

RESEARCH

Open Access



Standards-based mathematics curricula and the promotion of quantitative literacy in elementary school

Jesse L. M. Wilkins

Abstract

Background: Prior research has shown that students taught using *Standards*-based mathematics curricula tend to outperform students on measures of mathematics achievement. However, little research has focused particularly on the promotion of student quantitative literacy (QLT). In this study, the potential influence of the *Investigations in Number, Data, and Space* curriculum on student quantitative literacy is investigated. Quantitative literacy is conceptualized as a hierarchical three-factor model comprising the interrelationship among a student's mathematical beliefs, disposition, and cognition. This theoretical model is validated with elementary-aged students and used to investigate whether students' quantitative literacy is related to the use of the *Investigations* curriculum.

Results: The hierarchical three-factor QLT model was found to have relatively good fit for the sample of elementary-aged students, and all inter-factor relationships were found to be consistent with the proposed theoretical model of the quantitative literacy construct. On average, students in the school district using a *Standards*-based mathematics curriculum had increased levels of quantitative literacy when compared to students in the district not using the curriculum or using it for less time.

Conclusions: Based on the findings of this study, the *Investigations* mathematics curriculum has potential to promote students' development of quantitative literacy in elementary school. Furthermore, the results of the study provide additional validation for the theoretical quantitative literacy construct modeled as a second-order factor comprising the interrelationship among a student's mathematical beliefs, mathematical disposition, and mathematical cognition.

Keywords: Quantitative literacy, *Standards*-based Mathematics curricula, Mathematical beliefs, Mathematical disposition

The curriculum...should emphasize the mathematics processes and skills that support the quantitative literacy of students. (National Council of Teachers of Mathematics 2000, p. 16).

Background

Out of a call for reform in mathematics education (National Council of Teacher of Mathematics 1989; also see National Research Council 1989; American Association for the Advancement of Science 1990),

the National Science Foundation funded several projects during the 1990s to develop mathematics curricula that were consistent with the standards outlined in the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teacher of Mathematics 1989) and later the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics 2000). Research evaluating the impact of these curricula has consistently found that students who learn mathematics using these curricula outperform students who do not on measures of mathematics achievement, problem-solving, and reasoning (Post et al. 2008; Mokros 2003; Harris et al. 2001; McCaffrey et al. 2001; Reys et al. 2003; Senk and Thompson 2003; Thompson and Senk 2001; Riordan and Noyce 2001; Carroll 1997). However, most of these studies

Correspondence: wilkins@vt.edu
School of Education/Teaching and Learning, War Memorial Hall, RM 300C,
Virginia Tech, 370 Drillfield Drive, Blacksburg, VA 24061, USA

focused primarily on measures of mathematics content and did not incorporate aspects of mathematical learning related to beliefs and attitudes. In other words, the impact of these curricula has not been evaluated directly for their impact on students' overall quantitative literacy (QLT).

The notion of quantitative literacy developed in the later part of the 20th century, also, out of a call for reform in mathematics education (American Association for the Advancement of Science 1990; National Council of Teacher of Mathematics 1989; National Research Council 1989; Cockcroft 1982; National Commission on Excellence in Education 1983). Although continued debate exists over what exactly constitutes a person's quantitative literacy, a review of the literature suggests that most would agree that it includes an everyday "walking around" knowledge of mathematics characterized by an ability to reason and think mathematically (Steen 1997, p. xvi; also see Cockcroft 1982; Steen 1999, 2001; National Council of Teacher of Mathematics 2000, 1989). However, different from mathematics, which is a discipline to be studied, quantitative literacy is a habit of mind that is further characterized by a person's beliefs about mathematics and their mathematical disposition (e.g., Wilkins, in press, 2010, 2000; Steen, 2001; National Council of Teachers of Mathematics 2000, 1989; Gal 1997; Atkins and Helms 1993; American Association for the Advancement of Science 1990). This expanded view of quantitative literacy embodies the mathematical goals for students advocated by the mathematics education reform begun in the later part of the 20th century (National Council of Teacher of Mathematics 1989, p. 5) and continuing into the 21st century (National Council of Teachers of Mathematics 2006, 2000; National Governors Association Center for Best Practices, Council of Chief State School Officers 2010).

In the USA, quantitative literacy is most often associated with upper secondary and post-secondary education as students prepare for adult citizenship (e.g., Gillman 2006a; Steen 2004; Madison and Steen 2003). However, before students reach secondary and post-secondary education, they are enculturated with beliefs and attitudes that shape the way they view and approach quantitative situations both in and out of school and in their current and future daily lives. For this reason, it is important to investigate potential influences on the early development of children's quantitative literacy. One such influence is the curriculum that is used in mathematics classrooms.

In this study, grounded in the notion of quantitative literacy outlined by Wilkins (in press, 2010), a measurement model of QLT is validated for a sample of elementary-aged students. This measurement model is then used to examine the level of quantitative literacy for these elementary-aged students as it relates to the use of a *Standards*-based mathematics

curriculum (i.e., *Investigations in Number, Data and Space*, [Russell et al. 2004]).

Quantitative literacy

Based on earlier work to define quantitative literacy (e.g., Steen 2001; Wilkins 2000; Gal 1997; Dossey 1997), Wilkins (in press, 2010) conceptualized a theoretical framework and measurement model of quantitative literacy to include not only the cognitive aspects of quantitative literacy but also its affective components. Wilkins (in press, 2010) validated a three-factor model of quantitative literacy that incorporated the multiple aspects of quantitative literacy into one measure. Based on this framework, a person's quantitative literacy is characterized by the inter-relationship among their (a) mathematical beliefs (b) mathematical disposition, and (c) mathematical cognition.

A person who is quantitatively literate tends to have beliefs about mathematics and statistics that are more humanistic in that they view quantitative information as accessible, interpretable, and producible by everyone, not just experts (Wilkins, in press, 2010, 2000; Steen 2001; Gal 1997). Hofer and Pintrich (1997; also see Hofer 2000) outlined personal epistemological beliefs into those related to the nature of knowledge and those related to the nature of knowing. Within these two categories, the nature of knowledge can be thought of in terms of the simplicity and certainty of knowledge, and the nature of knowing can be thought of in terms of the source and justification of knowledge. Wilkins (in press) used this framework to outline the views of mathematics that would be consistent or not consistent with a person who is quantitatively literate. A belief in the simplicity of mathematics would produce, for example, a view of mathematics as a set of predetermined facts and procedures as opposed to a network of interrelated concepts. A belief in the certainty of knowledge would produce a view of mathematics as static and known as opposed to constantly changing over time as new ideas are discovered. Source of knowledge relates to views about where mathematical knowledge originates from, externally or internally; for example, do mathematical ideas only come from textbooks, experts, or teachers or can they be created or re-created at some level by everyone? Justification of knowledge relates to how mathematical claims are warranted. For example, can mathematical ideas be justified by individuals or can they only be justified by experts, textbooks, or calculators? For a person who is quantitatively literate, mathematics is viewed more as a way of thinking and reasoning, as dynamic and open for discussion and new discoveries, instead of a static set of predetermined facts and formulas that are passed along in an authoritarian manner to be memorized.

