Taylor & Francis
Taylor & Francis Group

International Journal of Mathematical Education in Science and Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmes20

Temporal relationship between attitude toward mathematics and mathematics achievement

Henry Nsubuga Kiwanuka , Jan Van Damme , Wim Van den Noortgate & Chandra Reynolds

INTERNATIONAL JOURNAL OF MATHEMATICAL EDUCATION IN SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/0020739X.2020.1832268

Taylor & Francis
Taylor & Francis Group

Check for updates

Temporal relationship between attitude toward mathematics and mathematics achievement

Henry Nsubuga Kiwanuka[a], Jan Van Damme[b], Wim Van den Noortgate[c] and Chandra Reynolds[d,e]

[a]Faculty of Science, Uganda Martyrs University, Kampala, Uganda; [b]Centre for Educational Effectiveness and Evaluation, KU Leuven, Belgium; [c]Faculty of Psychology and Educational Science, Methodology of Educational Sciences Research Group, KU Leuven, Belgium; [d]Department of Psychology, University of California, Riverside, CA, USA; [e]St. Pius Church, Orange County, CA, USA

ABSTRACT

This study examined the temporal relationship between attitude toward mathematics and mathematics achievement among first-year secondary (i.e. middle) school students (grade 7, about 14–15 years) with a focus on sex differences in Central Uganda. Random-intercept cross-lagged panel models, based on structural equation modeling, were used to analyze data which were collected through students' questionnaires and mathematics tests at three measurement points. The proportion of variance at the class and school level was larger for the achievement measures than for the attitude measures. At these two levels, also the correlation coefficients between the two constructs were higher for boys than for girls. The study found evidence in support of the reciprocal-effects model as the best-fitting structural model within and across both sexes. Between the first and second measurement points, there was evidence for the self-enhancement model (attitude influencing achievement), while between the second and third points, the skill-development model (achievement influencing attitude) was supported.

ARTICLE HISTORY
Received 24 June 2020

KEYWORDS
Attitude toward mathematics; mathematics achievement; reciprocal-effects model; self-enhancement and skill-development models; Uganda

CLASSIFICATIONS MSC
Computing methodologies and applications; number theory; numerical methods; ordinary differential equations; social and behavioral sciences

1. Introduction

The relationship between attitude toward mathematics (ATM) and mathematics achievement (MA) has been established in the literature (Bhowmik & Banerjee (Roy), 2016; Dowker et al., 2019; Kundu & Ghose, 2016). However, the interpretation of this correlation remains unclear. One possibility is that positive ATM leads to higher MA. For lower secondary students to become mathematics literates, they must cultivate positive ATM and persist in choosing to take mathematics courses in upper secondary when the courses become optional. An alternative possibility is that high MA enhances positive ATM. Improving MA through the implementation of high national mathematics standards is a major strategy for reaching this goal. A third possibility is that both processes operate in a reciprocal relationship that unfolds over time.

CONTACT Henry Nsubuga Kiwanuka hkiwanuka@umu.ac.ug

Some previous studies in Uganda have investigated only students' MA (Kiwanuka et al., 2015), while others have examined only students' ATM(Khun-Inkeeree et al., 2016; Kiwanuka et al., 2016). Yet, no study has endeavored to establish the relationship between ATM and MA for Uganda secondary school students. To achieve this, this paper will try to capture the complexity of the interrelationships by employing dynamic longitudinal modeling, the random-intercept cross-lagged panel model (RI-CLPM), within a temporal framework (Berry & Willoughby, 2017). Thus, the paper seeks to examine a temporal relationship which is generally the timing between a factor and an outcome, in the hope to shed light on the causality of the relationship (Masih & Masih, 1997).

Mathematics is considered an important tool for many natural sciences and social science disciplines. In spite of this importance accorded to mathematics in society, most African students have performed persistently poorly in mathematics (Gitaari et al., 2013). In 2015, the Secondary Science and Mathematics (SESEMAT) programme was introduced in Uganda to address the then poor performance or high failure rate in science and mathematics at lower secondary level.

A first step in understanding and enhancing MA and ATM may be found in investigating the relationship between students' ATM and their MA over time. Hence, the goal of this study is to test for the unidirectional and reciprocal effects within a cross-lagged modeling framework and to explore possible differences between boys and girls in the temporal relationships by analyzing longitudinal (time-series) data on ATM and MA collected in Central Uganda at the beginning, half-way and end of school year 2012. Evaluating possible temporal relationships has implications for educational practices, diagnosis and intervention in school mathematics, and can also contribute to the understanding of the interplay between the affective and cognitive factors. For this reason, the current study could make a significant contribution to the educational effectiveness research. Indeed, attitude and achievement have been considered as important educational outcomes. Hence, the results of this study can assist educators in turn to help students to develop positive ATM and improve their performance in mathematics.

ATM is defined as a positive or negative emotional disposition toward mathematics (Zan & Di Martino, 2007). From a multidimensional perspective, ATM can be described as the emotion that students associate with mathematics, their beliefs about mathematics and their behaviour towards mathematics. Hence, ATM has cognitive, affective and behavioural components. The pfto the educational resent study defines ATM as an aggregated measure of mathematics self-confidence, perceived usefulness and enjoyment of mathematics. We have employed these attitudinal indicators which are much used in the literature (Dowker et al., 2019; Khun-Inkeeree et al., 2016; Kiwanuka et al., 2016; Van Damme et al., 2004) as important determinants of ATM. Achievement is defined by student ability in computations and solving problems, which can be measured by written tests (Evans, 2007).

The general relationship between attitude and achievement is based on the idea that a positive attitude of a student has towards a subject or task is expected to result in higher achievement level. Students learn more effectively when they like what they learn and they attain better in mathematics if they enjoy it (Ma & Kishor, 1997). Therefore, mathematics educators and teachers have traditionally directed their attention towards creating, developing and reinforcing positive attitudes towards any subject (Opolot-Okurut, 2005).

Nevertheless, research literature has failed to provide consistent findings regarding the ATM-MA relationship. While some studies report a statistically significant positive relationship between the two variables (Dowker et al., 2019; Kundu & Ghose, 2016), other studies have found a negative, or at least no clear ATM-MA relationship (Phonguttha et al., 2009).

To evaluate the magnitude of the ATM-MA relationship, Ma and Kishor (1997) conducted a meta-analysis on 113 studies. They found that the overall mean effect was positive, with a weaker association in primary schools and a stronger one in secondary schools. Researchers attribute the low correlation among primary school children to the fact that their attitudes, though positive, tend to deteriorate with age during childhood and adolescence (Dowker et al., 2019). Since the ATM-MA correlation generally has been found to be rather low, many researchers have concluded that ATM-MA relationship is weak and cannot be considered to be of statistical or practical significance and to have any useful implications for educational practices (Ma & Kishor, 1997).

Meta-analyses reviewing basic correlation research on this relationship have indicated that the strength of the relationship between attitudes toward and achievement in the subject may differ by student's socio-economic status, sex, age, grade, and ethnicity as well as by the subject domain studied and by the types of achievement and attitude measures used (Ma & Kishor, 1997). In fact, sex has been one of the most important moderators of the relationship between attitude and achievement. Mubeen et al. (2013) found that the relationship between ATM and MA was significant only for girls but not for boys. Bhowmik and Banerjee (Roy) (2016) reported a statistical difference between ATM and MA in terms of sex. However, in their meta-analysis, Ma and Kishor (1997) found that sex did not have a significant effect on the relationship between ATM and MA. So, a renewed effort to examine the role of sex in the relationship between attitude and achievement appears to be warranted.

Though a positive relationship between ATM and MA has been widely assumed to exist, by using correlation coefficients as measures of the relationship, there is no clear evidence on whether ATM is a cause of MA, or MA is a cause of ATM or if they interact with each other in a complicated manner, their relationship being subject to multiple influences (Ma & Kishor, 1997; Ma & Xu, 2004). Unfortunately, past research in this field provides little evidence on the existence and direction of a temporal relationship between the two variables. So, it is important for educators to understand the nature of this relationship.

