Educational Studies in Mathematics (2024) 115:387–406
https://doi.org/10.1007/s10649-024-10299-4

Proposing and testing a model relating students' graph selection and graph reasoning for dynamic situations

Heather Lynn Johnson[1] · **Courtney Donovan**[1] · **Robert Knurek**[1] · **Kristin A. Whitmore**[1] · **Livvia Bechtold**[1]

Accepted: 7 January 2024 / Published online: 15 February 2024
© The Author(s) 2024

Abstract
Using a mixed methods approach, we explore a relationship between students' graph reasoning and graph selection via a fully online assessment. Our population includes 673 students enrolled in college algebra, an introductory undergraduate mathematics course, across four U.S. postsecondary institutions. The assessment is accessible on computers, tablets, and mobile phones. There are six items; for each, students are to view a video animation of a dynamic situation (e.g., a toy car moving along a square track), declare their understanding of the situation, select a Cartesian graph to represent a relationship between given attributes in the situation, and enter text to explain their graph choice. To theorize students' graph reasoning, we draw on Thompson's theory of quantitative reasoning, which explains students' conceptions of attributes as being possible to measure. To code students' written responses, we appeal to Johnson and colleagues' graph reasoning framework, which distinguishes students' quantitative reasoning about one or more attributes capable of varying (Covariation, Variation) from students' reasoning about observable elements in a situation (Motion, Iconic). Quantitizing those qualitative codes, we examine connections between the latent variables of students' graph reasoning and graph selection. Using structural equation modeling, we report a significant finding: Students' graph reasoning explains 40% of the variance in their graph selection (standardized regression weight is 0.64, $p < 0.001$). Furthermore, our results demonstrate that students' quantitative forms of graph reasoning (i.e., variational and covariational reasoning) influence the accuracy of their graph selection.

Keywords Quantitative reasoning · Graphs · Structural equation modeling · Mixed methods · College algebra · Covariation · Variation · Qualitative analysis

By conducting studies that employ quantitative methods, mathematics education researchers can test hypotheses generated from smaller scale studies, and thereby work to engender systemic change (Kilpatrick, 2001). Developing and validating assessments, researchers

✉ Heather Lynn Johnson
heather.johnson@ucdenver.edu

[1] School of Education and Human Development, University of Colorado Denver, Campus Box 106, P.O. Box 173364, Denver, CO 80217-3364, USA

have tested hypotheses related to students' whole number multiplicative reasoning (Kosko, 2019; Tzur et al., 2022), and students' and teachers' fractional (Izsák et al., 2019; Norton & Wilkins, 2009; Pantziara & Philippou, 2012) and covariational reasoning (Carlson et al., 2010; Thompson et al., 2017). With this mixed methods study, we expand these efforts to examine undergraduate students' graph reasoning, with a focus on quantitative and covariational reasoning (Carlson et al., 2002; Thompson, 1994, 2022; Thompson & Carlson, 2017). The purpose of our study is to propose and test a conceptual model relating students' graph reasoning and graph selection on a fully online assessment.

Bakker et al. (2021) have identified online assessment as a research challenge deserving further attention in mathematics education. Our study addresses this challenge, investigating an understudied population, U.S. college algebra students. College algebra is one of the earliest credit-bearing U.S. undergraduate mathematics courses. It is notorious for being packed with content (Gordon, 2008), and many course textbooks emphasize the application of procedures over concept development (Mesa et al., 2012). Not only are graphs central to college algebra, the value of graph reasoning extends beyond school mathematics, to include interpretation of trends of change and variation depicted in graphs in the news media. In smaller scale qualitative studies, researchers have examined students' graph reasoning, spanning middle, secondary, advanced undergraduate students, and preservice mathematics teachers (e.g., Carlson et al., 2002; Ellis et al., 2016; Johnson et al., 2020; Moore et al., 2019b). Using an online assessment, we examine early undergraduate students' graph reasoning and draw connections to their graph selection.

We have developed and validated a fully online measure of students' graph selection and reasoning for dynamic situations (MGSRDS) (Johnson et al., 2018, 2021). By "dynamic situations," we mean situations in which some object is changing, such as a toy car moving along a square track. The six-item MGSRDS contains video animations of situations, followed by a set of graph choices and a text box to explain the graph choice. By "graph reasoning," we mean people's reasoning regarding the sketching and/or interpreting of graphs; for our purposes, we focus on Cartesian graphs. To operationalize the construct of graph reasoning, we use the framework from Johnson et al. (2020), which distinguishes students' quantitative reasoning about one or more attributes capable of varying from students' reasoning about observable elements in a situation. By "graph selection," we mean the process of choosing a graph given a fixed set of options. Students' graph choices and text responses on the MGSRDS serve as observable indicators for the larger constructs of graph selection and graph reasoning.

Our large sample size ($n = 673$) allows for the use of rigorous statistical techniques, such as structural equation modeling (SEM). SEM is a regression-based technique that allows researchers to analyze the correlational and predictive nature of latent variables, which are not directly observable (Byrne, 2016). Measurement is key to SEM; assessments that undergird researchers' use of SEM need to be valid and reliable measures of the constructs in their model. Using SEM, researchers can confirm and/or modify a conceptual model relating different variables, representing constructs grounded in theoretical and empirical research (Kline, 2023). It is a valuable technique to explore relationships that may be assumed (e.g., it may seem predictable for there to be a relationship between graph reasoning and selection), but not strongly supported with statistical models.

Our research question is: To what extent does evidence of students' graph reasoning relate to the accuracy of their graph selection? Two research hypotheses underlie this question. The first is broad: Students' graph reasoning will influence the accuracy of their graph selection on the MGSRDS. To test the first hypothesis, we use SEM. The second is more specific: Students' quantitative graph reasoning (i.e., variational and covariational

reasoning) will relate to more accurate graph selections. To test the second hypothesis, we use chi-square tests of independence.

To organize this article, we begin with theoretical and conceptual background to undergird our conceptual model. Next, we address the design and validation of the MGSRDS, to explain how we measure the constructs of graph reasoning and graph selection. Then we describe our methods and results, reporting significant findings for both research hypotheses. We conclude with discussion and implications.

1 Theoretical and conceptual background

Thompson's theory of quantitative reasoning (Thompson, 1994, 2022; Thompson & Carlson, 2017) explains quantities in terms of students' conceptions of attributes as being possible to measure. For example, a student may conceive of a toy car's distance traveled along a square track as something possible to measure. Per Thompson's theory, conceiving of some attribute as a quantity is not the same as finding amounts of measure of that attribute. For example, students may imagine the toy car leaving a trail of dust as it moves along and conceive of the car's total distance traveled as the length of that trail (see also Frank, 2017). Hence, students can engage in quantitative reasoning even if they do not actually measure some attribute of an object. Through quantitative reasoning, students mathematize attributes (Johnson, 2022), and can sketch or select graphs to represent relationships between those attributes.

1.1 Students' graph reasoning

Decades ago, researchers identified two early conceptions that students may have for graphs that represent attributes in dynamic situations (Clement, 1989; Kerslake, 1977; Leinhardt et al., 1990). One was that graphs represent an iconic interpretation of a physical situation (Clement, 1989; Leinhardt et al., 1990). For example, a student might expect a graph of a Ferris wheel situation to look round like the Ferris wheel itself. Another was that graphs would represent the physical motion in a dynamic situation (Kerslake, 1977). For example, a student might expect a graph representing a person's bike trip to show the literal path that the person rode. While both conceptions reflected students' attempts to make sense of graphs, the physical constraints imposed did not necessarily align with the structure of the Cartesian coordinate system, in which graphs could look quite different from objects or motion in a physical situation (see also Lee et al., 2020).

