

RESEARCH

Open Access



TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties

Emmanuel Adu-tutu Bofah* and Markku S. Hannula

*Correspondence:
emmanuel.bofah@hotmail.
com
Department of Teacher
Education, University
of Helsinki, PO Box 9,
Siltavuorenpenger 5,
00014 Helsinki, Finland

Abstract

Relationships among motivational constructs from the 2011 Trends in International Mathematics and Science Study (TIMSS 2011) were investigated for eight-graders in all the five participating African countries, representing 38,806 (49 % girls). First, we investigated the psychometric properties (factor structure, reliabilities, method effect, and measurement invariance—country and gender) of the mathematics motivational constructs across the five educational systems. There was empirical support for the multidimensionality of the construct and the TIMSS 2011 motivational construct was largely invariant across cultures. Furthermore, a series of confirmatory factor analyses revealed that there is a need to control method effects associated with negatively worded items in the measurement model. There was support suggesting that in many cultures responses to negatively worded items are systematically different. The factor structures and reliabilities (i.e., confidence and the like mathematics scales) were affected by negatively worded items. Second, the relationships between the constructs, achievements and background variables such as parental education, gender and students' educational aspirations were investigated. We identified several significant relationships between self-belief and mathematics achievement. Differences in the latent mean achievement and the motivational construct were similar to those that have been described in the literature as “paradoxical” and “perplexing”. Nations with high mathematics achievement seem to have students with more negative mathematics self-belief. Some results extend, whereas others refute the findings of previous research. For instance, the relationship between students' mathematics confidence and mathematics achievement was lower than the relationship between the value of mathematics and achievement in some countries and it was the reverse in others. However, consistent with cultural stereotypes, boys rated their mathematics competence higher than girls. The findings are discussed with reference to implications for cross-cultural research and practice.

Keywords: Africa, Mathematics and gender difference, Liking mathematics, Mathematics confidence, Measurement invariance, Negative worded method effect, Parental involvement, Teacher responsiveness, Trends in International Mathematics and Science Study (TIMSS), Value for mathematics

Background

The performance of African countries participating in international benchmark tests, (e.g., TIMSS) is a major concern amongst educators and policymakers. This lukewarm performance raises questions about the effectiveness of the periodical curriculum and educational reforms in most of these countries (Ndlovu and Mji 2012). Ghana and South Africa, two African countries, participating in TIMSS 2003 to TIMSS 2011, have shown a significant improvement within that period. On the average Ghana and South Africa improved by 51 and 65 scaled score points respectively for mathematics (Reddy et al. 2014). However, the performance of students from the five African countries participating in TIMSS 2011 was amongst the very lowest. All the African countries participating in TIMSS 2011 recorded an average mathematics achievement below the TIMSS center point of 500, with all but Tunisia falling below the Low (400) International Benchmark. Moreover, for students in Ghana, Morocco, South Africa and Botswana the percentage of students with achievements too low for estimation exceeded 25 % (Mullis et al. 2012).

Researchers rarely cross-culturally compare the TIMSS motivational constructs (including the psychometric properties) and the achievement levels of African nations participating in TIMSS in a single study. Most cross-cultural studies involving TIMSS motivational constructs and achievement have been predominately between countries such as the United States, European countries, Japan, China and other East Asian Countries (e.g., Zhu and Leung 2011; Shen and Tam 2008; Liu and Meng 2010). The outcomes of these studies have produced their own controversies. The findings from these studies often have been described as *paradoxical* (Shen and Tam 2008) or *perplexing* (Marsh et al. 2013), in the sense that for instance self-concept has a negative relationship with achievement at the country level and a positive relation at the individual level. Moreover, Shen and Tam (2008) found that the top mathematics performing nations such as Chinese Taipei, Korea, Japan, Hong Kong, and the Netherlands comparatively reported a lower levels of affect (e.g., enjoyment for learning mathematics), whereas low-performing countries such as South Africa, Ghana, Botswana, Morocco, and Egypt generally report a higher levels of affect (e.g., enjoyment for learning mathematics) (see also Liu and Meng 2010).