1.1. Models of ATM-MA association

Since a major objective of mathematics education is to increase students' MA (Opolot-Okurut, 2005), the ideas that a change in attitudes may somehow 'cause' change in achievement is important to investigate. In order to conclude that there is such a directional relationships between attitudes and achievement, Byrne (1986) proposed three conditions to be fulfilled (a) there must be a statistical relationship between achievement and attitudes, (b) there must be a clear time precedence established, and (c) SEM should be used as a method of analysis. Two other important criteria were added: (d) both attitude and achievement constructs should be measured at least twice, and (e) all latent constructs in the SEM models should be inferred on the basis of multiple indicators (Marsh et al., 1999) . Previous researches have proposed four causal models describing the temporal relationship between ATM and MA, and these models are (1) No cross-effects model, (2) Achievement predominant model, (3) Attitude predominant model, (4) Reciprocal-effects model.

Evidence for the *no cross-effects model* has been reported by Phonguttha et al. (2009) who indicated no relationship between ATM and MA. Mubeen et al. (2013) in Pakistan surveyed 500 students of class IX and X and found that the correlation between ATM and MA was not significant. In a two-wave longitudinal study of the temporal relationships between attitudes toward science and science achievement for rural 7th- and 8th-grade students in the United States (US), Mattern and Schau (2002) observed sex differences. The no cross-effects model exhibited the best fit for girls, in that, their achievement was not linked to their prior attitudes, but was only affected by their prior achievement.

At the country level, results from TIMSS 2011 revealed that East Asian students outperformed students from other countries in mathematics, while they displayed relatively negative mathematics attitudes. This could be attributed to the educational system and the families that have very high expectations which the students struggle to fulfil and face (Mullis et al., 2012). These paradoxical results at the country level have led some researchers to believe that neither attitude nor achievement influences each other.

Second is the *achievement predominant model* in which previous achievement has direct effects on later achievement and attitudes. This concurs with the skill-development model, where students' skill in an academic subject is a cause of their later attitudes toward the subject (Chiu, 2012). Researchers have indicated that achievement influences attitudes (Herbert & Stipek, 2005). Also, Ma and Xu (2004) conducted a study to evaluate the causal ordering between ATM and MA of secondary school students. Their results showed that achievement exhibited a temporal predominance over attitude across the entire secondary school.

A third possible model for the relationship is the *attitude predominant model* in which previous attitudes have direct effects on later attitudes and achievement, but prior achievement affects only later achievement. This model is consistent with the self-enhancement model, where students' attitudes toward a subject influence their later achievement in that subject (Chiu, 2012). Ma and Kishor (1997) found a unidirectional relationship that assumes a temporal predominance of attitude over achievement. Manoah et al. (2011) found that ATM has a positive and direct significant causal effect on the performance in mathematics of junior secondary school students.

The most complex model in which all other models are nested is the *reciprocal-effects model* (also called the *cross-effects model* by Mattern and Schau (2002)). Chiu (2012) proposed that a more realistic compromise between self-enhancement and skill-development models (Chiu, 2012) was a reciprocal effects model in which, prior academic self-concept affects subsequent achievement and prior achievement affects subsequent academic self-concept. Similarly, some researchers (Williams et al., 2005) have argued against any unidirectional temporal relationship between attitude and achievement, in favour of a reciprocal temporal relationship, that is, not only do attitudes affect achievement, but achievement affects attitudes as well. That means that students who have better attitudes towards mathematics demonstrate higher achievement, and students who have higher achievement exhibit better attitudes (Evans, 2007). Mattern and Schau (2002) found that the reciprocal-effects model was the best fitting model for rural 7th- and 8th-grade students in the United States of America.

Many studies found evidence for the four alternative models as illustrated in the above descriptions. However, it is worthy pointing out that some of the researchers, except

Mattern and Schau (2002), only looked at one of these models without comparing the model fit with those of the other models. So, the evidence given by a particular study would be weaker than when a study found that one model is the best model. To overcome this limitation, we have employed RI-CLPM based on longitudinal SEM to compare the model fits of these alternative nested models.

1.2. Research questions

The main research question addressed in this study is the following: What is the nature of the temporal relationship between mathematics achievement and attitudes toward mathematics (i.e. mathematics self-confidence, perceived usefulness and enjoyment of mathematics) for boys and girls in Central Uganda? In order to answer this research question, we examined the following specific questions.

(a) What is the magnitude of the ATM-MA relationship?
(b) Do changes in ATM lead to changes in subsequent MA?
(c) Do changes in MA lead to changes in subsequent ATM?
(d) Is there a difference between boys and girls in the relationships mentioned in the previous questions, controlling for individual background?

2. Methodology

2.1. Participants

Our study used a multistage sampling design. At the first stage, four districts in Central Uganda were purposively chosen. These were: Kampala and Wakiso, which are urban and populated with people from different parts of the country, and Mpigi and Mukono, which are semi-rural but reachable. At the second stage, 60schools were selected from a total of 376 schools in the four districts, with 25 semi-rural and 35 urban schools, thus forming two strata. These schools followed the same curriculum, but with a variety of learning climates and teacher practices. At the third stage, classes were selected according to the number of 7th gradeclasses at a school. Four, three and two classes were randomly selected from schools with five, four and three classes, respectively. In schools with only one and two classes, all classes were selected. The participating sample consisted of 4819 first-year secondary school students (grade 7; about 14–15 years old) who participated in at least one measurement occasion. They were grouped in 78 classes of 49 schools. Data came from student information of Wave 1 ($N = 4768$), Wave 2 ($N = 4531$) and Wave 3 ($N = 4244$). Table 1 describes numbers of sampled schools and classes in the target and participating sample. Out of the 60 targeted schools, 11 did not cooperate or comply with our research procedures. Table 2 describes the categories of the participating schools.

2.2. Procedure

The study involved three waves of data collection during school year 2012witha four-month interval between adjacent measurement occasions. At each measurement occasion, the student questionnaire was administered along with a mathematics test. The researcher and/or his assistants administered the test and questionnaire to the students with the help

Table 1. Number of schools and classes in the target and participating sample groups.

Total no. of 7th grade classes	Target group		Participating sample classes	
	No. of schools	No. of classes per school	No of participating Schools	No. of participating classes
5	4	4	2	8
4	8	3	6	18
3	8	2	5	10
2	7	2	6	12
1	33	1	30	30
Total:	60		49	78

Table 2. Categories of participating schools.

	By category	No. of schools
By type	Government	12
	Private	37
By location	Urban	28
	Semi-urban	21
By sex	All boys	2
	All girls	2
	Mixed	45

of the mathematics teacher(s) during mathematics class time. The students were assured of confidentiality and were told that the data collected in the study would be used only for research purposes.

2.3. Measures

Attitude toward mathematics (ATM) and mathematics achievement (MA) and sex were the major variables in our study.

ATM was measured by three indicators: mathematics self-confidence (CONF; 10 items), perceived usefulness of mathematics (USE; 7 items), and enjoyment of mathematics (ENJOY; 6 items) at Times 1, 2, and 3. Students responded to the indicators on a five-point Likert scale (1 = 'strongly disagree' to 5 = 'strongly agree'). The indicators of CONF and USE were adapted from the Fennema-Sherman Mathematics Attitudes Scale (Fennema & Sherman, 1976), and ENJOY was based on the Attitudes-Toward Mathematics Inventory (Tapia & Marsh, 2004). The mathematics self-confidence scale measures students' confidence in their ability to learn and perform well in mathematics tasks (Fennema & Sherman, 1976). The perceived usefulness of mathematics measures students' beliefs about the value/importance of mathematics in their life now and in the future (Fennema & Sherman, 1976). The enjoyment of mathematics scale measures the extent to which the students enjoy mathematics classes and the subject matter itself (Tapia & Marsh, 2004). For some examples of items, see Appendix. Some items were removed because of low inter-total correlation so as to increase the Cronbach's alphas of the scales.

MA was measured by three mathematics tests administered to grade 7 students at the beginning, half-way and end of school year 2012. The tests were designed to measure numeration, algebra, fractions, geometry and word problems, assessing the same domain. Each test item was scored 1 in case of a correct response or 0 in case of a wrong response.