Working within the Cartesian coordinate system, students can distinguish attributes of physical objects represented in dynamic situations (e.g., a toy car's distance from a stationary object) and conceive of how they may measure those attributes and how those attributes may vary. Variational reasoning refers to students' conceptions of variation in a single attribute (Thompson & Carlson, 2017). For example, a student may conceive of how the height of a Ferris wheel cart can increase and decrease during a revolution of the wheel. Covariational reasoning refers to students' conceptions of relationships between attributes that they conceive to be possible to measure and capable of varying (Carlson et al., 2002; Thompson & Carlson, 2017). For example, a student may conceive of a relationship between the height of a Ferris wheel cart from the ground and the total distance the cart travels in one revolution of the wheel. Thompson and Carlson (2017) have posited frameworks identifying levels in students' variational and covariational reasoning. Gross

variation and gross coordination of values are two early levels in each framework. With gross variation, students can conceive of the direction of change in an attribute (e.g., height increases, then decreases). With gross coordination of values, students can conceive of a loose relationship between the direction of change in attributes (e.g., height increases and decreases while distance continues to increase). When we refer to variational or covariational reasoning, we mean there is evidence of at least gross variation or gross coordination of values.

Distinguishing attributes of objects could help students to sketch and interpret graphs, and to make sense of why graphs often have observable differences from physical features or motion in a dynamic situation (Johnson et al., 2020). To characterize students' graph reasoning, Johnson et al. (2020) put forward a framework, laying out four forms of graph reasoning for dynamic situations: Covariation, Variation, Motion, and Iconic. Covariation and Variation referred to at least a gross coordination of values or gross variation in an attribute, per Thompson and Carlson's (2017) frameworks. Covariation indicated reasoning in terms of relationships between quantities (e.g., a graph represents a relationship between height and distance). Variation indicated reasoning in terms of single quantities (e.g., a "height" graph), with experiential time (Thompson & Carlson, 2017) as an implicit variable. Motion and Iconic referred to the physical movement (Kerslake, 1977) or physical features (Clement, 1989; Leinhardt et al., 1990) in a situation. Motion indicated reasoning such that graphs resemble an object's physical movement in a dynamic situation (e.g., a graph shows a path of a moving object). Iconic indicated reasoning such that graphs resemble physical objects or observable attributes of those objects (e.g., if an object is curved, then a graph associated with that object also would curve). The framework represented a range of graph reasoning, distinguishing students' quantitative reasoning about one or more attributes capable of varying (Covariation, Variation) from students' reasoning about observable elements in a situation (Motion, Iconic).

1.2 Students' graph selection

On a task or assessment, students can engage in graph selection when they are faced with a fixed set of options from which to choose. The result of students' graph selection is the option that students pick, which an outside observer or instructor may deem to be correct or incorrect. Yet, graph selection contains processes that can be unobservable. Students may have preconceived notions about what a graph "should" look like, based on their own encounters with graphs in school or other media (Johnson et al., 2020). For instance, students may expect graphs to follow conventions to which they are accustomed (see also Moore et al., 2014; Moore et al., 2019a), even if the conventions are tacit for the students. Furthermore, students may eliminate graph options prior to selecting a single graph. In addition, students may not think any graph in a given set of choices is viable, but are forced to opt for one. The result of students' graph selection (i.e., their final answer) does not reveal the fuller process. Hence, we treat graph selection as a latent construct, including unobservable processes that contribute to an observable answer choice.

1.3 Relating students' graph selection and graph reasoning

Previously, researchers drew connections between undergraduate students' graph selection and their covariational reasoning. The Precalculus Concept Assessment (PCA) (Carlson et al., 2010), used to assess students' calculus readiness, had an intended population of

college algebra and precalculus students. The PCA contained a multiple choice graphing item, converted from an interview-based task. Given a static image of a spherical bottle, with written instructions to assume the bottle is filling with water at a constant rate, students were to select a graph that best represented a relationship between the height and amount of water. There were 5 graph choices, all monotonically increasing; two pairs had curved graphs with similar physical features but different concavities, and one graph was linear. In a validation study for the PCA, all students who selected the correct graph also demonstrated the ability to engage in covariational reasoning. Yet, there was a range of reasoning linked to students' graph selections.

Given the importance of quantitative and covariational reasoning for students' conceptions of key mathematical ideas, including function (Thompson & Carlson, 2017), and the prevalence of Cartesian graphs in college algebra, we have developed the MGSRDS. Because the MGSRDS is fully online, it allows for the incorporation of digital elements. In addition to written descriptions of dynamic situations, there are video animations of the situations. Furthermore, alongside graph choices, there is a text box for students to explain their graph choice. This has allowed us to examine a range of graph reasoning and link that reasoning to students' graph selection in the MGSRDS implementation.

2 The MGSRDS: design and validation

To use SEM, researchers must address issues of measurement, which includes establishing validity and reliability for measures of constructs involved (Kline, 2023). Next, we discuss the design and validation of the MGSRDS.

2.1 Measure design: assessing students' graph reasoning and selection

To assess students' graph reasoning and selection, we use the MGSRDS (Johnson et al., 2018, 2021). There are six assessment items, each containing a different dynamic situation: a turning Ferris wheel, a person walking toward a tree and back, a fishbowl filling with water, a toy car moving along a square track, a changing cone expanding and contracting, and two insects crawling back and forth along a path. Each item has four parts: a video animation depicting the situation, a student declaration of understanding (i.e., Do you understand this situation?), four graph choices, and a text box to explain the graph choice. The MGSRDS is accessible across mobile phones, tablets, and computers (Fig. 1), with each assessment item appearing in random order.

Table 1 shows the six assessment items (Ferris Wheel, Nat & Tree, Fishbowl, Changing Cone, Toy Car, Ant & Ladybug). For the first five items, one of the attributes varies in its direction of change (e.g., increases and decreases), while the other attribute's direction of change stays constant (e.g., only increases), and there is a correct and partially correct graph choice (one piecewise linear and one curved). Each partially correct graph choice represents gross covariation (Thompson & Carlson, 2017) in the attributes, but does not accurately represent a relationship between values of each attribute. To illustrate, on the Ferris Wheel item, the cart (represented by the small red dot, Table 1, upper left) moves counterclockwise for one revolution. Both graphs B (piecewise linear) and D (curved) represent gross covariation in the attributes (i.e., the cart's distance from the ground increases, decreases, then increases, while the cart's total distance traveled continues to increase). However, given a small change in distance, the associated amounts of height are greater