A positive disposition toward mathematics would be characteristic of a person who is quantitatively literate

(Wilkins, in press, 2010, 2000; Steen 2001; National Council of Teachers of Mathematics 2000, 1989). In this case, a positive disposition would be characterized by a willingness to engage and persist in quantitative situations (Wilkins, in press, 2010). Grounded in the expectancy value theory of achievement motivation (Eccles et al. 1983), a quantitatively literate person would have increased expectations for success in quantitative situations, they would have an increased value in mathematics and statistics for its social utility and would have an increased level of intrinsic interest in quantitative situations (Wilkins, in press, 2010). A person's expectations for success with a given task and the values associated with the success of the task are related to how persistent a person is with that task (Eccles et al. 1983; Eccles and Wigfield 1995). This interrelationship among expectations and values becomes an essential component of a person's way of thinking and habit of mind, and thus, an important part of a person's quantitative literacy.

Mathematical cognition as it relates to quantitative literacy includes a functional knowledge of mathematics content—an ability to handle mathematics that one might find in everyday life; that is, mathematics necessary to balance a checkbook, shop for groceries, read a map or bus schedule, or read the newspaper. Such a facility with mathematics includes a breadth of knowledge as outlined, for example, by the National Council of Teachers of Mathematics (2000)—number and operations, geometry and measurement, data analysis, and algebra (see also Dossey 1997); or more generally an understanding of pattern, dimension, quantity, uncertainty, shape, and change (Steen 1990). Mathematical cognition further extends beyond content knowledge to include a person's ability to reason and think mathematically—to evaluate and synthesize information, make conjectures and draw logical conclusions, and apply knowledge to real-world problems and situations (Steen 2001; Wilkins 2000; National Council of Teachers of Mathematics 2000, 1989).

Purpose of study

This study has two interrelated purposes: (1) to create and validate a model of quantitative literacy for elementary-aged students and (2) to use this model to examine the relationship between student quantitative literacy and exposure to a *Standards*-based mathematics curriculum. In this study, quantitative literacy is modeled as a second-order factor that accounts for the interrelationship among three first-order factors (Wilkins, in press, 2010): a person's (a) mathematical beliefs, (b) mathematical dispositions, and (c) mathematical cognition. This theoretical model has been validated for secondary (Wilkins, 2010) and undergraduate (Wilkins, in press) students. The purpose of this study is to examine the construct validity of this hierarchical three-factor model of quantitative literacy

for three cohorts of elementary-aged students. This model is then used to investigate whether elementary students, who are in a school district that has adopted and implemented a mathematics curriculum consistent with the goals of the National Council of Teachers of Mathematics (2000, 1989; cf. National Governors Association Center for Best Practices, Council of Chief State School Officers 2010), have increased quantitative literacy over students in the district not using the curriculum or for a shorter amount of time.

Methods

Data

This study involved 2490 fourth graders from one school district located in the southeastern USA. The sample was made up of three cohorts of students from three consecutive school years. Only students who were in the school district for at least two successive years (grades 3 and 4) were included in the study. The school district included 16 elementary schools.¹ The district had recently implemented the *Investigations in Number Data & Space* curriculum (Russell et al. 2004) in the elementary schools. The curriculum was implemented in shifts beginning with grades K-1 in school year 2000–2001, grades 2–3 in school year 2001–2002, and grades 4–5 in school year 2002–2003. Because of this implementation schedule, the first cohort of students was not involved with *Investigations*; the second cohort of students was involved with *Investigations* for 2 years, in grades 3 and 4; the third cohort of students was involved with *Investigations* for 4 years, in grades 1–4.

Teachers in the school district were provided with 153 h of professional development related to the implementation of the new curriculum during the summer prior to implementation and during the subsequent 2 years. Based on reports from the district mathematics supervisors, all schools used the *Investigations* curriculum but at differing levels of implementation.

The students in the first cohort ($n = 815$) were fourth graders in the school year 2001–2002 and were surveyed and tested in the spring of 2002, prior to the implementation of *Investigations* in the district. This sample was 49.2 % female and 50.8 % male. The sample was 80.6 % White, 13.4 % Black, and 6.0 % Other (Asian, Hispanic, American Indian). Of the sample, 18.7 % of the students received free or reduced lunch, with 0.6 % missing. The students in the second cohort ($n = 837$) were fourth graders in the school year 2002–2003 and were surveyed and tested in the spring of 2003, after implementation of *Investigations*. This sample was 48.7 % female and 51.3 % male. The sample was 80.5 % White, 11.9 % Black, and 7.5 % Other (Asian, Hispanic, American Indian). Of the sample, 19.0 % of the students received free or reduced lunch. The students in the third cohort

($n = 838$) were fourth graders in the school year 2003–2004 and were surveyed and tested, in the spring of 2004, also after implementation of *Investigations*. This sample was 48.8 % female and 51.2 % male. The sample was 79.5 % White, 12.9 % Black, and 7.6 % Other (Asian, Hispanic, American Indian). Of the sample, 17.1 % of the students received free or reduced lunch.

Measures

Student data were gathered using the *New Standards Reference Examination* (NSRE, for discussion of the exam see Briars and Resnick 2000; Wiley and Resnick 1997), published by Harcourt Educational Measurement (1996, 1997, 1998, 1999),² and the *Elementary Mathematics Attitude Survey* (EMAS; Wilkins, 2004a). Each of the variables used in the study is described below. Descriptive statistics for the variables for the whole sample are presented in Table 1. This study uses secondary data analysis as the data were collected for other purposes. However, although the measures for the beliefs component had to be modified because there was not a complete set of belief measures, the data included measures consistent with the earlier models of quantitative literacy (Wilkins, in press, 2010).

Table 1 Descriptive statistics for observed variables for full sample

	<i>N</i>	<i>M</i>	<i>SD</i>
Cognition			
Problem	2459	5.66	3.76
Concept	2459	32.83	8.48
Skills	2459	13.28	2.78
Disposition			
Interest	2211	3.92	1.01
Self-concept	2211	3.82	0.85
Utility	2211	4.05	0.67
Beliefs			
B1	2173	3.27	1.25
B2	2198	1.83	1.14
B3	2186	2.68	1.20
B4	2190	2.64	1.17
B5	2189	3.70	1.13
B6	2193	2.69	1.11
Background			
Gender	2490	0.49	0.50
White	2490	0.80	0.40
Black	2490	0.13	0.33
Other	2490	0.07	0.26
Free/reduced lunch	2485	0.18	0.39
Reading	2297	645.37	43.26

Mathematical cognition

Student mathematical cognition was assessed using the NSRE (Form E). The NSRE assesses students’ mathematical knowledge in the different content strands of mathematics advocated by the NCTM (see Wiley and Resnick 1997; Briars and Resnick 2000). In addition, this knowledge is assessed at three process levels: problem-solving, concepts, and skills. The items on the test use multiple formats including multiple-choice and open-ended. Given that the test assesses children’s content knowledge across the strands of the NCTM and also assesses children’s problem-solving and reasoning, conceptual understanding, and skills, it provides a good measure of the mathematical cognition component of quantitative literacy. The raw scores from each of the three process levels were used as the three measures of mathematical cognition: *problem*, *concept*, and *skills*. The same test form was used for all three cohorts; thus, the raw scores were comparable across the three cohorts.