The raw scores of the tests were analyzed using Item Response Theory (IRT), in order to estimate the proficiency of the students and study the psychometric quality of the items. To link the tests to each other, we combined the data from our group (4,819 students) with those of another group (720 students from another study who answered part of the items of all three tests). The two-parameter logistic IRT model was used to estimate the difficulty and discrimination parameters with computer program BILOG-MG (Zimowski et al., 1996). The study of the item quality resulted in a reduction of the items. Out of 103 items, one item was removed because it did not correlate with the other items, eight items were removed because of too low or too high difficulty estimates and very poor discrimination estimates. The 94 good items were calibrated to create a common IRT metric scale so that the tests were comparable across the three measurement points. No subscales were computed because of the limited number of items for some subscales. The student proficiency estimates (which will be referred to later as IRT scores) at the first time point were scaled to a mean of 50 and a standard deviation of 10 using T-transformation, and this scale was used as the base achievement at the subsequent time points.

2.4. Statistical analyses

We first calculated descriptive statistics which are shown in Tables 3 and 4 in the results section. The relationships between ATM and MA were analyzed using cross-lagged panel models. In order to conduct the cross-lagged longitudinal analyses on ATM and MA, we employed multilevel structural equation modeling (SEM) techniques using *Full Information Maximum Likelihood* with robust standard errors (MLR; Mplus 8.1; Muthen & Muthen, 2017) to estimate the models. The model thus accounted for the simplex-cross lagged nature of stability change among individuals, while accounting for the nesting of individuals within classrooms and schools.

As several indicators for the latent variable of ATM were used in the tested models, we took a two-step approach. In the first step, measurement invariance was studied, using PROC CALIS (SAS 9.4 software, 2013). In a second step, we tested a model to

Table 3. Bivariate correlations, means (M), and standard deviations (SD) of variables used in this study.

Variable	1.	2.	3.	4	5.	6.	7.	8.	9.	10.	11.	12.
1. MA1	1											
2. CONF1	.18	1										
3. USE1	.17	.44	1									
4. ENJOY1	.16	.52	.35	1								
5. MA2	.41	.28	.26	.23	1							
6. CONF2	.16	.23	.18	.19	.22	1						
7. USE2	.10	.15	.17	.12	.16	.40	1					
8. ENJOY2	.13	.20	.14	.20	.19	.54	.44	1				
9. MA3	.37	.24	.25	.20	.63	.19	.16	.18	1			
10. CONF3	.17	.23	.16	.21	.27	.22	.14	.18	.26	1		
11. USE3	.10	.13	.18	.14	.22	.16	.20	.15	.21	.43	1	
12. ENJOY3	.15	.19	.12	.20	.23	.21	.16	.21	.23	.54	.50	1
Mean	50.00	67.80	75.19	68.49	50.86	62.41	80.74	65.75	51.21	60.76	76.46	62.97
SD	10.00	18.94	18.14	19.88	15.05	17.73	19.41	19.11	16.24	17.15	18.58	19.66
N	4768	4768	4768	4768	4351	4351	4351	4351	4244	4244	4244	4244

Notes: All correlations are significant at $p < .001$. MA1, MA2, MA3: Math achievement at Time 1, 2, & 3, respectively; CONF1, CONF2, CONF3: Math self-confidence at Time 1, 2, & 3, respectively; USE1, USE2, USE3: Usefulness of math at Time 1, 2, & 3, respectively; ENJOY1, ENJOY2, ENJOY3: Enjoyment of math at Time 1, 2, & 3, respectively.

Table 4. Means and standard deviations of mathematics achievement and attitude toward mathematics by sex across the three measurement occasions.

	1st measurement point			2nd measurement point			3rd measurement point		
	N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Mathematics Achievement									
Boys	2146	50.95	10.13	1956	52.61	15.70	1889	53.54	16.46
Girls	2622	49.22	9.82	2395	49.43	14.34	2355	49.34	15.81
Total	4768	50.00	10.00	4351	50.86	15.05	4244	51.21	16.24
Attitude toward Mathematics									
Boys	2146	78.09	13.12	1956	65.53	17.19	1889	63.79	16.57
Girls	2622	78.32	13.64	2395	66.93	16.28	2355	63.49	16.18
Total	4768	78.22	13.41	4351	66.30	16.71	4244	63.62	16.35

Note: All mathematics achievement values are based on the IRT transformed data.

study the structural relationship between the observed and unobserved variables. The measurement models for the latent construct for ATM were tested for model fit to establish measurement invariance prior to evaluating cross-lagged models in boys and girls (Vandenberg & Lance, 2000). We therefore tested four nested measurement invariance models over time, organized in an hierarchical ordering with decreasing numbers of parameters (and increasing degrees of freedom) along with various constraints. A model was tested with no constraints, that is, all parameters were allowed to be different (Model 1a). The measurement errors of each indicator were allowed to covary with the corresponding error terms at other time points (Kline, 2016). Also, a model with equal error variances but different factor loadings (Model 1b) was tested. Another model that was tested assumed equal factor loadings but different error variances (Model 1c). A last model that was tested included equal factor loadings and equal error variances (Model 1d).

By means of a series of cross-lagged SEMs (see Figure 1), we tested several competing structural models to determine the best-fitting model for the temporal relationship between ATM and MA among students. In view of the practical interpretability of cross-lagged panel models (Berry & Willoughby, 2017), we added a trait-level ATM and MA factors with equated loadings across time, especially, to disaggregate between-person variance from time-related within-person change in rank ordering. A model without cross-lagged structural paths (ab1 = ab2 = ba1 = ba2 = 0) but with stability paths with autoregressive effects over time, (i.e. sa1, sa2 and sb1, sb2paths estimated) was specified (Model 2). This model was compared to more complex models incorporating cross-lagged paths. Next, we tested a causal model which added to the baseline model cross-lagged structural paths from MA (Time 1 & 2) to ATM (Time 2 & 3), respectively (Model 3; ba1 and ba2 path estimated). Subsequently, we tested a causal model which added to the baseline model cross-lagged paths from ATM to MA (ab1 and ab2, Model 4). Lastly, we estimated a reciprocal-causation model which added cross-lagged paths to the baseline model both from MA to ATM, and from ATM to MA (ab1, ab2, ba1, and ba2; Model 5). Our primary interest was to model the individual level structure while accounting for the variance and covariance structures of the classroom and school levels. The variance and covariances at the classroom and school levels were modelled as autoregressive associations across MA and ATM (see Figure 1(b,c) respectively). Although of interest, we were unable to directly