The cultural backgrounds, school resources, and teaching conditions of students in the African countries that participated in TIMSS 2011 differ in many respects (e.g., school types, educational resources, socialization norms, daily experiences). For instance, schools in the Arab countries are single-sex with male teachers teaching male classes and female teachers teaching female classes whereas in Ghana, Botswana and South Africa schools are predominantly co-educational although single-sex schools exist. For more detail information about the students assessed in TIMSS 2011 and their respective countries' educational systems, see Martin and Mullis (2012); the World Data on Education (UNESCO International Bureau of Education 2011) and UNESCO Institute for Statistics (UNESCO Statistics 2014).

Research focusing on African countries in TIMSS, is very important because to the best of our knowledge there is no rigorous method that evaluates the combined psychometric properties of TIMSS motivational constructs in these countries. Moreover, knowledge of the relations between affects and achievement is important in informing teaching practice because it has important and wide-ranging implications for educational interventions. For instance, many affective enhancement programs as well

as educational policy statements throughout the world are based on the fact that an improvement in affects (e.g., self-concept) will lead to better academic achievement (Bofah 2015).

This study will add to the existing literature on students' motivational constructs and help in the generalizability of these constructs. This is due to the fact that most of the literature on psychometric properties on motivational constructs is based on western research samples with undisputed excellent psychometric properties. As the TIMSS data has a cross-cultural perspective, it is a good source to use if researchers want to establish generalized theories, as research studies that integrate different cultural perspectives are crucial for establishing more useful and universal theories (Leung and Zhang 1995; Van de Vijver 2000).

In the present study, we first analyzed the psychometric properties of the TIMSS 2011 motivational constructs in an African context. Specifically, to ascertain whether the psychometric properties cut across these countries. Second, the study investigated the relationships between the constructs to ascertain the universality of the relationship across the five African countries. The study also investigated the existence of systematic country/cultural differences in the latent factor means for these constructs and mathematics achievements. Moreover, since school types are mostly single-sex in Arab countries (Morocco and Tunisia), gender differences in the constructs and achievement were investigated. Finally, the relationship between these constructs and other background variables such as, parental education, students' educational aspirations, and mathematics achievement were investigated. These variables were introduced into the analysis as *correlates*. The "...term *correlate* mean variables that are posited to be correlated with the latent groups without specifying a causal ordering" (Marsh et al. 2009, p. 7).

Motivational constructs in affect: cross cultural perspective

Cross-cultural comparative studies such as TIMSS and the Programme for International Student Assessment (PISA) have gained considerable attention recently. However, the problems associating with TIMSS and PISA studies stem from the fact that the target populations have unique cultures, school systems and cognitive structures (Metsämuuronen, 2012a, b). Moreover, there has been much discussion in the literature concerning the applicability of frameworks and the instruments used in measuring affective constructs in cross-cultural research such as TIMSS and PISA (e.g., Van de Vijver 2000; Steenkamp and Baumgartner 1998). First, the theories on which these studies are based are products of western cultures without cognizance of the fact that "the self" is a cultural construction that varies from culture to culture (Shen and Tam 2008; Markus and Kitayama 1991; Heine 2001). For instance, students' responses to mathematics self-belief items are based on their comparisons with their peers in the same school or class—frame of reference effect (e.g., Marsh 2007). Therefore, country-level analyses may not reflect individual students' self-belief (Marsh et al. 2013; Marsh 2007) and generalizability of findings may be problematic (Ho et al. 2000). Second, the discussion stems from the fact that importing and adapting instruments developed in western cultures into new cultural settings is problematic irrespective of their reliabilities in the original settings (e.g., Bofah and Hannula 2014, 2015). Third, research into the relationship between students' self-beliefs and their relationship to achievement at the country level goes far beyond

psychological theories, because these theories operate at the individual level, not at the country level (Shen and Tam 2008). Moreover, theories and studies associated with these constructs are products of western cultures and societies where the populations are ethnically homogenous.

Studies have shown that the scales (e.g., mathematics self-concept) that had been used in PISA and TIMSS studies were less reliable in East Asia, the Middle East, and some parts of Europe, when compared with North America, where the scales were originally constructed (Marsh et al. 2013; Metsämuuronen 2012a, b; Rutkowski and Rutkowski, 2010). Other studies (Bofah and Hannula 2014, 2015; Tuohilampi et al. 2013) have also reported similar findings with other imported instruments—the students' views of mathematics instruments (VOM). Bofah and Hannula (2014, 2015), in a confirmatory analysis of VOM instruments for responses by twelve grade Ghanaian students there was no support for an a priori factor structure. For example, the mathematics self-concept among Finnish students seems to have an underlying structure (e.g., ability and success), whereas for the Ghanaian students the construct is a single entity. They concluded that, the statistical method used in the previous VOM studies; translation of the VOM instruments and the dramatic cultural differences between these two countries may be the possible cause for this disparity. This study supports the necessity for the use of more robust approaches such as structural equation modeling in cross-cultural research. The failure of the original scale highlights the problems associated with the adapting of instruments to other research settings.