Mathematical disposition

Student mathematical disposition was assessed using items from the EMAS. This survey was designed using items from other sources (e.g., Third International Mathematics and Science Study [TIMSS], Gonzalez and Smith 1998; Second International Mathematics Study, Westbury and Thalathoti 1989; Fennema and Sherman 1976). Items were chosen to be appropriate for Grade 4 students. This survey consisted of 48 items rated on a 5-point Likert scale from *Strongly Agree* to *Strongly Disagree*. Consistent with Wilkins (in press, 2010), measures of student mathematics self-concept, interest, and utility value were created from the survey and included as part of the mathematics disposition factor. The self-concept scale contained six items (e.g., “I usually do well in math”, “Math is harder for me than most people”) with strong internal consistency ($\alpha_1 = 0.81, \alpha_2 = 0.83, \alpha_3 = 0.84, \alpha_{FULL} = 0.83$). The interest scale contained five items (e.g., “Math is interesting to me,” “I like math”) with strong internal consistency ($\alpha_1 = 0.91, \alpha_2 = 0.92, \alpha_3 = 0.90, \alpha_{FULL} = 0.91$). The utility value scale contained eight items (e.g., “I will use math in many ways as an adult”, “Math is useful in everyone’s life”) with good internal consistency ($\alpha_1 = 0.77, \alpha_2 = 0.74, \alpha_3 = 0.76, \alpha_{FULL} = 0.76$).

Mathematical beliefs

The four components associated with epistemological beliefs related to mathematics (see Wilkins, in press; Hofer 2000) were not explicitly assessed within the EMAS. However, six items on the survey did have face validity in terms of the simplicity of knowledge component. These six individual items were used to measure the beliefs component: (B1) Learning math is mostly memorizing; (B2) Math problems should be solved the

same way by everyone; (B3) In math, memorizing facts is more important than being a good problem solver; (B4) Math is a set of rules; (B5) There is always a rule to follow in solving a math problem; (B6) Math helps one to think according to strict rules. These six items did not overlap with the items used in the disposition factors. Given the age of the students, it was felt that these items were accessible to the children and provided a good representation for the beliefs component. The internal consistency of these six items is relatively low ($\alpha_1 = 0.51$, $\alpha_2 = 0.58$, $\alpha_3 = 0.56$, $\alpha_{FULL} = 0.56$), but along with the face validity of the items, provides some evidence that the scale will be useful as a measure of the simplicity of mathematical knowledge.

Curriculum

Because of the implementation schedule of the *Investigations* curriculum, the three cohorts of students had differing amounts of exposure to the curriculum. For Cohort 1, the district was not using *Investigations* for students in the grades being surveyed. For Cohorts 2 and 3, the district was using *Investigations*. More specifically, for Cohort 2 the curriculum had been implemented for 2 years (in grades 3 and 4), and for Cohort 3, the curriculum had been implemented for 4 years (in grades 1–4). By comparing the three cohorts of students, it is possible to investigate the relationship between curriculum exposure and level of quantitative literacy. In order to compare the level of quantitative literacy across the three cohorts, students were dummy coded by cohort to create three new variables: Cohort 1, Cohort 2, and Cohort 3 (e.g., 1 = member of Cohort 1, 0 = not a member of Cohort 1). By using this coding scheme, it is possible to test whether students in the school district using *Investigations* had higher quantitative literacy scores than students in the district not using *Investigations*. Moreover, because the three cohorts are coded separately, Cohorts 2 and Cohorts 3 can also be compared to test whether increased exposure leads to increased quantitative literacy. For the students in Cohort 1, it is not assumed that they did not have opportunities to learn mathematics and develop their quantitative literacy; it is only assumed that they were not using the *Investigations* curriculum.

Background measures

Student background measures included gender, race/ethnicity, free/reduced lunch status, and prior knowledge (reading). The three cohorts of students were collected from the same school district in three consecutive school years and thus likely provide three very similar samples of students for comparison. However, background variables were used to statistically control for differences that may exist across cohorts. Gender was

coded 0 = male, 1 = female. Students' race/ethnicity was dummy coded into three categories: White, Black, and Other (Asian, Hispanic, American Indian). Free/reduced lunch status was used as a proxy for socioeconomic status (0 = regular lunch, 1 = free/reduced). Stanford 9 reading total scaled scores were included as a control for prior knowledge.

Analysis

In order to test for differences in quantitative literacy across the three different cohorts, it was necessary to build and test a baseline measurement model of quantitative literacy (QLT model) for these elementary-aged students. The building of the QLT model in this study followed a model generation scenario (Joreskog 1993). Quantitative literacy was posited as a second-order three-factor model (Wilkins, in press, 2010) and then tested for its fit to the data from the first cohort of students using a confirmatory factor analysis (CFA). Based on model fit, a series of post hoc respecifications were made to the model and retested to create a model that better fit the data. This new model was then cross-validated with data from the second and third cohorts of students. Model respecifications were guided by fit statistics, but changes were made only if they reflected meaningful relationships or were consistent with theory. The goal of these initial analyses was to create a baseline measurement model that was acceptable across all three cohorts.

Following model generation, a multiple-groups analysis was conducted to test for model structure invariance across the three cohorts of students. In order to validly compare quantitative literacy scores, it is necessary to test for model invariance to determine whether the same construct has been measured for all three groups. Model invariance provides evidence that any differences found between groups are not merely artifacts of measurement differences in the models.

Finally, using the QLT model, a test for differences in latent quantitative literacy scores by cohort was conducted. In this study, a quasi-experimental design was used. Student latent quantitative literacy scores were regressed on dummy coded variables representing cohort membership (i.e., opportunity to learn using *Investigations*) while controlling for student background characteristics.

Model creation and testing were conducted using AMOS 22.0.0 (IBM Corporation 2013). Judgment of model fit was guided by four fit indexes (Kline 2011; Byrne 2010): (a) the model chi-square statistic, (b) the comparative fit index (CFI), (c) the root mean square error of approximation (RMSEA) with its 90 % confidence interval, and (d) the standardized root mean square residual (SRMR). The chi-square statistic has been found to be sensitive to sample size and model

complexity, and thus, other indexes are also considered that take into account sample size and model complexity. The CFI takes on values between 0 and 1 with values closer to 1 showing better fit; values greater than 0.90 usually show adequate model fit (Kline 1998; Hoyle 1995) with values close to 0.95 showing good model fit (Hu and Bentler 1999); although Marsh et al. (2004) have cautioned against applying 0.95 as a stringent cutoff. The RMSEA and SRMR take on values between 0 and 1 with values closer to 0 showing better model fit. Values for the SRMR less than 0.05 represent good model fit (Byrne 2010), although Hu and Bentler (1999) suggest that values as high as 0.08 show good model fit. Values of the RMSEA less than 0.05 show good model fit, and values between 0.05 and 0.08 represent reasonable model fit (Browne and Cudeck 1993). The 90 % confidence interval (CI) of the RMSEA with its associated p value is used to test hypotheses of close fit ($p > 0.05$) and poor fit (upper bound of CI < 0.10).