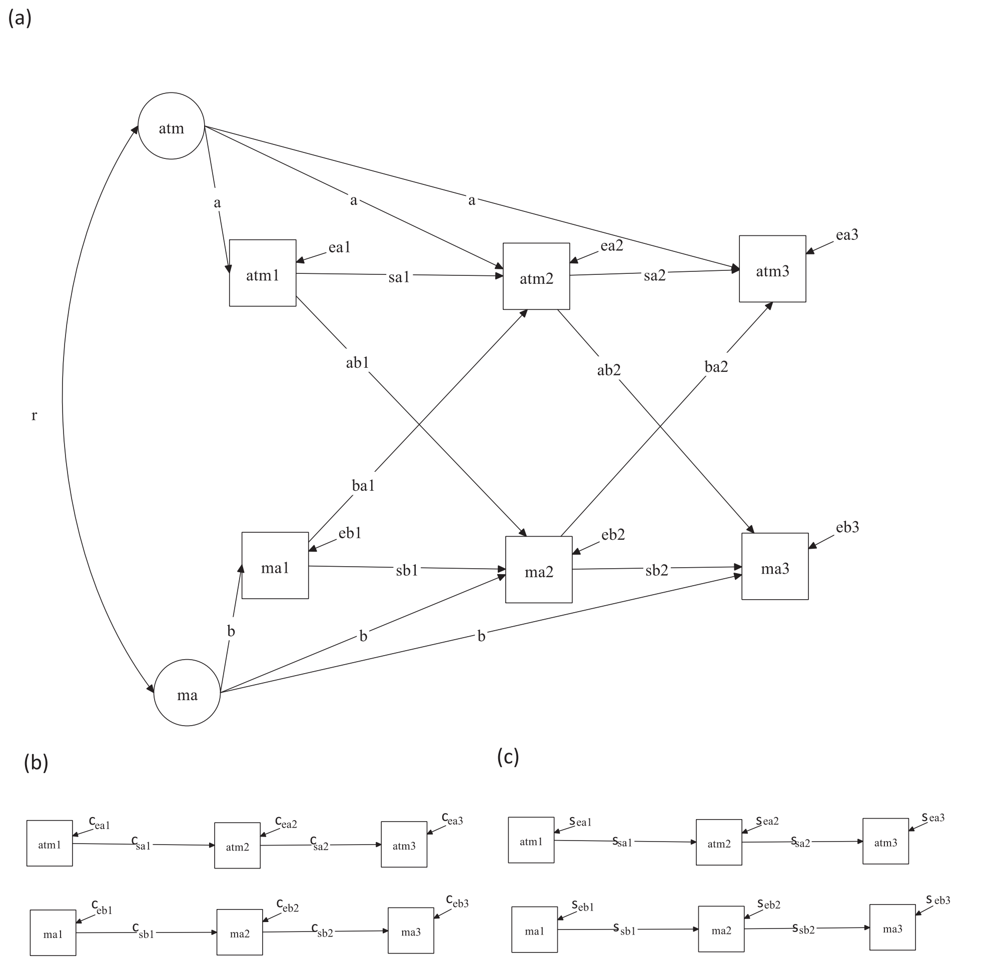

Figure 1. Cross-lagged simplex model of attitude toward mathematics and mathematics achievement across time: (a) individual level, (b) class level, (c) school level.

model the cross-lagged associations at the classroom and school levels given the limited number of units at these levels.

Several goodness of fit indices have been proposed to assess the structural equation model. The following relative fit criteria were used: (a) difference chi-square; (b) goodness-of-fit index (GFI) > 0.90 (Hooper et al., 2008); (c) comparative fit index (CFI) ≥ 0.95 (Hu & Bentler, 1999); (d) Root Mean Square Error of Approximation (RMSEA) < 0.07 (Steiger, 2007); and (e) Standardized Root Mean Square Residual (SRMR) < 0.08 (Hu & Bentler, 1999).

In PROC CALIS, different nested models are compared by a chi-square difference test (Jöreskog & Sörbom, 1996). The difference between two nested models has itself a chi-square distribution with a number of degrees of freedom equal to the corresponding difference between the degrees of freedom of the separate models. In Mplus, where

MLR was employed, chi-square difference tests were computed using the obtained log-likelihood values, scaling correlation factors and differences in the number of parameters (www.statmodel.com).

In the primary analyses evaluating simplex cross-lagged associations, we used Full Information Maximum Likelihood with robust standard errors (MLR; Mplus 8.1; Muthen & Muthen, 2017). We fitted multilevel models for the total sample, adjusting the ATM1, ATM2, ATM3, and MA1, MA2 and MA3 scores for boys and girls respectively, to estimate the correlations of ATM and MA on the student, classroom and school level to account for the hierarchical variance-covariance structure.

3. Results

Relations among the variables considered are presented in Table 3 and are the basis of subsequent analyses. Table 3 which gives also some other descriptive statistics shows that all indicators were positively and significantly correlated with each other.

Table 4 shows the descriptive statistics for MA and ATM of boys and girls. On average, the mean values for the outcome measures of MA and ATM were fairly stable over the three measurement points, although there was a small increase in the means of achievement for boys. The boys performed better than the girls in mathematics. Overall, there was a small growth in achievement for the students. It is important to note that the change was so small that it is probably not worth mentioning. There was a decrease in the means of ATM among both boys and girls, especially from Time 1 to Time 2, indicating a common decline in attitude for both sexes. By looking at the subscales (results not shown in the tables), we found that the girls had higher means of mathematics self-confidence and usefulness of mathematics at Time 1 and 2. And, the boys had a slightly higher mean in enjoyment of mathematics at Time 3. In general, the standard deviations for both ATM and MA increase over time. That suggests that the students are increasingly heterogeneous both in attitudes and achievement.

The results of the analysis showed the Cronbach's alpha, α, as follows: mathematics self-confidence (CONF; $\alpha_{T1} = .71$, $\alpha_{T2} = .69$, $\alpha_{T3} = .70$), perceived usefulness of math (USE; $\alpha_{T1} = .61$, $\alpha_{T2} = .67$, $\alpha_{T3} = .70$) and enjoyment of math (ENJOY; $\alpha_{T1} = .56$, $\alpha_{T2} = .60$, $\alpha_{T3} = .60$). A confirmatory factor analysis (CFA) was used to establish the reliability of the scales. The model fit was evaluated by the fit indices: χ^2 =188.19, RMSEA = .051, CFI = .982, SRMR = .044. These results show a good model fit. The component loadings were used as weights to combine the items to form one indicator.

The first test (MA1; $\alpha = .58$) consisted of 23 items in total of which 10 were multiple-choice questions and 13 items were (derived from 8) open questions. The 2nd test (MA2; $\alpha = .84$) and 3rd test (MA3; $\alpha = .83$) consisted of 40 items, which were multiple choice questions. The 2nd and 3rd tests included TIMSS mathematics items which were made publicly available.

3.1. Missing data

Missing data are inevitable in longitudinal studies such as the present study. We observe that the percentage of missing data for the attitudinal indicators was higher at the second

Table 5. Percentage of missing values in the sample of 4819 students.

	Time 1	Time 2	Time 3
Participants	1.1%	9.7%	11.9%
Math achievement	8.7%	6.1%	4.0%
Math self-confidence	5.3%	8.7%	7.5%
Usefulness of Math	4.5%	7.5%	6.6%
Enjoyment of Math	3.8%	7.3%	6.5%

measurement point, that of achievement kept on decreasing over time, whereas that of the participants made a drastic rise at the second and third-time points as shown in Table 5.

In our study, we found that missingness was at random. So, we used a Full Information Maximum Likelihood estimation method, which yields unbiased estimates under the assumption that data are missing at random (Enders & Bandalos, 2001).

Prior to the model analyses, correlations and variance components (at student, class and school) in ATM and MA are given in Table 6 (for boys and girls together) and Table 7(a) (for boys only) and Table 7(b) (for girls only) as estimated in Mplus 8.1 (Muthen & Muthen, 2017)

For both sex groups, the largest part of the variance (in MA and ATM) was situated at the student level. The proportions of the variances at the class level and at school level were larger for the achievement measures than for the attitude measures. We found statistically significant positive correlations among the constructs for both boys and girls. These correlations were generally larger at the class and the school level than at the student level. We also observe that the correlation coefficients between the three measurement occasions of ATM were higher for the boys than for the girls at both class and school level, and were almost the same at the student level. However, these constructs (ATM and MA) are acceptable as theoretically and empirically distinct although they are highly correlated at both class and school level (Hilton et al., 2004).

3.2. Measurement invariance results

It is worth noting that to test the measurement invariance, we made use of nested models with increasing degrees of freedom, as suggested by Vandenberg and Lance (2000). All measurement models were specified for all latent constructs at the same time, i.e. confidence, perceived usefulness and enjoyment. The purpose of the measurement model was to test for a model fit and to examine the relationship between measured variables (indicators) and latent constructs (Kline, 2016).

Table 8 gives the comparison of models. The first model tested with no constraints (Model 1a) indicated a significant relationship between the variables ($p < .001$). The constraint of having equal error variances (Model 1b) showed an improvement in the model fit ($\Delta\chi^2 = 35.78$, df $= 20$, $p = .016$) as compared to Model 1a. The model with equal factor loadings (Model 1c) fitted the data better than Model 1a ($\Delta\chi^2 = 75.15$, df $= 20$, $p < .001$). Lastly the model with equal factor loadings and equal variances for boys and girls fitted the data better than Model 1d and Model 1b. Because the$\Delta\chi^2$ value was not significant ($\Delta\chi^2 = 4.9$, df $= 9$, $p = .84$) the constrained model (Model 1d) did not differ statistically from Model 1c. This finding showed that the measurement of the latent

Table 6. Correlations and variance of ATM and MA at student, class and school level for boys and girls from the multivariate multilevel model.