TIMSS motivational constructs are broadly divided into *self-confidence*— a specific form of *self-concept* (often referred to as *competence belief*), *positive affect* (i.e., intrinsic motivation: student like learning mathematics), and *task value—including importance* (i.e., extrinsic motivation: students' value for mathematics) (see Marsh et al. 2013). Motivation has been studied in the literature mostly by distinguishing between intrinsic and extrinsic motivation (e.g., Ryan and Deci 2000). In affect research, terminology ambiguity and definition is a major challenge in the mathematics-related affective literature (Furinghetti and Pehkonen 2002; Hannula 2012). The concept of motivation, beliefs, attitudes, values, and views have been interchangeably defined in the literature (e.g. see Areepattamannil and Freeman 2008; Andre et al. 1999; Cokley et al. 2001; Middleton and Spanias 1999). Particularly in large-scale research like TIMSS, motivations have been frequently used rather than beliefs, attitudes or views of mathematics.

We will consider the Expectancy-value framework proposed by Eccles and her colleagues (e.g., Eccles et al. 1983; Wigfield and Eccles 2000; Jacobs and Eccles 2000) on motivation as a framework for understanding the relationships between the various TIMSS motivational constructs, students' performance in mathematics, and other background variables such as parental education, gender, and students' long term educational aspirations. The expectancy-value model (e.g., Eccles and Wigfield 2002; Eccles[Parsons], et al. 1983; Jacobs and Eccles 2000; Wigfield and Eccles 2000) posits that individuals excel in subjects in which they expect to succeed or which they value (Leaper et al. 2012). The model posits that students' competency beliefs and task-value are positively related, and they directly influence their academic performance and goals (Eccles and Wigfield 1995; Jacobs and Eccles 2000). The model further posits that parents' and teachers' expectations, parental involvement, values, and beliefs influence students'

motivation, students' goals (e.g., educational aspirations), their competence beliefs (e.g., self-confidence) and task values (e.g., value for mathematics) (Eccles et al. 1990; Jacobs and Bleeker 2004; Jacobs and Eccles 2000). As far as the relationship between affect and achievement is concerned, most studies have found that they have a positive relationship (Ma and Kishor 1997; Marsh et al. 2013; Wilkins et al. 2002; Vermeer et al. 2000); however, others have found a negative relationship (e.g., Wilkins 2004) especially in international comparative studies (Kifer 2002; Shen and Tam 2008; Shen and Pedulla 2000). In a meta-analysis study, Hattie (2009) demonstrated that mathematics-related affect (e.g., mathematics self-concept) is the best non-cognitive predictors of individual achievement in mathematics (see also Stankov and Lee 2014). A cross-cultural dimension has been demonstrated (e.g., self-concept) for the causal determinism between self-belief and achievement, more typically in the western hemisphere (e.g., Parker et al. 2014; Seaton et al. 2014; Williams and Williams 2010; Hannula et al. 2014) and recently in the African context (e.g., Bofah 2015).

Parental involvement, parental education, and teacher responsiveness

Parents' and teachers' expectations and attitudes have been found to be important in shaping students' self-concept and expectation for success (Eccles [Parsons] et al. 1983; Jacobs and Eccles 2000). Parents and teachers are socio-actors with many social networks that can shape children's education, which subsequently may have a significant impact on a child's education (Ames et al. 1993; Sheldon 2002; Marchant et al. 2001). The important role that parents and teachers play on children's school achievement has been thoroughly confirmed in a study by Marchant et al. (2001). Parental involvement have been known to influences children's values and future aspirations (e.g., Jacobs and Eccles 2000). Many studies on parental involvement have found that parental involvement in a child's education has a positive impact (Abu-Hilal 2001; Campbell and Mandel 1990; Epstein, 2008; Epstein 2010; Fan and Chen 2001; Ice and Hoover-Dempsey 2011; Gonzalez-DeHass et al. 2005) whereas others have found a negative (e.g., Desimone 1999; Fan and Chen 2001) or no impact (e.g., Topor et al. 2010; Reynolds 1992).