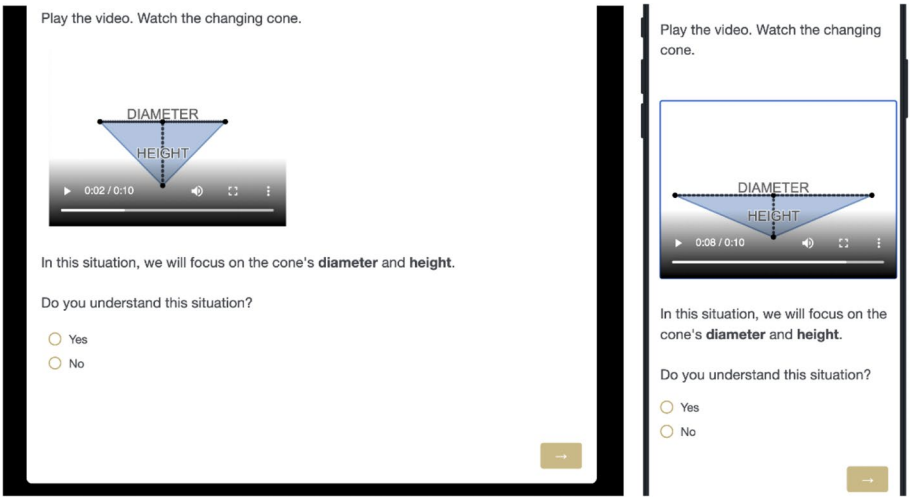
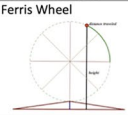
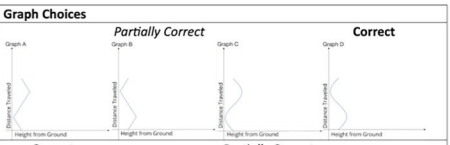
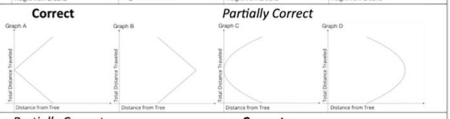
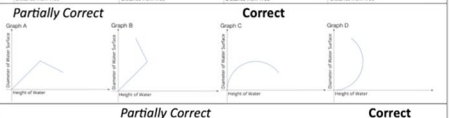
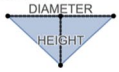
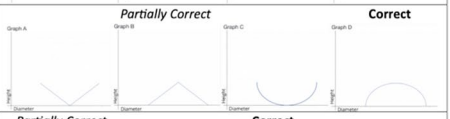
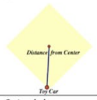
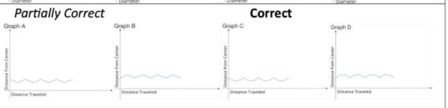
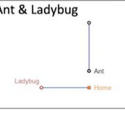
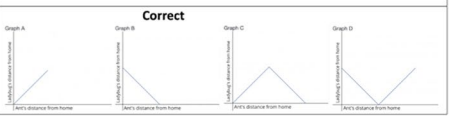


Fig. 1 Changing Cone item, optimized for computers/tablets (left) and mobile phones (right)

near the rightmost edge of the wheel and less near the top of the wheel. Hence, graph D is correct, and graph B is partially correct. The Ant & Ladybug item differs from the first five items; both attributes vary in their direction of change (i.e., both insects' distances from home increase and decrease), and there is no partially correct graph choice.

For each item, graph choices come in related pairs. For the first five items in Table 1, there are two piecewise linear graphs followed by two curved graphs. For the Ant &

Table 1 MGSRDS assessment items, directions, and graph choices

Assessment Item	Directions	Graph Choices
Ferris Wheel 	Select the graph that best represents a relationship between the Ferris wheel cart's **height from the ground** and the **distance traveled**, for one revolution of the Ferris wheel. Each graph shows the height from the ground on the horizontal axis and the distance traveled on the vertical axis. Explain why you chose the graph that you did.	*Partially Correct* **Correct** 
Nat & Tree	Select the graph that best represents a relationship between Nat's **distance from the tree** and Nat's **total distance traveled**. Each graph shows Nat's distance from the tree on the horizontal axis and Nat's total distance traveled on the vertical axis. Explain why you chose the graph that you did.	**Correct** *Partially Correct* 
Fishbowl	Select the graph that best represents a relationship between the **height of the water** and the **diameter of the water surface**. Each graph shows the height of the water on the horizontal axis and the diameter of the water surface on the vertical axis. Explain why you chose the graph that you did.	*Partially Correct* **Correct** 
Changing Cone 	Select the graph that best represents a relationship between the cone's **diameter** and **height**. Each graph shows the cone's diameter on the horizontal axis and cone's height on the vertical axis. Explain why you chose the graph that you did.	*Partially Correct* **Correct** 
Toy Car 	Select the graph that best represents a relationship between the toy car's **distance traveled around the track** and the toy car's **distance from the center of the track**. Each graph shows the toy car's distance traveled on the horizontal axis and toy car's distance from the center on the vertical axis. Explain why you chose the graph that you did.	*Partially Correct* **Correct** 
Ant & Ladybug 	Select the graph that best represents a relationship between the **ant's distance from home** and the **ladybug's distance from home**. Each graph shows the ant's distance from home on the horizontal axis and the ladybug's distance from home on the vertical axis. Explain why you chose the graph that you did.	**Correct** 

Ladybug item, there are two monotonic linear graphs followed by two piecewise linear graphs. This design move of paired graph choices shares similarities with graph choices on the PCA filling bottle item (Carlson et al., 2010). Across the MGSRDS items, the correct and partially correct choices span graphs A-D. Furthermore, to engender students' conceptions of change in progress, which Castillo-Garsow et al. (2013) call smooth images of change, none of the graph choices includes numerical amounts on the axes.

Theoretical considerations included whether objects in the situations (or observable attributes of those objects) resembled graphs (or portions of graphs) in the set of graph choices (see Johnson et al., 2020). Across the items, correct and partially correct graph choices never resembled the objects themselves, to guard against iconic reasoning leading to a correct graph choice (e.g., The incorrect graph A in the Changing Cone item resembles the cone itself.). For three items (Ferris Wheel, Nat & Tree, Fishbowl), portions of the correct graphs shared similarities with observable attributes of objects in the situations. The Ferris wheel and fishbowl had curved portions, like the correct graphs, and Nat's path was straight, like the correct graph. These items included unconventional graphs (see Moore et al., 2014; Moore et al., 2019a) as all or some of the graph choices, to guard against students selecting a correct graph just because it looked familiar. In each, the attribute whose direction of change was constant (e.g., distance traveled) was represented on the vertical axis, which contrasted with typical graphs representing elapsing time on the horizontal axis. For the Changing Cone and Toy Car items, the correct graphs did not share similarities with observable attributes of objects in the situations. The changing cone and the toy car's track had no curved portions, but the correct graphs were curved. These items included more conventional graphs, with the attribute whose direction of change was constant (e.g., diameter) being represented on the horizontal axis.

Theoretical considerations also included whether graph choices represented the physical movement and/or locations of objects in the animation (see Johnson et al., 2020), which is not a necessary feature of the Cartesian coordinate system. For the Nat & Tree item, the turning point on the two incorrect graphs (B and D) occurred in the right hand portion of the coordinate plane, similar to the location of the tree in the animation. However, if Nat were to move toward the tree, Nat's distance from the tree would shrink, not grow, as represented by graphs B and D. For the Ant & Ladybug item, the insects moved back and forth along straight paths, yet the correct graph was a single line segment. Furthermore, the incorrect graphs C and D showed "out and back" portions, like the insects that moved toward and away from home.

We view our theoretical considerations to share some overlap with students' spatial conceptions of coordinate systems (Lee et al., 2020; Paoletti et al., 2022), in which students conceive of coordinate systems as representing the physical space in which phenomena occur. For example, imagine coordinate systems overlaid on the dynamic situations shown in Table 1. For the Nat & Tree item, the lower right portion of the incorrect graph B resembles Nat's path to the tree. Furthermore, like Paoletti et al. (2022) contend, we assume that there is not necessarily a one-to-one correspondence between a particular graph and a form of student's graph reasoning. Therefore, while we anticipate that particular graph choices likely associate with certain forms of graph reasoning (e.g., Students engaging in iconic reasoning seem more likely to select the incorrect graph D for the Fishbowl because it resembles the contour of the bowl.), we leave room for different possibilities.