	Correlations Student level						Correlations Class level						Correlations School level						% variance			Raw variances		
Variables	1.	2.	3.	4.	5.	6.	1.	2.	3.	4.	5.	6.	1.	2.	3.	4.	5.	6.	Student	Class	School	Student	Class	School
1 MA1	1						1						1						87.2	3.9	8.9	85.31	3.82	8.75
4.ATM1	.10	1					.52	1					.10	1					89.4	4.7	5.9	161.07	8.55	10.55
3.MA2	.27	.19	1				.99	.61	1				.27	.19	1				66.7	13.8	19.4	144.79	30.04	42.12
5.ATM2	.11	.21	.20	1			.77	.46	.73	1			.11	.21	.20	1			94.7	2.2	3.1	266.14	6.16	8.73
3.MA3	.24	.17	.49	.17	1		.96	.50	.94	.86	1		.24	.17	.49	.17	1		69.5	13.6	16.8	174.36	34.12	42.24
6.ATM3	.11	.19	.20	.27	.18	1	.65	.51	.68	.57	.58	1	.11	.19	.20	.26	.18	1	87.2	4.2	8.6	231.57	11.26	22.83

Note: T1: Time point 1; T2: Time point 2; T3: Time point 3; MA: math achievement; ATM: attitude toward math.

Table 7. Correlations and variance of ATM and MA at student, class and school level for boys from the multivariate multilevel model.

	Correlations																		% variance			Raw variances		
Variables	Student level						Class level						School level											
	1.	2.	3.	4.	5.	6.	1.	2.	3.	4.	5.	6.	1.	2.	3.	4.	5.	6.	Student	Class	School	Student	Class	School
(a)																								
1 MA1	1						1						1						83.5	4.1	12.4	84.01	4.15	12.43
4.ATM1	.11	1					.63	1					.74	1					89.4	3.6	6.9	153.75	6.27	11.95
3.MA2	.27	.17	1				.96	.78	1				.91	.74	1				66.3	13.8	19.9	158.55	32.93	47.62
5.ATM2	.11	.22	.19	1			.77	.85	.77	1			.82	.70	.70	1			93.2	3.5	3.3	275.84	10.39	9.66
3.MA3	.24	.16	.47	.16	1		.98	.59	.90	.81	1		.89	.83	.92	.59	1		71.2	15.0	13.8	186.90	39.3	36.23
6.ATM3	.12	.17	.21	.27	.19	1	.60	.91	.67	.95	.63	1	.75	.48	.72	.52	.72	1	89.2	5.4	5.4	244.06	14.80	14.77
(b)																								
MA1	1						1						1						88.6	4.7	6.7	85.02	4.46	6.47
4.ATM1	.09	1					.46	1					.31	1					87.6	7.0	5.4	163.28	12.96	10.11
3.MA2	.26	.21	1				.98	.46	1				.73	.13	1				65.9	13.7	20.4	131.48	27.28	40.77
5.ATM2	.12	.21	.22	1			.74	.11	.71	1			.59	.58	.40	1			96.1	1.6	2.3	255.24	4.34	6.05
3.MA3	.22	.17	.50	.19	1		.89	.35	.92	.85	1		.75	.39	.75	.47	1		67.1	12.8	20.1	159.82	30.48	47.87
6.ATM3	.10	.21	.21	.24	.18	1	.49	.14	.57	.17	.37	1	.43	.39	.49	.57	.73	1	86.2	5.5	8.4	217.27	13.78	21.06

Table 8. Model comparisons.

	χ^2	df	GFI	CFI	SRMR	RMSEA	$\Delta\chi^2$	Δdf	Models
1. Model 1a	188.19	26	0.990	0.982	0.044	0.051			
2. Model 1b	223.97	46	0.988	0.981	0.045	0.040	35.78*	20	1a vs 1b
3. Model 1c	263.34	46	0.986	0.976	0.051	0.044	75.15**	20	1a vs 1c
4. Model 1d	268.24	55	0.986	0.977	0.051	0.040	44.27**	9	1b vs 1d
							4.9	9	1c vs 1d

Significance level: *$p < 0.05$ and **$p < 0.001$.
Model 1a = no constraints
Model 1b = constraints with equal error variances
Model 1c = equal factor loadings
Model 1d = equal factor loadings and equal variances.

attitude construct in the model was highly similar for both sexes, and this indicated the appropriateness of using this latent construct in the structural models for both boys and girls (Schermelleh-Engel et al., 2003).

3.3. Reciprocal effects models

In a single-group analysis for both sexes, the fit of each of these proposed nested structural models (no cross effects) was compared with that of the full cross-effects model. A model that described the data almost equally good but with fewer parameters estimated is considered a better model for reasons of parsimony (Hu & Bentler, 1999). The results are summarized in Table 9.

In addition to the no cross-effects model (Model 1), we estimated three subsequent models (Table 10). The no cross-effects model (Model 1) fits significantly worse than the full reciprocal effects model (Model 4) ($\Delta\chi^2 = 15.77$, $\Delta df = 4$, $p = .003$). The TLI index was below threshold while all other relative fit indices exceeded the criteria (Table 10). The achievement predominant model (Model 2), with paths connecting prior MA with subsequent ATM, was estimated and likewise the $\Delta\chi^2$ value was significant ($\Delta\chi^2 = 12.04$, $\Delta df = 2$, $p = .002$), suggesting additional paths may be needed to achieve a better fit. Again the TLI index was below threshold while all other fit indices exceeded the criteria (Table 10). The attitude predominant model (Model 3), with paths connecting prior ATM with subsequent MA, was estimated. This model showed a good fit to the data apart from TLI, however, the $\Delta\chi^2$ value was significant ($\Delta\chi^2 = 6.33$, $\Delta df = 2$, $p = .042$), suggesting that additional paths may be required. Comparing the fit criteria of Models 2 and 3,we observe that the attitude predominant model fits better to the data than the achievement predominant model.

Lastly, the reciprocal-effects model (Model 4), with paths connecting prior MA with subsequent ATM as well as prior ATM with subsequent MA, was estimated. This model showed the best relative fit to the data (Table 10). The trait-level ATM and MA factors are strongly correlated (r = .73) and constitute a systematic set of correlated influences across time. For the cross-effects, ATM1 had the highest and statistically significant direct effect on MA2, followed by MA2 on ATM3, suggesting some dynamical influences apart from trait-level correlations. The SMR fit criteria for level 2 and 3 (classroom and school), which constitute a relatively small amount of variance and covariance, suggest additional model specifications might be useful to explore (SRMR $>$ .34); however, the number of

Table 9. Model fit indices from the one-group estimation, adjusting for sex.

Full sample	Full / Comparison	L1 MLR correct	c1parms	p1 Equated	L0 MLR correct	c0 parms	pO diff scalling	cd chi-sq diff	TRddf	*p*	
Model 4 vs Saturated Model	−105,233.74	2.38	75	−105,296.56	3.04	49	1.12	112.27	26	0.000 full model	full cross-lag
Model 3 vs Model 4	−105,296.56	3.04	49	−105,303.68	3.08	47	2.25	6.33	2	0.042 ba = 0	drop cross-lag MA > ATM
Model 2 vs Model 4	−105,296.56	3.04	49	−105,308.29	3.09	47	1.95	12.04	2	0.002 ab = 0	drop cross-lag ATM > MA
Model 1 vs Model 4	−105,296.56	3.04	49	−105,312.98	3.13	45	2.08	15.77	4	0.003 ba = ab = 0	no cross-lag

Table 10. Model fit indices from the one- group estimation.