These statistics were also used to test for model invariance across the cohorts. Tests of model invariance compare fit statistics of a baseline model with subsequent models with additional constraints. For example, if the difference in the chi-square statistic from two models is not found to be statistically significant then the model is assumed to be invariant with respect to the given constraint. However, recent research has shown that the difference in chi-square is also sensitive to sample size and non-normality (Cheung and Rensvold 2002). Therefore, in this study, the difference in chi-square is reported, but final decisions of model invariance are based primarily on two criteria: (a) the fit of the constrained model to the data and (b) whether the difference in the CFI from one model to the next is less than 0.01 (Byrne and Stewart 2006).

Missing data existed for all variables in the sample (see Table 1). The amount of missing data differed by variable as different sets of variables were collected at different times. For the most part, student background variables were complete with the exception of the reading scores (7.8 %) and free/reduced lunch status for five students (0.2 %). Missing data were handled using pairwise deletion methods. Pairwise deletion removes missing data only for variables that are being used in a particular statistical computation. In this study, pairwise deletions were carried out by calculating the correlation matrix among all variables in the study; thus, correlations were calculated for available data for each pair of variables. This matrix was then used in AMOS to conduct the analyses described above.

If data are not missing completely at random, results from statistical analyses can be biased depending on the amount of data missing by different groups. For each of the cognitive measures, *problem*, *concept*, and *skills*,

1.2 % of the data was missing representing a relatively insignificant amount of missing data. For each of the disposition measures, *Self-Concept*, *Interest*, and *Utility*, 11.2 % of the data was missing. For the six items used to measure *Beliefs*, the percent of missing data ranged from 11.7 to 12.7 %. The amount of missing data for *Beliefs* and *Dispositions*, while not negligible, represents a reasonably small amount of missing data. The amount of missing data for these variables was examined by demographic groups (gender, race/ethnicity, and free/reduced lunch status); and missingness was further examined across variables created from the NSRE and EMAS as these surveys were administered at different times. A chi-square test of association found no relationship between amount of missing data and demographic group (i.e., gender, race/ethnicity, free/reduced lunch status). A test of association found no relationship between missingness for variables created from the NSRE and the EMAS (e.g., comparing missingness for the *PROBLEM* and *Self-Concept* variable). A t - test found no statistical difference in reading scores by missing data status for each of the *Beliefs* and *Disposition* variables. However, there was a statistically significant difference for the *Cognition* variables suggesting that, on average, the 31 students with missing data for the cognition variables tended to have lower reading scores (however, the small amount of missing data for the cognition variables seems to mitigate this finding). Taken all together, the evidence suggests that missing data cannot be systematically attributed to student demographics or completion of one or both of the NSRE or the EMAS, thus lessening the chance that statistical results will be biased due to missing data. However, it is important that results be interpreted in light of the potential biases that could be due to missing data.

Results and discussion

Establishing a baseline measurement model of quantitative literacy

In this study, quantitative literacy is modeled as a second-order factor (Wilkins, in press, 2010). That is, each of the three first-order factors (*Beliefs*, *Disposition*, and *Cognition*) is explained by a person's quantitative literacy (the second-order factor). Further, *Self-concept*, *Utility*, and *Interest* were used as measures of the three components related to student disposition (Wilkins, in press, 2010). The three sub-scores from the NSRE (*problem*, *concept*, and *skills*) were used as measures associated with students' mathematical cognition. Finally, the six items (B1–B6) identified to be associated with the simplicity component of epistemological beliefs associated with mathematics were used to form a latent measure, *Beliefs*.

It was first necessary to establish a baseline model that could be used to make comparisons across the three cohorts. Using the first cohort of students ($N = 815$), estimation of the QLT model described above resulted in an overall χ^2 value of 317.90, $p < 0.001$, with 51 degrees of freedom. Based on the χ^2 value, the exact-fit hypothesis was rejected. Together, the CFI = 0.89, SRMR = 0.069, and RMSEA = 0.08 also suggested inadequate model fit; however, the magnitude of the indexes suggested promise for the model and warranted inspection for potential model respecification. Furthermore, the pattern coefficients were all found to be statistically significant and consistent with the theorized model (cf. Wilkins, in press, 2010). An inspection of the standardized residual covariance matrix (SRCM) identified several large residuals associated with the covariances among the items within the *Beliefs* factor; the modification indices (MI) associated with these items also confirmed the possible underestimation of the covariances among these items and suggested that a model that estimated some of these covariances would better represent the data. An examination of the wording of the individual items with large residual covariances revealed three smaller clusters of items (items B1 and B3; items B4, B5, and B6; items B1, B4, and B5). Items B1 and B3 share a common theme related to equating mathematics to memorization; items B4, B5, and B6 share the common theme of mathematics as a set of rules; and items B1, B4, and B5 share a common theme of memorizing rules. Each of these sub-clusters reflects the larger notion of the simplicity of mathematics but also reflect a meaningful sub-theme within the larger *Beliefs* factor. Thus, it is reasonable that the items share residual covariance and it makes sense for this covariance to be estimated in the model. Other large residual covariances found between factors could not be substantively justified, and so covariances for these relationships were not specified in the model.

Thus, a series of model respecifications based on the SRCM and MI were made for the *Beliefs* factor and subsequent tests were conducted. Estimation of the final respecified model resulted in a χ^2 value of 211.94, $p < 0.001$, with 45 degrees of freedom. The reduction in the chi-square ($\Delta\chi^2 = 105.96$, $p < 0.001$, with 6 degrees of freedom) was statistically significant indicating a better fit to the data. Furthermore, the CFI = 0.934 and SRMR = 0.056 suggest adequate fit to the data. The RMSEA = 0.068, with 90 % CI = (0.059, 0.077), and $p_{close} = 0.001$, also suggests reasonable fit, although the close-fit hypothesis was rejected (i.e., $p_{close} < 0.05$), but the poor-fit hypothesis was also rejected (i.e., upper bound of the 90 % CI < 0.10). Again, all pattern coefficients were statistically significant and consistent with theory. This general QLT model is presented in Fig. 1 (note that the coefficients in the model are

estimated for all three groups combined into one sample and will be discussed later).

This respecified model was cross-validated using the data from the second cohort ($N = 837$). Estimation of the QLT model using these data resulted in an overall χ^2 value of 161.65, $p < 0.001$, with 45 degrees of freedom. Based on the χ^2 value, the exact-fit hypothesis was rejected. However, based on the RMSEA = 0.056, with 90 % CI = (0.047, 0.065), and $p_{close} = 0.148$, the close-fit hypothesis was not rejected (i.e., $p_{close} > 0.05$), and the poor-fit hypothesis was rejected (i.e., upper bound < 0.10). The CFI = 0.952 and SRMR = 0.048 suggest good fit to the data. An examination of the SRCM found no significant residual covariances among the items in the *Beliefs* factor. All pattern coefficients were statistically significant and consistent with theory. Furthermore, findings provide validation for the model respecifications made to the first model suggesting that the changes do not merely represent artifacts of the sample data.