2.2 Establishing validity and reliability of the MGSRDS: setting the stage for SEM

Following Devellis' (2003) process, we established validity and reliability for the MGSRDS (Donovan et al., accepted). Our steps included field administration, exploratory factor analysis, and Rasch modeling. We showed content validity through analysis of individual student interviews ($n = 31$) conducted via videoconference, with Johnson serving as the interviewer (Johnson et al., 2021). During the interviews, each student worked on the MGSRDS and talked about their thinking. A graduate research assistant (GRA) attended each interview to observe and take field notes and later produced verbatim transcripts. Through exploratory factor analysis (Comrey & Lee, 1992), we confirmed the validity of the internal structure of the MGSRDS. Via Rasch modeling (Bond & Fox, 2001), we gathered evidence to support valid and reliable scales of students' graph reasoning, graph selection, and item difficulty.

Table 2 shows the five graph reasoning codes used for qualitative analysis: Covariation (COV), Variation (VAR), Motion (MO), Iconic (IC), and Limited Evidence (LE). The first four codes align with the graph reasoning framework from Johnson et al. (2020). The fifth code describes responses providing limited evidence of reasoning. For each, there is a description of the code and a sample response to illustrate.

To show content validity, we gathered evidence that the MGSRDS was measuring the intended constructs of graph selection and reasoning. For graph selection, we used the correctness of students' graph choices for each item. For graph reasoning, we conducted qualitative analysis of student interviews using the codes in Table 2. This resulted in a total of 186 codes (31 students * 6 items/student). We used a consensus coding process, which was more time consuming, but supported deeper accuracy (Olson et al., 2016). First, Johnson coded students' reasoning on each item. Then, to guard against expectation bias, two GRAs also independently coded students' reasoning on each item. After independent coding, the team met to calibrate codes and come to agreement.

Conducting Rasch modeling (Bond & Fox, 2001), we tested the theoretical ordering of the graph reasoning framework from Johnson et al. (2020), the coding for graph selection, and the item difficulty levels (Donovan et al., accepted). Results showed that the graph reasoning codes form a hierarchical scale in the order theoretically proposed by Johnson et al. (2020): COV, VAR, MO, and IC. Responses demonstrating LE (e.g., "Makes the most sense to me") worked well as a zero code to form a five code scale. To create a graph selection scale, coding for three levels (incorrect/partially correct/correct) worked better than coding for two levels only (incorrect/correct). Results showed that the graph selection codes (incorrect/partially correct/correct) form a hierarchical scale. We organized the items in Table 1 in the order of difficulty level of graph selection, with the Ferris Wheel being the easiest to select the correct graph and the Ant & Ladybug being the most difficult.

Reliability refers to how consistently a tool or person is being measured; Cronbach's alpha is the standard analysis to explore internal consistency of items for a person's response (Tavakol & Dennick, 2011). The graph selection scale has an alpha of 0.55, indicating low reliability. This means that across items, students are selecting a range of correct and incorrect graph choices. The graph reasoning scale has an alpha of 0.83, indicating strong reliability across a student's written response and the qualitative reasoning code. By establishing reliability and validity for the MGSRDS, we demonstrate the viability of our use of SEM.

Table 2 Graph reasoning codes, adapted from the framework from Johnson et al. (2020)

Reasoning Code	Description	Sample response from the MGSRDS
Covariation (COV)	Students coordinate amounts and/or directions of change in one quantity with direction(s) of change in another quantity	The diameter keeps increasing while the height increases and then decreases. (Changing Cone)
Variation (VAR)	Students conceive of variation in the direction of change and/or amounts of change in a single quantity	I chose that graph because the ants distance from home decreases to zero and then rises again. (Ant & Ladybug)
Motion (MO)	Students conceive of a graph as a close approximation (or literal representation) of the movement of a physical object	She goes and returns like in the graph I picked. (Nat & Tree)
Iconic (IC)	Students conceive of a graph as a close approximation (or literal representation) of the shape of a physical object	It is a square and will have edges in the graph not curves. (Toy Car)
Limited Evidence (LE)	Students provide limited evidence of their reasoning. They may state that they don't know, they guessed, they chose the best option, or provide an off task response	I picked graph b because it accurately represents the video shown. (Fishbowl)

3 Methods

This study is part of a larger, U.S. National Science Foundation-funded research project to promote mathematical reasoning and instructional transformation in college algebra. Using mixed methods, we integrate qualitative and quantitative approaches in our design and analysis (Tashakkori & Creswell, 2007). Per Leech and Onwuegbuzie's (2009) typology, our study is a fully mixed, concurrent, equal status design. This means that we mix qualitative and quantitative approaches across the research objective, the types of data, the data analysis processes, and the inferences drawn from analyses. Figure 2 shows the logic model for our study.

3.1 Methodological background: SEM

SEM is based on classical test theory (DeVellis, 2006), where one assumes that a person's observed score (e.g., their written response explaining their reason for selecting a graph) is equal to their true score (e.g., their actual graph reasoning) plus error (e.g., the gap between what they wrote and their actual reasoning ability). Unlike multiple regression techniques that "wash out" such variance by forcing a mean or total score, SEM accounts for variance between student responses across items (Kline, 2023). Furthermore, SEM is a rigorous statistical technique, requiring a larger sample size than multiple regression models. A benefit of structural equation models is that they provide statistical evidence to support the viability of larger, systematic patterns stemming from qualitative studies or smaller quantitative studies.

SEM must be grounded in theory, whereas multiple regression and other intermediate statistical analysis can explore connections without strongly supported theoretical underpinnings (Kline, 2023). After creating a model based on theory, researchers calculate statistics to examine model fit, to ensure that patterns in the data fit the patterns in the hypothesized model. If model fit is not satisfactory, researchers then adapt or reevaluate the model. A goal of SEM is to provide a model that makes theoretical sense, is as bare-bones as possible, and has an acceptably close correspondence to the data (Kline, 2023).

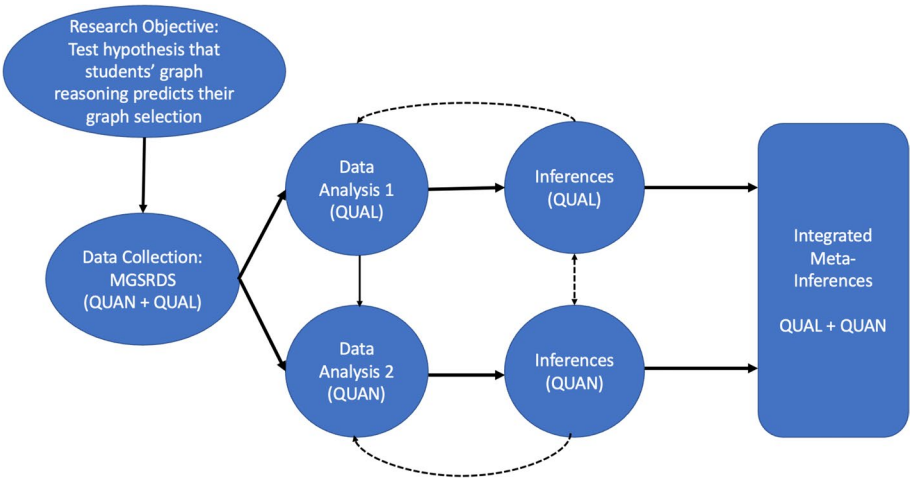


Fig. 2 Logic model

3.2 Using SEM to test a conceptual model

Using SEM, we tested a conceptual model relating the constructs of graph reasoning and graph selection. For graph reasoning, we used students' written explanations of why they chose a graph as proxies for their reasoning. For graph selection, we used the correctness of students' choices as a proxy for their selection. Drawing on the Johnson et al. (2020) framework, we made two hypotheses. First, we hypothesized students' graph reasoning to influence the accuracy of their graph selection. Second, we hypothesized that students who demonstrated quantitative forms of graph reasoning (i.e., covariational and variational reasoning) would have greater accuracy in their graph selection.