Model	TRd$\Delta\chi^2$ (vs Model 4)	Δdf	*p*	AIC	BIC	N-adjusted BIC	RMSEA	CFI	TLI	SRMR	Model
1.No cross-effects	15.77	4	0.003	210,715.95	211,007.57	210,864.57	0.026	0.928	0.878	0.027	1
2.Achievement predominant (MA- > ATM)	12.04	2	0.002	210,710.58	211,015.16	210,865.81	0.027	0.927	0.867	0.026	2
3.Attitude predominant (ATM- > MA)	6.33	2	0.042	210,701.36	211,005.93	210,856.59	0.026	0.934	0.881	0.024	3
4.Reciprocal-effects (ATM- > MA, MA- > ATM)	-	-	-	210,691.12	211,008.66	210,852.98	0.026	0.936	0.875	0.020	4

Note: CFI: Bentler Comparative Fit Index; TLI: Tucker Lewis Index; RMSEA: Root Mean Square Error of Approximation; TRdχ^2: adjusted difference in chi-square between reciprocal-effects models and model of interest; Δdf: difference in degrees of freedom between reciprocal-effects model and model of interest.

parameters estimated under the current model is kept limited, particularly at level 3, due to the small number of units available as indicated earlier.

3.4. Sex differences

To explore sex differences in structural paths, we fitted separate models for boys and girls. The correlations between mathematics achievement and attitudes at the student level are similar between boys and girls, but are lower for girls at the class and school levels, and the stability of attitudes is weaker for girls (see Table 7(a) versus (b)). For both boys and girls, the reciprocal effects model showed the best, or comparable, relative fits compared to the other proposed models (see Tables 11 and 12). However, the difference chi-square tests suggest that for girls, the attitude predominant model (Model 3), with paths connecting prior ATM with subsequent MA, did not differ from the full reciprocal model ($\Delta\chi^2 = 3.80$, $\Delta df = 2, p = .149$), suggesting that the inclusion of the cross-paths from MA to ATM are not required. The difference chi-square tests suggest that for boys, the achievement predominant model (Model 2), with paths connecting prior MA with subsequent ATM, did not differ statistically from the full reciprocal model ($\Delta\chi^2 = 5.87$, $\Delta df = 2$, $p = .053$), suggesting that the inclusion of the cross-paths from ATM to MA are not required.

In the full reciprocal effects model, the differences between boys and girls were negligible, despite that differing predominance models (Model 2 or 3) could be supported for each group as described above. The path coefficients are shown in Figure 2. Although the fit indices favoured different models, the full model parameters suggest similarity for boys and girls overall. At the student level, the results of the reciprocal model are similar between boys and girls. The trait-level ATM and MA factors are strongly correlated ($r > .69$) and constitute a systematic set of correlated influences across time. For both sexes, the effect of ATM2 on MA3 was the weakest. The direct effect of ATM2 on ATM3 is stronger than that of ATM1 on ATM2. The direct effect of MA1 on MA2 is weaker than that of MA2 on MA3. For the reciprocal associations, the effect of ATM1 on MA2 is stronger than that of ATM2 on MA3. The effect of MA1 on ATM2 is weaker than that of MA2 on ATM3. The results are similar for mathematics achievement stability at the class and school levels. However, for both sexes, the results are similar for only stability of attitudes at the school level and different at the class level.

4. Discussion

In this study, we used a three-wave panel design and SEMs to examine longitudinal directionality and temporal relatedness between attitude toward math and math achievement for boys and girls. The main finding was that the reciprocal-effects model was the best-fitting model for both sexes. This is consistent with the study of Mattern and Schau (2002) who reported the reciprocal-effects model as best-fitting model for 7th- and 8th-grade boys and girls. They also reported that this model was invariant between the sexes. The fact that both the effect of prior attitude on later attitude and that of prior achievement on later achievement were significant across the three points show that previous mathematics attitude and achievement are strong and important predictors of subsequent attitudes and attainments in mathematics. More importantly, almost all cross-effects of prior attitude on later achievement and prior achievement on later attitude were significant. In particular,

Table 11. Model comparisons from separate analyses of boys and girls.

Full sample	**L1** Full/ Comparison	**c1** MLR correct	**p1** parms	**LO** Equated	**c0** MLR correct	**p0** parms	**cd** diff scalling	**TRd** chi-sq diff	df	*p*	
Boys											
Model 4 vs Saturated Model	−47,426.24	1.66	69	−47,468.49	1.88	43	1.30	65.14	26	0.000 full model	full cross-lag
Model 3 vs Model 4	−47,468.49	1.88	43	−47,472.98	1.92	41	1.13	7.98	2	0.019 ba = 0	drop cross-lag MA > ATM
Model 2 vs Model 4	−47,468.49	1.88	43	−47,472.19	1.91	41	1.26	5.87	2	0.053 ab = 0	drop cross-lag ATM > MA
Model 1 vs Model 4	−47,468.49	1.88	43	−47,475.84	1.95	39	1.18	12.46	4	0.014 ba = ab = 0	no cross-lag
Girls											
Model 4 vs Saturated Model	−57,835.69	2.31	69	−57,878.55	2.77	43	1.55	55.33	26	0.000 full model	full cross-lag
Model 3 vs Model 4	−57,878.55	2.77	43	−57,883.27	2.79	41	2.48	3.80	2	0.149 ba = 0	drop cross-lag MA > ATM
Model 2 vs Model 4	−57,878.55	2.77	43	−57,890.05	2.84	41	1.54	14.97	2	0.001 ab = 0	drop cross-lag ATM > MA
Model 1 vs Model 4	−57,878.55	2.77	43	−57,892.41	2.85	39	2.04	13.56	4	0.009 ba = ab = 0	no cross-lag

Table 12. Model fit indices from separate analyses of boys and girls.

Model	TRd$\Delta\chi^2$ (vs Model 4)	Δdf (vs Model 4)	*p*	AIC	BIC	N-adjusted BIC	RMSEA	CFI	TLI	SRMR (Level 1)	Model
Boys											
1.No cross-effects	12.46	4	0.014	95,029.67	95,251.25	95,127.34	0.027	0.939	0.909	0.030	1
2.Achievement predominant (MA- > ATM)	5.87	2	0.053	95,026.38	95,259.33	95,129.07	0.027	0.945	0.912	0.026	2
3.Attitude predominant (ATM- > MA)	7.98	2	0.019	95,027.90	95,260.90	95,130.63	0.027	0.943	0.908	0.028	3
4.Reciprocal-effects (ATM- > MA, MA- > ATM)	-	-	-	95,022.98	95,267.28	95,130.67	0.026	0.950	0.914	0.021	4
Girls											
1.No cross-effects	13.56	4	0.009	115,862.81	116,092.24	115,968.32	0.022	0.947	0.921	0.035	1
2.Achievement predominant (MA- > ATM)	14.97	2	0.001	115,862.10	116,103.29	115,973.02	0.024	0.945	0.911	0.034	2
3.Attitude predominant (ATM- > MA)	3.80	2	0.149	115,848.54	116,089.73	115,959.46	0.020	0.960	0.935	0.030	3
4.Reciprocal-effects (ATM- > MA, MA- > ATM)	-	-	-	115,843.09	116,096.05	115,959.43	0.021	0.962	0.934	0.026	4

Note: CFI = Bentler Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; TRdχ^2 = adjusted difference in chi-square between reciprocal-effects models and model of interest; Δdf = difference in degrees of freedom between reciprocal-effects model and model of interest.