The QLT model was then estimated using the data from the third cohort ($N = 838$). Estimation of the model using these data resulted in an overall χ^2 value of 203.53, $p < 0.001$, with 45 degrees of freedom. Based on the χ^2 value, the exact-fit hypothesis was rejected. However, the RMSEA = 0.065, with 90 % CI = (0.056, 0.074), and $p_{close} = 0.003$, suggests reasonable fit; the close-fit hypothesis was rejected (i.e., $p_{close} < 0.05$), but the poor-fit hypothesis was rejected (i.e., upper bound < 0.10). The CFI = 0.942 and SRMR = 0.056 suggest reasonably good fit to the data. An examination of the SRCM found no significant residual covariances among the items in the *Beliefs* factor. All pattern coefficients were statistically significant and consistent with theory. Again, estimation of this model with the third sample of data provides additional validation for the model.

Testing model invariance across the three cohorts

A multiple-groups analysis was conducted to compare the measurement and structural invariance of the QLT model across the three cohorts. Based on the earlier model testing, a baseline multigroup model of quantitative literacy consistent with theory (Wilkins, in press, 2010) has now been created and found to fit the data from each cohort reasonably well. At this point, it is important to establish structural invariance for the model across the three cohorts. This provides evidence that the model operates similarly for each cohort, and thus, differences in the cohorts can be tested with confidence that the differences are more likely due to interventions and not an artifact of measurement differences. Testing for invariance of first- and second-order factor loadings and factor variances is usually sufficient to declare model invariance (Byrne 2010). First, a test for configural invariance was conducted to create a baseline for

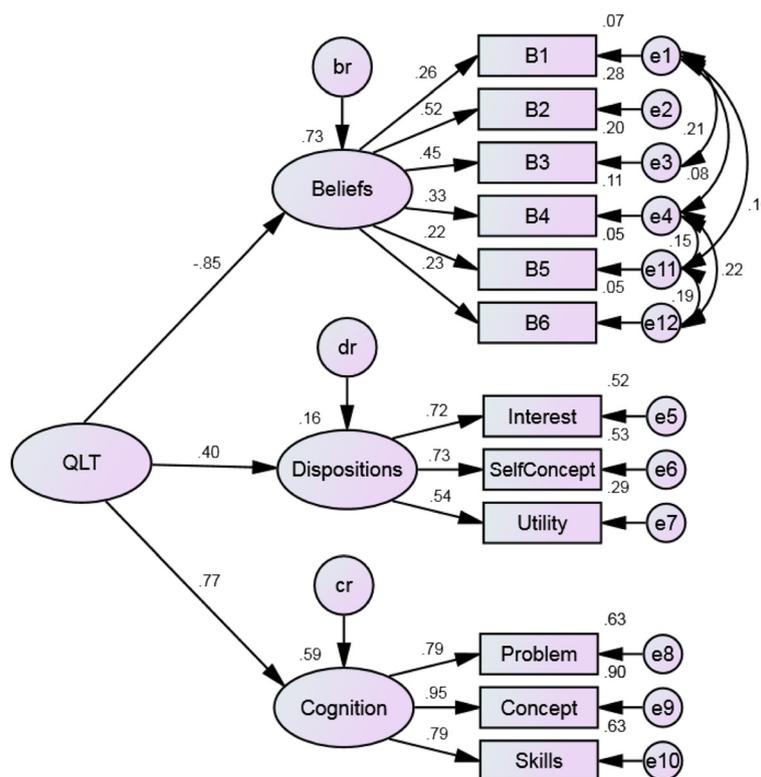


Fig. 1 Measurement model of quantitative literacy

subsequent comparisons. Next, tests for invariance of first-order factor loadings, second-order factor loadings, and structural variances were conducted. Beyond this, tests for invariance of structural residuals and measurement residuals were also conducted.

A test of configural invariance was conducted by estimating the QLT model for all three cohorts simultaneously. In this test, no constraints are imposed on the model. The goal of this test is to determine if the baseline model structure is the same for all three cohorts; further, this model provides baseline statistics for subsequent model comparisons with more strict constraints (see Table 2). Estimation of this unconstrained model (Model 1) resulted in the following fit indexes (see Table 2): $\chi^2 = 577.12$, $p < 0.001$, with 135 degrees of freedom; CFI = 0.942; and RMSEA = 0.036, with 90 % CI = (0.033, 0.039), and $pclose = 1.00$. Overall these indexes provide evidence of good fit to the data and evidence of configural invariance.

Given configural invariance, a series of tests was conducted (see Table 2). For each test, increased constraints were added to the model, first constraining the first-order factor loadings to be equal across the cohorts (Model 2). This was followed by constraining the second-order factor loadings to be equal across the three cohorts (Model 3). In Model 4, factor variances were constrained. Finally, structural residuals and measurement residuals were constrained in Model 5 and Model 6, respectively. In each

case, the subsequent model was compared to the preceding model based on model fit and change in the CFI. Although, the change in chi-square value was significant for Model 2 and 6, in both cases, the Model was found to have good fit and the change in the CFI was less than 0.01. In all other comparisons, the models were found to have good fit, non-significant differences in chi-square and negligible differences in the CFIs. Based on these tests, the QLT model was determined to have structural and measurement invariance across the three cohorts.

A final model of quantitative literacy

Given model invariance across the three cohorts, the three samples of students were combined to estimate model fit for the entire sample. Using this sample of students, estimation of the QLT model resulted in an overall χ^2 value of 430.19, $p < 0.001$, with 45 degrees of freedom. Based on the χ^2 value, the exact-fit hypothesis was rejected. The RMSEA = 0.059, with 90 % CI = (0.054, 0.064), and $pclose = 0.002$, suggests reasonably good model fit, and the poor-fit hypothesis was rejected (upper bound of 90 % CI < 0.10). Furthermore, the CFI = 0.950 and the SRMR = 0.048 suggest good fit to the data. Overall these statistics provide evidence of good model fit and construct validity of the model. This final model is presented in Fig. 1. All pattern coefficients³ were statistically significant (see Fig. 1 and Table 3) and

Table 2 Fit statistics for testing invariance of second-order factor model of QLT across cohorts

Model	χ^2	df	CFI	SRMR	RMSEA (90 % CI) (pclose)	Model comparison	Δ CFI	$\Delta\chi^2$	Δ df
Model 1 configural invariance	577.12***	135	0.942	0.056	0.036 (0.033–0.039) (1.00)	–			
Model 2 first-order factor loadings invariant	618.31***	153	0.939	0.061	0.035 (0.032–0.038) (1.00)	2 vs. 1	0.003	41.19***	18
Model 3 first- and second-order factor loadings invariant	621.67***	157	0.940	0.061	0.034 (0.032–0.037) (1.00)	3 vs. 2	0.001	3.36	4
Model 4 first- and second-order factor loadings and structural variances invariant	623.68***	159	0.940	0.062	0.034 (0.031–0.037) (1.00)	4 vs. 3	0.000	2.02	2
Model 5 first- and second-order factor loadings, structural variances, and structural residuals invariant	628.66***	165	0.940	0.062	0.034 (0.031–0.036) (1.00)	5 vs. 4	0.000	4.98	6
Model 6 first- and second-order factor loadings, structural variances, structural residuals, and measurement residuals invariant	713.26***	201	0.933	0.062	0.032 (0.029–0.035) (1.00)	6 vs. 5	0.007	84.98***	36

*** $p < 0.001$

consistent with the theorized model (Wilkins, in press, 2010). That is, students’ quantitative literacy was characterized by an interrelationship among mathematical cognition, disposition, and beliefs in the simplicity of mathematical knowledge, e.g., increased mathematical cognition is related to positive disposition and decreased beliefs in the simplicity of mathematics and inversely.