There is evidence to support the argument that students tend to have positive self-confidence if there is positive support and encouragement from their parents to accomplish their goals and aspirations (Jacobs and Eccles 2000; Ormrod 2011). Meta-analytic studies (e.g., Fan and Chen 2001; Jeynes 2003, 2007) have found a positive relationship between parental involvement and academic achievements and a longitudinal study by Ice and Hoover-Dempsey (2011) revealed a significant positive relationship between parental involvement and student achievement outcomes. Keith and his colleagues (1993) examined the effects of parental involvement and the achievement of 21,814 eighth-graders from the National Educational Longitudinal Study (NELS) data. The results indicated that parental involvement had an important influence on 8th-grade students' mathematics achievement. The findings confirmed that parental involvement with students' homework might have influenced the results. Moreover, Sharp (1995) found that parental involvement was consistently associated with mathematics achievement. In a study concerned with the effects of parental involvement on eighth-graders achievement, Bhanot and Jovanovic (2009) found that children's self-perception about their mathematics competency are functions of their parents' and teachers' perceptions of the children's

competency. Davis-Kean and her colleagues (2007) in a longitudinal study analyzed how parents' values and attitudes affect children's mathematics performance and later interest, and how these attitudes vary with the child's gender. Between 1987 and 2000, using a data set of more than 800 children and their parents, they found that parental attitudes were important determinants of children's mathematics performance and later interests. The study found that girls' interest in mathematics decreases as their fathers' gender stereotypes increase, whereas the contrary was true for boys. The study further reported that parents provided a more mathematics-supportive environment for boys than girls, including the purchase of toys and time spent on activities.

The literature on higher parental education claims that parental education is significantly associated with students' academic performance (e.g., Davis-Kean 2005; Haveman and Wolfe 1995; Klebanov et al. 1994). Moreover, parents who are recipients of higher education place a great deal of emphasis on the value and importance of education, while committing themselves more to their children's school activities—for example helping with their homework (e.g., Sy 2006). Chen and Stevenson (1995) found that students whose fathers had postgraduate education scored almost 10 points higher on an achievement test than students whose fathers had junior high school education or less.

Teachers are surrogate parents with a similar socializing influence as students' parents (Herring and Wahler 2003). As Herring and Wahler (2003, p. 120) put it: "a responsive teacher knows when to praise, to instruct, to acknowledge, to warn, and to penalize his or her students. As a result of these well-timed reactions and approaches, most children are ready to work and ready to learn". Research has shown that a close positive student-teacher relationship is associated with high student academic performance (Hughes et al. 2005; Hamre and Pianta 2001; Wright et al. 1997; Connor et al. 2005, 2006). Moreover, teacher responsiveness has been found to be significantly associated with children classroom responsiveness and negativity (Herring and Wahler 2003), and a lower positive attitude toward mathematics is a possible function of a lack of teacher supportiveness and an unsuitable classroom environment (Middleton and Spanias 1999). For example, Connor and colleagues found that students whose teachers were more responsive while also spending more time on academic activities had better academic success. Wright et al. (1997) found the teacher effect is a dominant factor affecting student academic gain more than the classroom context variables of heterogeneity among students and class sizes. However, the teacher-student relationships in the classroom and teaching practices have been found to be culturally specific (Connor et al. 2006; Page 1987; Lerkkanen et al. 2012). For example, Page (1987) found that teachers' definitions of students are reflections of the educational culture of a school, their perceptions of students' social constructions (e.g., race, SES) and students' academic status in the school (e.g., high- or low-ability track).

Gender and educational aspirations

The literature concerning the relationship between gender, motivation, and mathematics achievement is abundant. The relationship between motivational beliefs and performance for male and female students' is important in unlocking the gendered mechanisms that affect mathematics performance and participation (e.g., Simpkins and Davis-Kean 2005; Watt et al. 2012; Nagy et al. 2010). Social and cultural barriers have

been known to influence gendered motivational belief and achievement. These barriers have been found to vary within and across nations (e.g., Hyde and Mertz 2009).