We grounded our hypotheses in Thompson's theory of quantitative reasoning (Thompson, 1994, 2022; Thompson & Carlson, 2017), as well as results of empirical studies investigating participants' covariational reasoning. Conducting smaller-scale qualitative studies, researchers identified the utility of covariational reasoning for students' graphing activity, with participants spanning middle grades (e.g., Ellis et al., 2016), secondary (e.g., Johnson et al., 2020), and university students (e.g., Carlson et al., 2002; Moore et al., 2019b). In an empirical study, Thompson et al. (2017) asked teachers from the U.S. and South Korea ($n = 487$) to sketch a Cartesian graph to represent a relationship between two variables that were varying unsystematically. They found strong evidence that teachers' placement of an initial point on a graph influenced their graph's accuracy. Furthermore, in a validation study for the PCA, Carlson et al. (2010) found that students who selected the correct graph on a bottle problem item also were engaging in covariational reasoning. Hence, we viewed our hypotheses to have a strong foundation in the literature.

3.3 Data collection, setting, and participants

Our sample included 673 students' responses to the MGSRDS from students enrolled in a college algebra course, across four U.S. postsecondary institutions. At the start of the study, each institution was designated as Hispanic Serving, which meant that at least 25% of the undergraduate population identified as Hispanic or Latino. Furthermore, at least 45% of undergraduate students at each institution identified as students of color. The courses varied in meeting times and delivery methods (in person, remote synchronous via video conference, and online asynchronous), in part due to the global pandemic. We collected responses throughout four consecutive fall and spring semesters (Fall 2020, Spring 2021, Fall 2021, Spring 2022). To facilitate data collection, the project team developed an online module that instructors could include in their course learning management system. The module included a video introducing the research project, a video describing the MGSRDS, and a link for access. Students completed the MGSRDS near the end of each semester, either during class or on their own time; individual student demographics were not collected.

3.4 Qualitative analysis: coding students' written explanations and graph choices

As we had done when establishing the validity and reliability of the MGSRDS, we used consensus coding (Olson et al., 2016) for graph reasoning, with three GRAs working together under Johnson's supervision. GRAs began by familiarizing themselves with

each form of graph reasoning in the Johnson et al., (2020 framework), along with sample student responses evidencing each form of reasoning. Given identical sample data sets, individual GRAs coded using a priori codes in Table 2. Next, all three GRAs met to discuss their codes, taking reflective notes to capture their questions and insights, then meeting with Johnson to calibrate their codes. After training, analysis occurred in four rounds, with coding occurring after each semester of administration. Two GRAs coded each written explanation, and after coding they met to calibrate any codes for which they disagreed. If they could not agree, they brought in a third GRA, or as a last step, consulted Johnson to resolve the disagreement. Hence, the consensus coding process yielded 100% interrater agreement for the set of 4038 responses (673 students * 6 responses/student).

GRAs coded students' graph choices along three categories: correct, partially correct, and incorrect (see Table 1), based on their multiple choice response. To guard against potential bias in coding, we used spreadsheets that separated students' graph selection from their graph reasoning. After coding was complete, we linked students' graph selection codes to their graph reasoning codes to examine patterns in responses.

3.5 Quantitative analysis: scoring the MGSRDS, conducting SEM

To address our first hypothesis, we conducted SEM. To prepare the data set, we quantitized (Sandelowski et al., 2009) qualitative codes for students' graph reasoning and selection. For graph reasoning, we used the five-point scale shown in Table 2 (4-COV, 3-VAR, 2-MO, 1-IC, 0-LE). For graph selection, we used a three-point scale (2-correct, 1-partially correct, 0-incorrect).

Power analysis revealed that a minimum of 200 was required to detect a small effect size; hence our sample size ($n = 673$) was sufficiently large. Using IBM SPSS AMOS 27 (IBM Corp., 2020) with raw scores, we assessed model fit with three statistics: Chi-square goodness of fit, the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA). With the CFI, we estimated the relative fit of the model in comparison to a null baseline model of uncorrelated variables. CFI values should range from 0–1 with values of 0.90 and above being sufficient evidence of good fit (Bentler & Bonett, 1980). With the RMSEA, we examined absolute fit, which considers how far our conceptual model is from the ideal, with values of 0.08 and below being acceptable fit (Browne & Cudeck, 1992). After assessing model fit, we examined whether assessment items contributed to the intended constructs of graph reasoning and graph selection. To verify, items should contribute at values of 0.30 or above, with higher values indicating stronger contributions (Leech et al., 2014).

3.6 Quantitative and qualitative analysis: integrated meta-inferences

To address our second hypothesis, we examined what the qualitative and quantitative data analyses could indicate when taken together, to develop integrated meta-inferences (Tashakkori et al., 2020). SEM addressed high level, systematic patterns in the data (i.e., treating graph reasoning and selection as holistic constructs). To explore specific patterns in students' graph reasoning and selection on each item (e.g., COV + correct, COV + partially correct, COV + incorrect), we conducted six chi-square tests of independence. In the chi-square analyses, we examined whether the observed values in the data differed from statistical expected values, noted at 95% confidence (Leech et al., 2014). For example, if

the observed "COV + correct" values differed from the expected value, we could claim that students demonstrating covariational reasoning were more likely to select the correct graph than was statistically expected. This allowed us to explore whether different forms of students' graph reasoning on individual items contributed to greater accuracy in graph selections. After conducting chi-square analyses, we reflected on the potential for theory to explain our results.

4 Results

4.1 Hypothesis 1: graph reasoning and graph selection: a structural equation model

Figure 3 shows our structural equation model. The constructs of graph reasoning and graph selection are represented by the circles in the middle. The arrow going from graph reasoning to graph selection indicates the prediction pathway. The rectangles on the left and right represent the six MGSRDS items. From graph reasoning and graph selection, the arrows extending to each item indicate that each item contributes to each construct.

Fit indices showed that data fit the model well, with $\chi 2$ (54) = 75.17, $p = 0.03$; CFI = 0.99; RMSEA = 0.02. The CFI showed evidence of good fit, being well above the threshold of 0.90. The RMSEA also showed evidence of good fit, being well below the threshold of 0.08. Furthermore, the standardized regression weight of the graph reasoning → graph selection pathway was 0.64, $p < 0.001$. This meant that 40% of the variance in the graph selection construct was being explained by the graph reasoning construct. The graph reasoning → graph selection pathway statistic of 0.64 (Table 3) showed a moderate effect size. Based on the units in this model, a one unit increase on the reasoning scale (e.g., MO to VAR) would predict a 0.64 unit increase in accurate graph selection, moving almost into the next step on the correctness scale (e.g., from incorrect to partially correct).