Predicting student quantitative literacy

The relationship between students’ quantitative literacy and the opportunity to learn mathematics using a *Standards*-based mathematics curriculum was investigated at the district level using a series of regression

Table 3 Standardized regression weights for measurement model of quantitative literacy

	QLT	BELF	COG	DISP
Beliefs	-0.853			
Cognition	0.769			
Disposition	0.400			
B1		0.257		
B2		0.525		
B3		0.448		
B4		0.334		
B5		0.219		
B6		0.230		
Problem			0.794	
Concept			0.949	
Skills			0.792	
Interest				0.720
Self-Concept				0.728
Utility				0.543

All regression weights were statistically significant ($\alpha < 0.001$). QLT = quantitative literacy, BELF = beliefs, COG = cognition, DISP = disposition. $\chi^2 (45) = 440.19$, $p < 0.001$; CFI = 0.950; SRMR = 0.048; RMSEA = 0.059, 90 % CI = (0.054, 0.064) pclose = 0.002

models. Quantitative literacy was measured as a latent construct as described by the measurement model above (see Fig. 1, Table 2).

First, a Baseline Model, regressing student quantitative literacy on student background variables was estimated (see Table 4). Because the design of the study was quasi-experimental, this baseline model served to statistically control for differences in these variables that might exist across cohorts. Prior knowledge was found to be a statistically significant predictor of student QLT scores ($\beta = 0.664$; $b = 0.003$, $SE < 0.001$, $p < 0.001$). That is, on average, students with higher reading scores were found to have increased quantitative literacy and inversely. Reading scores were found to be the strongest predictor of QLT. Students receiving free or reduced lunch, on average, were found to have lower QLT scores ($\beta = -0.161$; $b = -0.092$, $SE = 0.012$, $p < 0.001$). Gender was found to be a statistically significant predictor of QLT ($\beta = -0.033$; $b = 0.015$, $SE = 0.006$, $p < 0.05$), indicating that, on average, females have lower QLT scores. Furthermore, compared to students categorized as White, on average, students categorized as Black were found to have lower QLT scores, and this difference was statistically significant ($\beta = -0.135$; $b = -0.089$, $SE = 0.013$, $p < 0.001$). Students categorized as Other were not found to be statistically different from students categorized as White ($\beta = 0.023$; $b = -0.020$, $SE = 0.013$). Overall, these background variables explained 62.5 % of the variance in the latent QLT scores.

In order to investigate the relationship between QLT scores and exposure to *Investigations*, a second model, the Curriculum Model, was estimated (see Table 4). This model included dummy coded variables that made it possible to compare students in Cohorts 2 and 3 to students in Cohort 1. By not including the Cohort 1 variable in the model, coefficients for the Cohort 2 and Cohort 3 variables represented differences in QLT scores compared to Cohort 1. After controlling for student background

Table 4 Regression of student quantitative literacy on background variables and opportunity to use a *Standards*-based curriculum

	Baseline model			Curriculum model		
	β	<i>b</i>	SE	β	<i>b</i>	SE
Gender	-0.033	0.015*	0.006	-0.033	-0.014*	0.006
Black (compared to White)	-0.135	-0.089***	0.013	-0.136	-0.090***	0.012
Other (compared to White)	0.023	0.020	0.013	0.020	0.017	0.013
Free/reduced Lunch	-0.161	-0.092***	0.012	-0.159	-0.090***	0.012
Prior achievement	0.664	0.003***	0.000	0.668	0.003***	0.000
Cohort 2 (compared to Cohort 1)				0.041	0.019*	0.008
Cohort 3 (compared to Cohort 1)				0.102	0.048***	0.009
R^2	0.625			0.640		

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

variables, the difference in QLT scores between Cohort 1 and Cohort 2 was found to be statistically significant ($\beta = 0.041$; $b = 0.019$, $SE = 0.008$, $p < 0.05$) in favor of Cohort 2. The difference in QLT scores between Cohort 1 and Cohort 3 was also found to be statistically significant ($\beta = 0.102$; $b = 0.048$, $SE = 0.009$, $p < 0.001$) in favor of Cohort 3. That is, on average, students in the school district who were surveyed after implementation of *Investigations* were found to have increased quantitative literacy when compared to students surveyed prior to the implementation of *Investigations*. The Curriculum Model explained 64.0 % of the variance in latent QLT scores. After controlling for background variables, the curriculum variables explained an additional 2.4 % of the variance in latent QLT scores relative to the variance explained by the Background Model. Furthermore, an additional test comparing QLT scores for Cohort 2 and Cohort 3 found the difference to be statistically significant ($\beta = 0.062$; $b = 0.029$, $SE = 0.008$, $p < 0.001$) indicating an overall increase for Cohort 3 above and beyond the increased change made by Cohort 2.

Conclusions

The purposes of this study were to create and validate a QLT model for elementary-aged students and to use this model to examine the relationship between student quantitative literacy and exposure to a *Standards*-based mathematics curriculum. The results of the study provide evidence for the validity of the QLT model for the sample of elementary-aged children. In this model, the quantitative literacy construct is represented as a second-order factor comprising three first-order factors: mathematical beliefs, mathematical disposition, and mathematical cognition (see Fig. 1). Furthermore, the model highlights the interrelationship among these factors that is consistent with the theoretical notion of quantitative literacy proposed by Wilkins (in press, 2010). By using a sample of elementary-aged students, the results of this study provide further evidence for the validity of a generalized model of

quantitative literacy as previous research has validated the model for secondary and undergraduate students (Wilkins, in press, 2010).

Using the measurement model of quantitative literacy, QLT scores for students in the school district using *Investigations* were compared to scores for students in the same school district before the implementation of *Investigations*. Overall, Grade 4 students in the school district after implementation had increased QLT scores when compared to Grade 4 students in the district prior to the implementation of *Investigations*. Moreover, a comparison of students with 4 years of exposure to the curriculum to students with only 2 years of exposure, found, on average, a statistically significant increase in QLT scores for those students with 4 years of exposure. Compared with the amount of variance in QLT scores explained by the Background Model, the amount of additional variance explained by the Curriculum Model was relatively low, and its practical significance should be considered with caution. However, the evidence does suggest a positive relationship between the use of a *Standards*-based curriculum and the development of quantitative literacy. Compared with prior research that has only shown increased mathematics achievement (cf., Mokros 2003), findings from this study provide evidence that the use of a *Standards*-based curriculum may lead to increased overall quantitative literacy. By using a measurement model that portrays the holistic notion of quantitative literacy, instead of measuring different components and aggregating findings, we are better able to assess whether *Standards*-based curricula promote the whole of quantitative literacy.