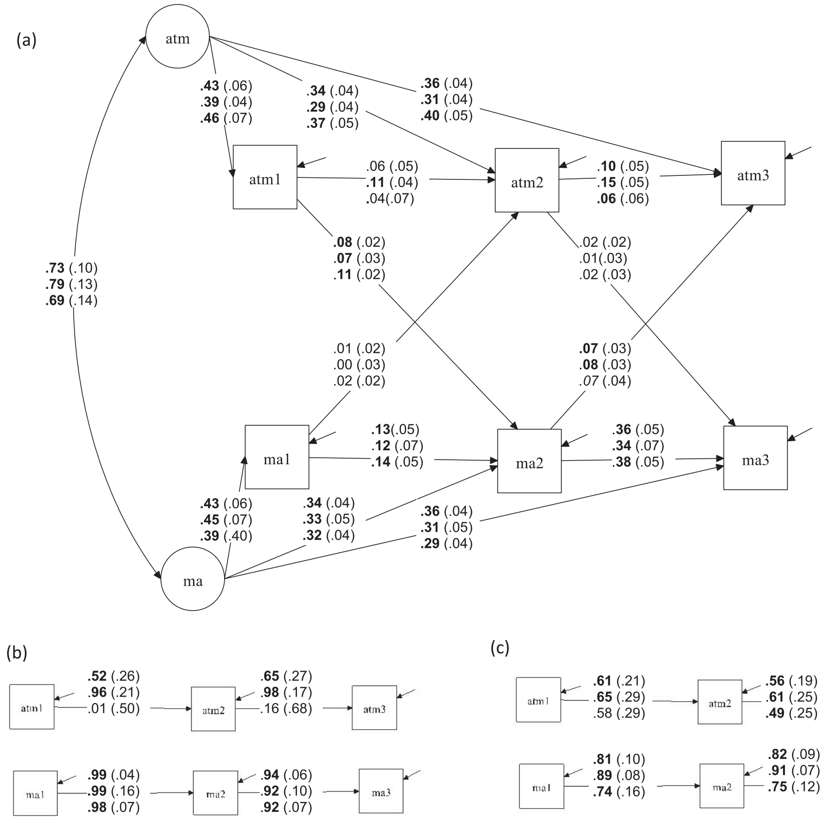

Figure 2. The reciprocal effects model standardized solution for the total sample(top), for boys (middle) and for girls (bottom), at the: (a) individual level, (b) class level, (c) school level. Note: The model fitted to the total sample adjusted for sex on the observed outcomes (MA1, MA2, MA3, ATM1, ATM2, ATM3), where ma = general mathematics achievement; ma1, ma2, ma3 = math achievement at Time 1, 2 & 3, respectively; atm = general attitude toward mathematics; atm1, atm2, atm3 = attitude toward math at Time 1, 2 & 3, respectively.

the results revealed that between the first and second measurement points, attitude had a statistically significant direct effect on achievement which was larger than the effect of achievement on attitude, which supported the attitude predominant model or the self-enhancement model. This is consistent with the findings of Manoah et al. (2011) who reported that attitude for boys and girls has a significant effect on performance, that is, the more positive attitudes students had, the more likely they were to achieve higher in mathematics. The observation that attitude at Time 1 had a positive direct effect on achievement at Time 3 indicates that students' attitude at the beginning of their secondary education predicted their mathematics performance during the whole school year.

Later on, between the second and third measurement points, achievement had a higher positive effect on attitude than attitude had on achievement, which supported the achievement predominant model or the skill-development model. That is, the students' mid-year attitudes have only a tiny effect on their achievement at the end of the year. This suggests that, when students have acclimatized themselves in school, in the middle of the year, achievement becomes a more important determinant of their attitude. At any rate, the effect of attitude on achievement was very low though significant. An unexpected result was that students' mathematics performance at the beginning of their secondary schooling did not significantly influence their end of the year attitude. That is, the effect of MA1 on ATM 3 is diminished by the intervening effect of mid-year performance which was controlled for. The ATM at the end of the year depends largely on the mid-year achievement, which is affected by the achievement at the beginning of the year. A part of the explanation is the low bivariate correlation between MA1 and ATM3, which is .11 and .09 for boys and girls, respectively, at the student level.

4.1. The best-fitting model for boys and girls

We found no evidence of sex differences in the causal relationship between ATM and MA. This finding re-echoes the results of prior research (Ma & Xu, 2004). Both boys and girls showed similar reciprocal effects and stability effects between attitude and achievement, and the reciprocal effects model was the best-fitting structural model for both sexes together and separately. This is in line with the finding of Mattern and Schau (2002). This result indicates that both boys and girls who initially performed well in mathematics continued to do well in mathematics and reported positive ATM. Similarly, both boys and girls who initially exhibit positive ATM subsequently achieve higher MA.

We found significant paths from attitude to achievement and vice versa. However, there was no meaningful difference in the structural paths between boys and girls. Inconsistent with the findings of Mattern and Schau (2002), the boys' and girls' initial attitude had higher influence on the mid-year attitude, which in turn had less impact on the attitude at the end of the school year. However, the boys' and girls' initial achievement had less influence on the mid-year achievement, which in turn had higher impact on the performance at the end of the school year.

4.2. Strengths, limitations and further research

This research posed some strengths that lend credibility to the results, which are in line with the suggestion by Byrne (1986): the use of a longitudinal study (three measurement points) and of structural equation modeling. In particular, this research underscored the use of the cross-lagged longitudinal model, i.e. the random-intercept cross-lagged panel model. It is important to note that this model allowed us to separate the within-person processes from the between-person differences from a multilevel perspective. Additionally, it enabled us to estimate the autoregressive and cross-lagged effects at three levels. More so, the study has some relevant theoretical and educational implications.

However, the study had several limitations. First, some of the cross-lagged effects found in this study were relatively weak. For example, the effect of attitude at Time 2 to achievement at Time 3 was the weakest of the significant effects. One possible cause for this finding could be the relatively short periods between the measurement points. Therefore, longer time lags with possibly more measurement occasions would be valuable in future research.

Second, another limitation is the external validity threat. External validity refers to the extent to which the results of the study can be generalized to other populations, conditions, experimenters and so forth (Gracetter & Forzano, 2011). The findings of this study are based on a single sample of 7th-graders (first-year secondary school students) from Central Uganda. They might have different attitudes than students in other grade levels, or in other parts of Uganda, or in other countries. This may raise problems in generalizing to other regions and countries. Future research is needed to replicate this study. It would also be informative to explore the models with students of different grade levels, because the causal relations may vary with the students' age/grade.

The third limitation is the questionable reliability of the scales since the three attitudinal indicators had low Cronbach's alpha coefficients. The alpha values of self-confidence scale were just on the cut-off point of 0.70 for reliability that was proposed by Majeed et al. (2013), whereas the alpha values of the scales of perceived usefulness and enjoyment of math were below this cut-off point. The items of these scales were retained on the grounds that for research purposes a smaller reliability of instruments can still be regarded as acceptable (Suhr & Shay, 2009, suggested a cut-off of .60).

4.3. Theoretical implications

The finding that stands out in this study is that the temporal relationship between ATM and MA is predominantly bidirectional from achievement to attitude and also from attitude to achievement. The effect sizes between the first and second point, and between the second and third, support the self-enhancement and skill-development models, respectively. This points to the mutual dominance of cognitive and affective factors. This indicates that both attitude and achievement play a role in their reciprocal relationship among students when they enter lower secondary school. Hopefully, this reciprocal relationship between ATM and MA continues throughout lower secondary education until when mathematics is elective in the upper secondary. As expected, students who expressed self-confidence in their ability to learn mathematics also enjoyed/liked mathematics more. And those students, who liked mathematics more, valued it more highly and had a higher achievement. Our results indicate that the temporal relatedness between attitude and achievement is sex invariant. Again, our results do not support the findings of Ma and Xu (2004) that prior attitude did not meaningfully predict later achievement.

4.4. Educational implications

The research results provide some guidelines for educational and counselling interventions. The descriptive statistics showed a common decline in attitude for both sexes. We found that a decrease in achievement seems to be one of the reasons leading to the decline in attitude. Other possible reasons could be ineffective teaching, peer influence, lack of confidence or shyness. However, some studies (e.g. Ma & Kishor, 1997) have indicated that attitudes of students become less positive as they grow up and become negative at high-school age. Nevertheless, if students achieve well in mathematics, their attitudes can be positively enhanced. Indeed SEM results showed that higher achievement had positive effects on attitude. Since attitudes can be changed, one of the strategies of mathematics educators can be to alter negative attitudes and strengthen more positive attitudes in view of improving students' mathematics performance. Although our study did not consider any direct or indirect causal relationship between each of the three attitudinal indicators and mathematics achievement, mathematics teaching should promote the three indicators. Therefore interventions are more likely to be appropriate, where teachers can identify students who may have problems and intervene in time.