Next, we examined each MGSRDS item to determine how items contributed to the intended constructs of graph selection and graph reasoning (see Table 3). All items contributed significantly ($p < 0.007$) to both constructs. For graph reasoning, all items contributed at values above 0.56. For graph selection, items contributed at somewhat lower values; all were below 0.47, and two items (Nat & Tree, Ant & Ladybug) contributed below the recommended threshold of 0.30. Per Leech et al. (2014), if items contributed below 0.30 and were not significant, they might be potential 'problem items' that would contribute less

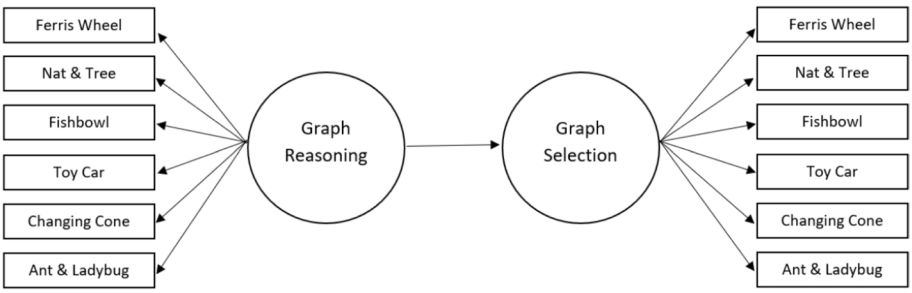


Fig. 3 Graph reasoning and selection: structural equation model

Table 3 Pathway statistics

Pathway	Std. Reg. Weight	p
Reasoning → Correctness	0.64	< 0.001
Ferris Wheel → Reasoning	0.71	< 0.001
Nat & Tree → Reasoning	0.71	< 0.001
Fishbowl → Reasoning	0.76	< 0.001
Toy Car → Reasoning	0.66	< 0.001
Changing Cone → Reasoning	0.62	< 0.001
Ant & Ladybug → Reasoning	0.56	< 0.001
Ferris Wheel → Correctness	0.35	< 0.001
Nat & Tree → Correctness	0.15	0.006
Fishbowl → Correctness	0.37	< 0.001
Toy Car → Correctness	0.43	< 0.001
Changing Cone → Correctness	0.47	< 0.001
Ant & Ladybug → Correctness	0.29	< 0.001

to the construct. Although two items contributed to graph selection below 0.30, they both were significant. Furthermore, removing them did not improve model fit, so we chose to keep all MGSRDS items. Hence, with our SEM analysis, we confirmed our first hypothesis, that students' graph reasoning influences their graph selection on the MGSRDS.

4.2 Hypothesis 2: linking graph selection and forms of graph reasoning

Table 4 shows frequencies of students' graph reasoning and graph selection, pairing each graph reasoning code (COV, VAR, MO, IC, LE) with each graph selection code (Correct, partially correct, incorrect). Using six chi-square tests of independence, we examine whether the association between graph selection and graph reasoning for each item is statistically supported along with which observed values differ from expected values. Five items, indicated with asterisks in Table 4, show significant associations (Ferris Wheel: $\chi 2$ (8) = 50.21, $p < 0.001$; Fishbowl: $\chi 2$ (8) = 31.55, $p < 0.001$; Toy Car: $\chi 2$ (8) = 42.50, $p < 0.001$; Changing Cone: $\chi 2$ (8) = 60.17, $p < 0.001$; Ant & Ladybug: $\chi 2$ (4) = 12.87, $p = 0.01$). Notably, those associations include only COV, VAR, and LE reasoning codes.

For the Ferris Wheel, Fishbowl, Toy Car, and Changing Cone items, more LE codes associated with incorrect graph selection, and for the Ferris Wheel, Fishbowl, and Changing Cone, fewer LE codes associated with correct graph selection. This meant that students who demonstrated limited evidence of reasoning were more likely to get these items incorrect (and for the Ferris Wheel, Fishbowl, and Changing Cone, less likely to get the items correct). Also, for the Ferris Wheel and Toy Car, fewer LE codes associated with partially correct graph selection.

For the Ferris Wheel and Toy Car items, more VAR codes associated with correct graph selection, and fewer VAR codes associated with incorrect graph selections. This meant that students who demonstrated variational reasoning were more likely to get these items correct (and less likely to get these items incorrect). For the Fishbowl, Changing Cone, and Ant & Ladybug items, more COV codes associated with correct graph selection, and for the Fishbowl and Changing Cone, fewer COV codes associated with incorrect graph selection. This meant that students who demonstrated covariational reasoning were more likely

Table 4 Frequencies of students' graph reasoning and selection across MGSRDS assessment items

Graph Selection					
Assessment Item	Reasoning Code	Correct (2)	Partially Correct (1)	Incorrect (0)	Total
Ferris Wheel (FW)***	COV (4)	85	14	30	129
	VAR (3)	99	5	22	126
	MO (2)	79	11	41	131
	IC (1)	50	4	30	84
	LE (0)	86	27	90	203
	Total	399	61	213	673
Nat & Tree (NT)	COV (4)	56	13	28	97
	VAR (3)	71	12	42	125
	MO (2)	116	24	94	234
	IC (1)	27	7	11	45
	LE (0)	70	29	73	172
	Total	340	85	248	673
Fishbowl (FB)***	COV (4)	138	55	61	254
	VAR (3)	49	31	39	119
	MO (2)	20	13	28	61
	IC (1)	23	11	13	47
	LE (0)	64	42	86	192
	Total	294	152	227	673
Changing Cone (CO)***	COV (4)	86	108	62	256
	VAR (3)	22	30	20	72
	MO (2)	21	30	26	77
	IC (1)	11	21	34	66
	LE (0)	28	62	112	202
	Total	168	251	254	673
Toy Car (TC)***	COV (4)	11	17	14	42
	VAR (3)	38	56	34	128
	MO (2)	34	102	71	207
	IC (1)	13	48	38	99
	LE (0)	27	62	108	197
	Total	123	285	265	673
Ant & Ladybug (AL)*	COV (4)	21		52	73
	VAR (3)	3		32	35
	MO (2)	59		242	301
	IC (1)	10		54	64
	LE (0)	25		175	200
	Total	118		555	673

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

to get these items correct (and for the Fishbowl and Changing Cone, less likely to get these items incorrect).

These results confirm our second hypothesis, that students' quantitative graph reasoning (i.e., variational and covariational reasoning) relates to more accurate graph selections. In doing so, we acknowledge nuances with this claim. First, we interpret the association between VAR codes and accuracy in graph selection on the Ferris Wheel and Toy Car items to indicate that variational reasoning may be sufficient to get these items correct. One reason may be that students are thinking of total distance, represented in each of the items, as if it were experiential time (Thompson & Carlson, 2017), and hence when they explain their reasoning, it seems sufficient just to mention the attribute that is changing direction.

Second, we interpret the design of the Fishbowl and Changing Cone situations to contribute in part to the association between COV codes and accuracy in graph selection. In particular, we consider whether there is an "actor" in the situation that links each of the attributes. In the Ferris Wheel, Nat & Tree, Toy Car, and Ant & Ladybug situations, there is an actor that can be represented by a single point, to which both attributes are linked. For example, in the Ferris Wheel item, the cart (i.e., the "actor") is represented by a single point, and the height from the ground and the total distance traveled are linked to the cart. In contrast, for the Fishbowl and Changing Cone situations, there is no actor linking the attributes, only the situation itself.