In addition to the measurement of quantitative literacy, the findings from this study have implications for assessment and curriculum development. From an assessment perspective, this study provides an example of how quantitative literacy can be assessed in a holistic manner. Most often, quantitative literacy is assessed with a focus solely on mathematical and statistical content knowledge using

tasks that contextualize the mathematics in real-world or problem-based situations (e.g., TIMSS [see Orpwood and Garden 1998; Mullis et al. 1998], and Programme for International Student Assessment [PISA, Organisation for Economic Cooperation and Development OECD 2000]). While this form of assessment does test an important aspect of being quantitatively literate, at best, the affective components of quantitative literacy are measured as a secondary consideration and discussed merely as they relate to achievement but not as they interrelate (e.g., TIMSS, PISA). In addition, many colleges and universities have developed quantitative literacy courses and standards (Gillman 2006a; Steen 2004), but again these standards are often assessed based on achievement tests of contextualized mathematics (e.g., Comaz and Martin 2006; Gillman 2006b). Findings from this study suggest that quantitative literacy can be assessed holistically as a single construct that considers the interrelationship among the different aspects of quantitative literacy.

From a curriculum perspective, *Investigations* seems to have the potential to promote quantitative literacy in elementary-aged students. Without further analysis, it cannot be generalized that *Investigations* is a “quantitative literacy curriculum,” but it is possible that *Investigations* could serve as a model that could be further studied to understand the components of a curriculum that could potentially promote quantitative literacy. In order to promote the notion of quantitative literacy outlined in this study, it would be important to understand how best to motivate children in a way that builds quantitative literacy as a habit of mind. That is, to help students create beliefs about mathematics as accessible, constructible, and justifiable by all, coupled with a value for mathematics and a confidence and willingness to engage in quantitative situations.

Limitations and future research

While the conclusions of the current study seem promising, it is important to situate the findings within a larger research trajectory to test mathematics education interventions (cf., Sloane 2008). In other words, it is important to highlight what the current study does not tell us but how the findings add to the research literature. First, this study is limited by its unit of analysis. The unit of analysis is the student, and the “treatment” happens at the level of the school district. From the findings of this study, it is not clear how students within different schools compared across the three cohorts⁴ nor is it clear how students within different classrooms or with different teachers compared across the three cohorts.⁵ It is quite possible for the findings to be different or even opposite at different levels of analysis (see e.g., Wilkins, 2004b). For example, some schools could have shown an increase across the cohorts, while others could have shown a

decrease, but on average, the school district could still show an increase. Similarly, this could have happened at the classroom level. Documenting the patterns of differences across schools and classrooms would enhance the current findings. In addition, by documenting these patterns of differences, they could potentially be modeled using other school- or teacher-level variables further adding to the understanding of how the intervention contributed to students’ overall quantitative literacy.

In this study, implementation (i.e., treatment) means that the curriculum was adopted by the school district, and thus, it can only be said that students are in a district that adopted the curriculum. In this study, there was no statistical or design-based control for the fidelity of curriculum implementation. Thus, it cannot be assumed that all teachers used the curriculum in the same way nor that all students received the same opportunities to use the curriculum, except to say, that at least they are in a school district and school that has adopted *Investigations*. Teachers were involved in professional development experiences to help them implement the curriculum appropriately, but beyond this, except for documentation by the mathematics supervisors that teachers in the schools were using the curriculum, there was no systematic documentation of implementation.

Sloane (2008) points out that prior to effectiveness studies of interventions in mathematics education, it is important, if not necessary, to conduct smaller scale studies to determine the feasibility of such larger scale studies. For example, these smaller scale studies may contribute to the larger research trajectory by developing measures of constructs and by providing preliminary tests of interventions through student-level quasi-experimental studies. Despite its limitations, the current study did just that. In this study, a measurement model of quantitative literacy was tested, and preliminary evidence from the study that students in a school district using *Investigations* had higher levels of quantitative literacy was documented. Future studies would enhance these findings by expanding the scope to include a multilevel design that considers the effects of classroom and school as interacting units of analysis. Beyond the research design, having teachers randomly assigned to treatment groups representing controlled professional development and curriculum implementation would make it possible to test the efficacy of the intervention. The findings of the current study provide the groundwork for additional research with this broader scope and research design.

Endnotes

¹In the first year that data was collected there were only 15 elementary schools in the district.

²Harcourt Brace, Inc. merged with Pearson in 2008.

³Structure coefficients were also examined. Structure coefficients are measures of association between variables. It is important that the correlations between variables not assigned to higher-order factors be less than those assigned to a factor. For example, the correlations between the three lower-order factors of cognition and the *Cognition* factor should be greater than the correlations between these three variables and the other higher-order factors, *Beliefs* and *Dispositions*. This was found to be the case for the factors and observed variables in the model.

⁴The small number of schools in the study makes a school-level analysis limited in statistical power. However, based on the use of a slopes-as-outcomes model (Burstein, 1980), the regression analysis was conducted separately for each school and regression coefficients compared across schools. One school's coefficients seemed to suggest that it was an outlier which was consistent with an earlier personal communication with the district mathematics supervisor. Once this school was removed from this analysis, the average school-level differences were relatively consistent with the district-level findings.

⁵Recall that it was not possible to link students and teachers across the different cohorts.

Abbreviations

CCSSO: Council of Chief State School Officers; CFA: Confirmatory Factor Analysis; CFI: Comparative Fit Index; CI: confidence interval; EMAS: Elementary Mathematics Attitude Survey; MI: modification index; NCTM: National Council of Teachers of Mathematics; NGA: National Governor's Association; NSRE: New Standards Reference Exam; OECD: Organisation for Economic Cooperation and Development; PISA: Programme for International Student Assessment; QLT: quantitative literacy; RMSEA: root mean square error of approximation; SIMS: Second International Mathematics Study; SRCM: standardized residual covariance matrix; SRMR: standardized root mean square residual; TIMSS: Third International Mathematics and Science Study.

Competing interests

The author declares that he has no competing interests.

Authors' contributions

JLMW is the sole author of the paper.

Acknowledgements

This study was supported in part by the National Science Foundation Grant #9911558. Any conclusions stated here are those of the author and do not necessarily reflect the position of the National Science Foundation. Virginia Tech's Open Access Subvention Fund supported the publication of this article.