Finally, positive effects of ATM and MA indicate that prior attitude and achievement significantly predicted later attitude and achievement of students. Consequently, we recommend that policy makers, school leaders, mathematics teachers, parents and counsellors, and other stakeholders engage in programmes designed to improve both attitude and achievement early in students' lower secondary education. Interventions that focus on developing MA or ATM alone should be avoided, because they might result in an improvement of MA with little effect on ATM or improvement in ATM with little effect on MA. Indeed, improvement in attitude and achievement in first-year of secondary education can lay a good foundation for the attitudinal and academic well-being of students in their schooling (Ma & Xu, 2004). All in all, improved achievement can promote positive ATM and improved ATM can in turn contribute to better performance.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186–1206. https://doi.org/10.1111/cdev.12660

Bhowmik, M., & Banerjee (Roy), B. (2016). A study on relationship between achievement in mathematics and attitude towards mathematics of secondary school students. *JRA International Journal of Education and Multidisciplinary Studies (ISSN 2455-2526)*, *4*(3), 402–408. https://doi.org/10.21013/jems.v4.n3.p7

Byrne, B. M. (1986). Self-concept/academic achievement relations: An investigation of dimensionality, stability, and causality. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement*, *18*(2), 173–186. https://doi.org/10.1037/h0079982

Chiu, M.-S. (2012). Differential psychological processes underlying the skill-development model and self-enhancement model across mathematics and science in 28 countries. *International Journal of Science and Mathematics Education*, *10*(3), 611–642. https://doi.org/10.1007/s10763-011-9309-9

Dowker, A., Cheriton, O., Horton, R., & Mark, W. (2019). Relationships between attitudes and performance in young children's mathematics. *Educational Studies in Mathematics*, *100*(3), 211–230. https://doi.org/10.1007/s10649-019-9880-5

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *The Mathematics Educator*, *17*(2), 24–30.

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning if mathematics by females and males. *Journal for Research in Mathematics Education*, *7*, 324–326.

Gitaari, E. M. E., Nyaga, G., Muthaa, G., & Reche, G. (2013). Factors contributing to students poor performance in mathematics in public secondary schools in Tharaka South district. Kenya. *Journal of Education and Practice*, *4*(7), 93–99.

Gracetter, F. J., & Forzano, L. B. (2011). *Research methods for the behavioral sciences* (4th Ed.). Wadsworth Cengage Learning.

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, *26*(3), 276–295. https://doi.org/101016/j.appdev.2005.02.007

Hilton, S. C., Schau, C., & Olsen, J. A. (2004). Survey of attitudes toward statistics: Factor structure invariance by gender and by administration time. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(1), 92–109. https://doi.org/10.1207/S15328007SEM1101_7

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Scientific Software International.

Khun-Inkeeree, H., Omar-Fauzee, M. S., & Othman, M. K. H. (2016). Students' attitude towards achievement in mathematics: A cross sectional study of year six students in Songklha Province, Thailand. *European Journal of Education Studies*, *2*(4), 88–99.

Kiwanuka, H., Van Damme, J., Van den Noortgate, W., Anumendem, D. N., & Namusisi, S. (2015). Factors affecting mathematics achievement of first-year secondary school student in Central Uganda. *South African Journal of Education*, *35*(3), 1–16. https://doi.org/10.15700/saje.v35n3a1106

Kiwanuka, H., Van Damme, J., Van den Noortgate, W., Anumendem, D. N., Vanlaar, G., Reynolds, C., & Namusisi, S. (2016). How do student and classroom characteristics affect attitude toward mathematics? *A Multivariate Multilevel Analysis. School Effectiveness and School Improvement*, *28*(1), 1–21. https://doi.org/10.1080/09243453.2016.1201123

Kline, R. B. (2016). *Principles and Practice of structural equation modeling, 4th edition*. The Guilford Press.

Kundu, A., & Ghose, A. (2016). The relationship between attitude towards and achievement in mathematics among higher secondary students. *International Journal of Multidisciplinary Research and Development*, *3*(6), 69–74.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, *28*(1), 26–47. https://doi.org/10.2307/749662

Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education*, *110*(3), 256–280. https://doi.org/10.1086/383074

Majeed, A. A., Darmawan, G. N., & Lynch, P. (2013). A confirmatory factor analysis of attitudes toward mathematics inventory (ATMI). *The Mathematics Educator*, *15*(1), 121–135.

Manoah, S. A., Indoshi, F. C., & Othuon, L. O. A. (2011). Influence of attitude on performance of students in mathematics curriculum. *Educational Research*, *2*(3), 965–981.

Marsh, H. M., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, *34*(3), 154–157. https://doi.org/10.1207/s15326985ep3403

Masih, A. M. M., & Masih, R. (1997). On the temporal causal relationship between energy consumption, real income, and prices: Some new evidence from Asian-energy dependent NICs based on a multivariate cointegration/vector error-correlation approach. *Journal of Policy Modeling*, *19*(4), 417–440. https://doi.org/10.1016/S0161-8938(96)00063-4

Mattern, N., & Schau, C. (2002). Gender differences in science attitude-achievement relationships over time among White middle-school students. *Journal of Research in Science Teaching*, *39*(4), 324–340. https://doi.org/10.1002/tea.10024

Mubeen, S., Saeed, S., & Arif, M. H. (2013). Attitude towards mathematics and academic achievement in mathematics among secondary level boys and girls. *IOSR Journal of Humanities and Social Science*, *6*(4), 38–41. https://doi.org/10.9790/0837-0643841

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International results in mathematics*. TIMSS & PIRLS International Study Center, Boston College. http://timssandpirls.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf.

Muthen, L. K., & Muthen, B. O. (2017). *Mplus User's Guide*. 8th Ed. Muthen & Muthen.
Opolot-Okurut, C. (2005). Student attitudes toward mathematics in Ugandan secondary schools. *African Journal of Research in Mathematics, Science and Technology Education*, *9*(2), 167–174. https://doi.org/10.1080/10288457.2005.10740587
Phonguttha, R., Tayraukham, S., & Nuangchalerm, P. (2009). Comparisons of mathematics achievement, attitude toward mathematics and analytical thinking between using the geometer's sketchpad program as media and conventional learning activities. *Australian Journal of Basic and Applied Science*, *3*(3), 3036–3039. https://doi.org/10.2139/22m.1285446
Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural modeling. *Personality and Individual Differences*, *42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017
Suhr, D., & Shay, M. (2009). *Guidelines for reliability, confirmatory and exploratory factor analysis.* https://www.lexjansen.com/wuss/2009/anl/ANL-SuhrShay.pdf
Schermelleh-Engel, K., Moosbrugger, H. & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2): 23–74.
Tapia, M., & Marsh IIG. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, *8*(2), 130–143.
Van Damme, J., Opdenakker, M.-C., & Van den Broeck, A. (2004). *Do classes and schools have an effect on attitudes towards mathematics?*http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2004/Papers/IRC2004_VanDamme_Opdenakker_etal.pdf.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002
Williams, T., Williams, K., Kastberg, D., & Jocelyn, L. (2005). Achievement and affect in OECD nations. *Oxford Review of Education, 31*(4), 517–545.
Zan, R., & Di Martino, P. (2007). Attitudes toward mathematics: Overcoming positive/negative dichotomy. *The Montana Mathematics Enthusiasts Monograph*, *3*, 157–168.
Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG: Multiple group IRT analysis and test maintenance for binary items*. Scientific Software International, Inc.

Appendix

Given are two representative items of each scale used for attitude toward math:
Scale: CONF = mathematics self-confidence
How much do you agree with the statement describing your confidence in math?

(1) I have a lot of self-confidence when it comes to math.
(2) Math does not scare me at all.

Scale: USE = perceived usefulness of mathematics
How do you agree with the following statements describing the usefulness of math?

(1) I study math because I know how useful it is.
(2) Math will not be important to me in my life's work. (-)

Scale: ENJOY = Enjoyment of mathematics
How much do you agree with the following statements describing your enjoyment of math?

(1) Math is enjoyable and stimulating to me.
(2) I find it hard to solve mathematical problems. (-)