5 Discussion and implications

5.1 Summary and support of hypothesis

At the outset, we asked: To what extent does evidence of students' graph reasoning relate to the accuracy of their graph selection? We confirmed two hypotheses. First, we found that students' graph reasoning significantly influenced their graph selection on the MGSRDS, with graph reasoning explaining 40% of the variance in graph selection. Second, we found that students' quantitative graph reasoning (i.e., variational and covariational reasoning) related to more accurate graph selections, with significant associations on five MGSRDS items.

While we are encouraged that students' graph reasoning can explain 40% of the variance related to the accuracy of their graph selection, we acknowledge that 60% is left unexplained. We provide a few reasons to account for this. The first is the setting itself; while students complete the MGSRDS as part of their course, they still may view the assessment to be outside the scope of college algebra and may be less invested in their responses. The second is our source of data; we use students' written responses as a proxy for their reasoning. Hence, students may engage in reasoning that their written work may not evidence. Third, our findings reveal only COV, VAR, and LE reasoning are associated with accurate (or inaccurate) graph selection. Thus, MO and IC reasoning less clearly associate with the accuracy of students' graph selection. An additional framework from a different theoretical perspective, such as students' perceptions of their competence with graphs, may help to explain more of the variance.

5.2 Structural equation model: impact of measurement error

Our results highlighted the benefits of latent variable modeling techniques (e.g., Ledgerwood & Shrout, 2011). At first, we used Pearson's correlation to examine a relationship between graph selection and graph reasoning, which meant we treated those constructs as observed,

rather than latent variables. We found only a small correlation between the observed mean scores, $r(671)=0.39$, $p<0.001$, indicating that only 15% of the variance in graph selection was being explained by graph reasoning. We concluded that measurement errors between and within items were influencing the relationship. Hence, we needed to treat these constructs as latent variables, which required the use of SEM. In our view, a primary source of the measurement error was due to our use of written responses as a proxy for students' graph reasoning, because students might be engaging in reasoning not captured by their written responses. Furthermore, by design, the MGSRDS forced students to select a graph choice, so given the option, students might not have selected any of the graphs as a viable choice. Our results highlighted the utility of latent variable modeling techniques for testing hypotheses generated from smaller scale studies of students' reasoning.

5.3 Implications for design and implementation of graphing tasks

To increase the complexity of graphing tasks, designers could include attributes that both vary in their direction of change (i.e., both increase and decrease). Notably, Thompson et al. (2017) suggested that such tasks may be best included in small scale teaching experiment studies (Steffe & Thompson, 2000). Because students were selecting, rather than sketching graphs on the MSGRDS, we decided to include such an item. The difficulty of the Ant & Ladybug item illuminated the challenge for students when both attributes vary in their direction of change. Despite differences between the Ant & Ladybug and other MGSRDS items, it was very close to fitting our structural equation model, and students demonstrating covariational reasoning were more likely to get the item correct than statistically expected. Thus, our results demonstrated the viability of including such tasks in research settings beyond teaching experiments.

Across the MGSRDS items, three included unconventional graphs (Moore et al., 2014), the Ferris Wheel, Nat & Tree, and Fishbowl. The unconventional graphs did not pass the vertical line test, a common technique students use to determine if a graph represents a function given by $y=f(x)$, with x and y represented on the horizontal and vertical axes, respectively. Moore et al. (2014) contended that breaking conventions could support students' quantitative reasoning. Notably, the Fishbowl item included the largest numbers of VAR/COV codes of all MGSRDS items. Furthermore, students' variational reasoning on the Ferris Wheel item and covariational reasoning on the Fishbowl item significantly associated with the accuracy of their graph selection. Hence, our results provided quantitative corroboration to support the findings from Moore et al. (2014).

Students' spatial conceptions of the Cartesian coordinate systems (Lee et al., 2020; Paoletti et al., 2022) likely are interwoven with their MO and IC reasoning. Paoletti et al. (2022) allude to this when they describe the possibility for a student to represent a relationship with a coordinate system different from the one intended (e.g., students using a spatial coordinate system instead of a quantitative one). In future studies, researchers can explore interconnections between students' quantitative and spatial conceptions of coordinate systems and the different forms of graph reasoning in the Johnson et al. (2020) framework.

6 Concluding remarks

Our results underscore the value of variational and covariational reasoning, even at early levels of gross variation and gross coordination of values, per Thompson and Carlson's (2017) framework. Our study has implications for incorporating a focus on covariational

reasoning into early undergraduate mathematics courses, such as college algebra, to foster students' conceptions of graphs as representing relationships between variables. Furthermore, instructors can use the MGSRDS in combination with the Johnson et al. (2020) framework to diagnose students' graph reasoning and selection. In future studies, researchers can investigate links between graph reasoning and graph selection for different student populations (e.g., secondary students).

This study adds to research efforts using quantitative methods to corroborate findings from qualitative studies investigating students and teachers' reasoning (Izsák et al., 2019; Kosko, 2019; Norton & Wilkins, 2009; Pantziara & Philippou, 2012; Tzur et al., 2022). We offer two recommendations for designing assessment items linking dynamic situations and graphs. First, it is useful to include items like the Fishbowl item, that do not incorporate an actor that can be represented by a single point, to which both attributes are linked. Second, if items do include such an actor, it is valuable to have both attributes vary in their direction of change, like the Ant & Ladybug item.

Bakker et al. (2021) posit a broad challenge, namely "to assess what we value, rather than valuing what we assess" (p. 11). The value of graph reasoning extends beyond school mathematics, to include students' roles as educated citizens. Employing a broad conceptualization of graph reasoning with the Johnson et al. (2020) framework, we have demonstrated that college algebra students' graph reasoning significantly contributes to their graph selection, with quantitative forms of graph reasoning (i.e., variational and covariational reasoning) significantly associating with increased accuracy in graph selection.

Funding This research was funded by the U.S. National Science foundation under DUE-2013186, Division of Undergraduate Education. Opinions and conclusions are those of the authors.

Data Availability The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

IBM Corp. (2020). *IBM SPSS statistics for windows, version 27.0.* IBM Corp.

Bakker, A., Cai, J., & Zenger, L. (2021). Future themes of mathematics education research: An international survey before and during the pandemic. *Educational Studies in Mathematics, 107*(1), 1–24. https://doi.org/10.1007/s10649-021-10049-w

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588. https://doi.org/10.1037/0033-2909.88.3.588

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Byrne, B. M. (2016). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming* (3rd ed.). Routledge.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education, 33*(5), 352–378. https://doi.org/10.2307/4149958

Carlson, M., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction, 28*(2), 113–145. https://doi.org/10.1080/07370001003676587

Castillo-Garsow, C., Johnson, H. L., & Moore, K. C. (2013). Chunky and smooth images of change. *For the Learning of Mathematics, 33*(3), 31–37.

Clement, J. (1989). The concept of variation and misconceptions in cartesian graphing. *Focus on Learning Problems in Mathematics, 11*(1–2), 77–87. https://eric.ed.gov/?id=EJ389508.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). SAGE Publications.

DeVellis, R. F. (2006). Classical test theory. *Medical Care, 44*(11 Suppl 3), S50–S59. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Donovan, C., Johnson, H. L., Knurek, R., Whitmore, K. A., & Bechtold, L. (accepted). Validating a measure of graph selection and graph reasoning for dynamic situations. *Journal of Mathematical Behavior*.

