answer:

x Hours in Culture	$f(x)$ Number of Bacteria (in millions)
1	1.5
3	6.0
5	24

Which of the following student responses demonstrates a correct process for finding the multiplicative rate of change (in the context of exponential growth)?

- Nancy: "Since the inputs are always changing by 2, I calculated the multiplicative rate of change by finding the difference between the output values of 24 and 6 then multiplied by 2."
- Arum: "Since the inputs are always changing by 2, I calculated the multiplicative rate of change by finding the ratio of the corresponding output values, 24 and 6."
- Rosa: "I calculated the multiplicative rate of change by first finding the ratio between 6 and 24, and then dividing by the difference between $x = 3$ and $x = 5$."
- Sean: "I calculated the multiplicative rate of change by finding the ratio between 24 and 6 and then finding the square root of this ratio."