Received: 10 August 2015 Accepted: 10 November 2015

Published online: 04 December 2015

References

- American Association for the Advancement of Science. (1990). *Science for all Americans*. New York: Oxford University Press.
- Atkins, M., & Helms, J. (1993). Getting serious about priorities in science education. *Studies in Science Education*, 21, 1–20.
- Briars, D. J., & Resnick, L. B. (2000). *Standards, assessments—and what else? The essential elements of standards-based school improvement (CSE Technical Report 528)*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Available online: <https://www.cse.ucla.edu/products/reports/TECH528.pdf>.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 445–455). Newbury Park: Sage.
- Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education*, 8, 158–233.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: basics concepts, applications, and programming* (2nd ed.). New York: Routledge.
- Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: a walk through the process. *Structural Equation Modeling*, 13(2), 287–321.
- Carroll, W. (1997). Results of third-grade students in a reform curriculum on the Illinois state mathematics test. *Journal for Research in Mathematics Education*, 28(2), 237–242.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255.
- Cockcroft, W. H. (1982). *Mathematics counts*. London: Her Majesty's Stationary office.
- Comaz, D., & Martin, W. O. (2006). Quantitative literacy as an integral component of mathematics curriculum, case at North Dakota State University. In R. Gillman (Ed.), *Current practices in quantitative literacy* (pp. 155–163). Washington, DC: Mathematical Association of America.
- Dossey, J. A. (1997). Defining and measuring quantitative literacy. In L. A. Steen (Ed.), *Why numbers count: quantitative literacy for tomorrow's America* (pp. 173–186). New York: College Entrance Examination Board.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of an actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, 21, 215–225.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motives* (pp. 75–146). San Francisco, CA: W. H. Freeman and Company.
- Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitude scales: instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, 7, 324–326.
- Gal, I. (1997). Numeracy: imperatives of a forgotten goal. In L. A. Steen (Ed.), *Why numbers count: quantitative literacy of tomorrow's America* (pp. 36–44). New York: College Entrance Examination Board.
- Gillman, R. (Ed.) (2006a). *Current practices in quantitative literacy*. Washington: Mathematical Association of America
- Gillman, R. (2006b). A case study of assessment practices in quantitative literacy. In R. Gillman (Ed.), *Current Practices in Quantitative Literacy* (pp. 165–169). Washington: Mathematical Association of America
- Gonzalez, E. J., & Smith, T. A. (Eds.). (1998). *User guide for the TIMSS international database: primary and middle school years, 1995 assessment*. Chestnut Hill, MA: TIMSS International Study Center.
- Harcourt Educational Measurement. (1996). *New standards reference exam—mathematics*. San Antonio: Harcourt, Inc.
- Harcourt Educational Measurement. (1997). *New standards reference exam—mathematics*. San Antonio: Harcourt, Inc.
- Harcourt Educational Measurement. (1998). *New standards reference exam—mathematics*. San Antonio: Harcourt, Inc.
- Harcourt Educational Measurement. (1999). *New standards reference exam—mathematics*. San Antonio: Harcourt, Inc.
- Harris, K., Marcus, R., McLaren, K., & Fey, J. (2001). Curriculum materials supporting problem-based teaching. *School Science and Mathematics*, 101(6), 310–318.
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, 25, 378–405.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and their relation to learning. *Review of Educational Research*, 67(1), 88–140.
- Hoyle, R. H. (1995). The structural equation modeling approach: basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 1–15). Thousand Oaks: Sage.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- IBM Corporation. (2013). *IBM SPSS AMOS, Version 22.0*. Armonk, NY: IBM Corporation.

- Joreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park: Sage.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.
- Madison, B. L., & Steen, L. A. (2003). *Quantitative literacy: why numeracy matters for schools and colleges*. Princeton: National Council on Education and the Disciplines. Retrieved from: http://www.maa.org/external_archive/QL/qltoc.html.
- Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341.
- McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Buglari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: the case of standards-based high school mathematics. *Journal for Research in Mathematics Education*, 32(5), 493–517.
- Mokros, J. (2003). Learning to reason numerically: the impact of *Investigations*. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: what are they? What do students learn?* (pp. 109–131). Mahwah: Lawrence Erlbaum Associates.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- National Commission on Excellence in Education. (1983). *A nation at risk: the imperative for education reform*. Washington: US Department of Education.
- National Council of Teacher of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston: NCTM.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: NCTM.
- National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: a quest for coherence*. Reston: NCTM.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards (mathematics)*. Washington: NGA Center, CCSSO.
- National Research Council. (1989). *Everybody counts: a report to the nation on the future of mathematics education*. Washington: National Academy Press.
- Organisation for Economic Cooperation and Development (OECD). (2000). *Measuring student knowledge and skills. The PISA 2000 measurement of reading, mathematical and scientific literacy*. Paris: OECD.
- Orpwood, G., & Garden, R. A. (1998). *Assessing mathematics and science literacy*. Vancouver: Pacific Educational Press.
- Post, T. R., Harwell, M. R., Davis, J. D., Maeda, Y., Cutler, A., Andersen, E., et al. (2008). Standards-based mathematics curricula and middle-grades students' performance on standardized achievement tests. *Journal for Research in Mathematics Education*, 39(2), 184–212.
- Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 34(1), 74–95.
- Riordan, J. E., & Noyce, P. E. (2001). The impact of two Standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education*, 32(4), 368–398.
- Russell, S. J., Tierney, C., Mokros, J., & Economopoulos, K. (2004). *Investigations in number, data, and space*. Glenview: Scott Foresman.
- Senk, S. L., & Thompson, D. R. (Eds.). (2003). *Standards-based school mathematics curricula: what are they? What do students learn?* Mahwah: Lawrence Erlbaum Associates.
- Sloane, F. C. (2008). Randomized trials in mathematics education: recalibrating the proposed high watermark. *Educational Researcher*, 37(9), 624–630.
- Steen, L. A. (Ed.). (1990). *On the shoulders of giants: new approaches to numeracy*. Washington: National Academy Press.
- Steen, L. A. (1997). Preface: the new literacy. In L. A. Steen (Ed.), *Why numbers count: quantitative literacy for tomorrow's America* (pp. xv–xxviii). New York: College Entrance Examination Board.
- Steen, L. A. (1999). Numeracy: the new literacy for a data-drenched society. *Educational Leadership*, 57(2), 8–13.
- Steen, L. A. (Ed.). (2001). *Mathematics and democracy: the case for quantitative literacy*. United States: The Woodrow Wilson National Fellowship Foundation.
- Steen, L. A. (2004). *Achieving quantitative literacy: an urgent challenge for higher education*. Washington: Mathematical Association of America.
- Thompson, D. R., & Senk, S. L. (2001). The effects of curriculum on achievement in second-year algebra: the example of the University of Chicago school mathematic project. *Journal for Research in Mathematics Education*, 32(1), 58–84.
- Westbury, I., & Thalathoti, V. V. (1989). *United States—population B*. Urbana: The Board of Trustees of the University of Illinois.
- Wiley, D. E., & Resnick, L. (1997). *The new standards reference examination standards-referenced scoring system (CSE Technical Report 470)*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Available online: <http://www.cse.ucla.edu/products/Reports/TECH470.pdf>.
- Wilkins, J. L. M. (in press). An assessment of the quantitative literacy of undergraduate students. *Journal of Experimental Education*.
- Wilkins, J. L. M. (2010). Modeling quantitative literacy. *Educational and Psychological Measurement*, 70(2), 267–290.
- Wilkins, J. L. M. (2004a). *Elementary Mathematics Attitude Survey (EMAS)*. Blacksburg, VA: Virginia Tech.
- Wilkins, J. L. M. (2004b). Mathematics and science self-concept: An international investigation. *Journal of Experimental Education*, 72(4), 331–346.
- Wilkins, J. L. M. (2000). Preparing for the 21st century: The status of quantitative literacy in the United States. *School Science and Mathematics*, 100(8), 405–418.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com