Ellis, A. B., Ozgur, Z., Kulow, T., Dogan, M. F., & Amidon, J. (2016). An exponential growth learning trajectory: Students' emerging understanding of exponential growth through covariation. *Mathematical Thinking and Learning, 18*(3), 151–181. https://doi.org/10.1080/10986065.2016.1183090

Frank, K. M. (2017). *Examining the development of students' covariational reasoning in the context of graphing* [Ph.D. Dissertation]. Arizona State University.

Gordon, S. P. (2008). What's wrong with college algebra? *Primus, 18*(6), 516–541. https://doi.org/10.1080/10511970701598752

Izsák, A., Jacobson, E., & Bradshaw, L. (2019). Surveying Middle-Grades Teachers' Reasoning About Fraction Arithmetic in Terms of Measured Quantities. *Journal for Research in Mathematics Education, 50*(2), 156–209. https://doi.org/10.5951/jresematheduc.50.2.0156

Johnson, H. L. (2022). An intellectual need for relationships: Engendering students' quantitative and covariational reasoning. In G. KaragözAkar, İÖ. Zembat, S. Arslan, & P. W. Thompson (Eds.), *Quantitative Reasoning in Mathematics and Science Education* (pp. 17–34). Springer International Publishing. https://doi.org/10.1007/978-3-031-14553-7_2

Johnson, H. L., Kalir, J., Olson, G., Gardner, A., Smith, A., & Wang, X. (2018). Networking theories to design a fully online assessment of students' covariational reasoning. In T. E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.), *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1343–1346). University of South Carolina & Clemson University.

Johnson, H. L., McClintock, E., & Gardner, A. (2020). Opportunities for reasoning: Digital task design to promote students' conceptions of graphs as representing relationships between quantities. *Digital Experiences in Mathematics Education, 6*(3), 340–366. https://doi.org/10.1007/s40751-020-00061-9

Johnson, H. L., Olson, G., Smith, A., Gardner, A., Wang, X., & Donovan, C. (2021). Validating an assessment of students' covariational reasoning. In D. Olanoff, K. Smith, & S. Spitzer (Eds.) *Proceedings of the 43rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education,* (pp. 63–67). Philadelphia, PA. https://www.pmena.org/pmenaproceedings/PMENA%2043%202021%20Proceedings.pdf

Kerslake, D. (1977). The understanding of graphs. *Mathematics in School, 6*(2), 22–25. http://www.jstor.org/stable/30212405.

Kilpatrick, J. (2001). Where's the evidence? *Journal for Research in Mathematics Education, 32*(4), 421–427. https://www.jstor.org/stable/749702.

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th Ed.). Guilford.

Kosko, K. W. (2019). A multiplicative reasoning assessment for fourth and fifth grade students. *Studies in Educational Evaluation, 60*, 32–42. https://doi.org/10.1016/j.stueduc.2018.11.003

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101*(6), 1174–1188. https://doi.org/10.1037/a0024776

Lee, H. Y., Hardison, H. L., & Paoletti, T. (2020). Two uses of coordinate systems [1]. *For the Learning of Mathematics, 40*(2), 32–37. https://www.jstor.org/stable/27091157.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity, 43*(2), 265–275. https://doi.org/10.1007/s11135-007-9105-3

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2014). *IBM SPSS for intermediate statistics: Use and interpretation* (5th ed.). Routledge.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*(1), 1–64. https://doi.org/10.3102/00346543060001001

Mesa, V., Suh, H., Blake, T., & Whittemore, T. (2012). Examples in college algebra textbooks: Opportunities for students' learning. *Primus, 23*(1), 76–105. https://doi.org/10.1080/10511970.2012.667515

Moore, K. C., Silverman, J., Paoletti, T., & LaForest, K. (2014). Breaking Conventions to Support Quantitative Reasoning. *Mathematics Teacher Educator, 2*(2), 141–157. https://doi.org/10.5951/mathteaceduc.2.2.0141

Moore, K. C., Silverman, J., Paoletti, T., Liss, D., & Musgrave, S. (2019a). Conventions, habits, and US teachers' meanings for graphs. *The Journal of Mathematical Behavior, 53*, 179–195. https://www.sciencedirect.com/science/article/pii/S0732312317302663.

Moore, K. C., Stevens, I. E., Paoletti, T., Hobson, N. L. F., & Liang, B. (2019b). Pre-service teachers' figurative and operative graphing actions. *The Journal of Mathematical Behavior, 56*, 100692. https://www.sciencedirect.com/science/article/pii/S0732312318300269.

Norton, A., & Wilkins, J. L. M. (2009). A quantitative analysis of children's splitting operations and fraction schemes. *The Journal of Mathematical Behavior, 28*(2), 150–161. https://doi.org/10.1016/j.jmathb.2009.06.002

Olson, J., McAllister, C., Grinnell, L., Gehrke Walters, K., & Appunn, F. (2016). Applying constant comparative method with multiple investigators and inter-coder reliability. *The Qualitative Report, 21*(1), 26–42. https://doi.org/10.46743/2160-3715/2016.2447

Pantziara, M., & Philippou, G. (2012). Levels of students' "conception" of fractions. *Educational Studies in Mathematics, 79*(1), 61–83. https://doi.org/10.1007/s10649-011-9338-x

Paoletti, T., Hardison, H. L., & Lee, H. Y. (2022). Students' static and emergent graphical shape thinking in spatial and quantitative coordinate systems. *For the Learning of Mathematics, 42*(2), 48–50. https://doi.org/https://flm-journal.org/index.php?do=show&lang=en&vol=42&num=2.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On Quantitizing. *Journal of Mixed Methods Research, 3*(3), 208–222. https://doi.org/10.1177/1558689809334210

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In R. A. Lesh & A. E. Kelly (Eds.), *Research design in mathematics and science education* (pp. 267–306). Lawrence Erlbaum Associates.

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The New Era of mixed methods. *Journal of Mixed Methods Research, 1*(1), 3–7. https://doi.org/10.1177/2345678906293042

Tashakkori, A., Burke Johnson, R., & Teddlie, C. (2020). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences* (2nd Ed.). SAGE Publications.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *The Journal of the International Association of Medical Science Educators: JIAMSE, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Thompson, P. W. (1994). The development of the concept of speed and its relationship to concepts of rate. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 179–234). State University of New York Press.

Thompson, P. W. (2022). Quantitative reasoning as an educational lens. In G. KaragözAkar, İÖ. Zembat, S. Arslan, & P. W. Thompson (Eds.), *Quantitative Reasoning in Mathematics and Science Education* (pp. 1–16). Springer International Publishing. https://doi.org/10.1007/978-3-031-14553-7_1

Thompson, P. W., & Carlson, M. P. (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.

Thompson, P. W., Hatfield, N. J., Yoon, H., Joshua, S., & Byerley, C. (2017). Covariational reasoning among U.S. and South Korean secondary mathematics teachers. *The Journal of Mathematical Behavior, 48*(Supplement C), 95–111. https://doi.org/10.1016/j.jmathb.2017.08.001

Tzur, R., Johnson, H. L., Davis, A., Hodkowski, N. M., Jorgensen, C., Wei, B., & Norton, A. (2022). A stage-sensitive written measure of multiplicative double counting for grades 3–8. *Studies in Educational Evaluation, 74*, 101152. https://doi.org/10.1016/j.stueduc.2022.101152

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.