Researchers have found that gender differences in mathematics have decreased and in some instances they are non-existent (Else-Quest et al. 2010; Ma 2008; Marks 2008; Mullis et al. 2012). However, some studies have shown a higher achievement for girls (e.g., Ma 2008; Vermeer et al. 2000), for boys (e.g., Ma 2008; Marks 2008; Cheema and Galluzzo 2013; Guiso et al. 2008; Fryer and Levitt 2010) and no difference between the sexes (e.g., Cheema and Galluzzo 2013; Marks 2008; Chen 2002). Studies across different countries have consistently identified a cultural dimension for gender differences in mathematics-related affect and performance (Hyde and Mertz 2009; Forgasz et al. 2015) with studies indicating that gender significantly influences students' mathematics self-concept, mathematics self-confidence, and the perceived usefulness of mathematics (e.g., Belcher et al. 2006). For instance, studies have reported lower mathematics confidence, less liking for mathematics, a lower mathematics self-concept, and a lower value for mathematics for girls in science and mathematics tasks (Frenzel et al. 2010; Hyde et al. 1990; Else-Quest et al. 2010; Marsh et al. 2013; OECD 2015; Nagy et al. 2010; Watt 2004; Watt et al. 2012). Moreover, other studies have consistently shown gender differences favoring males in their mathematics ability even when there is no evidence of any achievement differences between the sexes (Frenzel et al. 2010; Nagy et al. 2010; Watt 2004). However, other studies have found that mathematics-related affect regarding the value placed on mathematics do not differ as a function of gender (e.g., Jacobs et al. 2002; Watt 2004).

In the Arabic world, consistent with the TIMSS findings (see also, Marsh et al. 2013; Ayalon and Livneh 2013), studies have found higher achievements favoring girls in Middle Eastern Arabic-speaking countries (e.g., Abu-Hilal 2001; Fryer and Levitt 2010; Mullis et al. 2012) but favoring boys in Arabic countries in Africa (Ayalon and Livneh 2013). A possible explanation for the Middle Eastern Arabic-speaking scenario is that girls are more restricted to their homes than boys, and therefore they have more time to focus on their schoolwork than boys (Abu-Hilal 2001). Other reasons given are that it may be a result of the single-sex schooling system in these countries, which are normally the types of schools found in the Arab world (e.g., Fryer and Levitt 2010).

On the relationship between parental involvement and gender, studies have found that, in general, girls enjoy higher parental involvement with their schooling than boys (Gunderson et al. 2012; Jacobs and Bleeker 2004). Cultural stereotypes (e.g. mathematics is a male subject) have been shown to influence parents' perceptions of their children's abilities. For example, some studies (e.g., Jacobs and Bleeker 2004) have indicated that parents are more involved in their daughters' mathematics and science activities than in their sons' activities, resulting in an increased interest in children's mathematics and science activities (see also Hoover-Dempsey and Sandler 1995). Other studies found that girls are less encouraged to partake in mathematics/science related activities than boys and parents perceive boys as more competent in mathematics than girls as well as perceiving mathematics as a male subject (e.g., Eccles[Parsons] et al. 1983). Bhanot and Jovanovic (2009) found that girls value science more when their parents believed that science was important for them but this relationship was disconnected in the case of boys. See Hoover-Dempsey and Sandler (1995), and Hoover-Dempsey et al. (2005) for an in-depth review of "Why do parents become involved in children's education?"

Several meta-analyses have attempted to corroborate the patterns of sex differences in the classroom in respect to teacher-student responsiveness (Kelly 1988; Jones and Dindia 2004). Although the authors in the meta-analyses were cautious in their findings, nevertheless the outcome was that teacher-responsiveness favors male students. For instance, Jones and Dindia (2004) concluded that gender may not be the sole attribute but other factors such as school subject, student classroom behavior, and achievement, and the sex of the teacher may moderate the relationship. However, Cornbleth and Korth (1980) and Brady and Eisler (1999) found no sex differences in teacher-student responsiveness in the classroom. Given the mixed results, Canada and Pringle (1995) suggested that the social context within which the classroom interactions occur (i.e., whether the classroom does or does not include men) can influence the teacher responsiveness. Canada and her colleagues observed that teacher responsiveness varied as a function of gender composition of the classroom.

As far as students' educational aspirations are concerned, it is a known fact that as the level of education rises; individuals with lower degrees are at a disadvantage in terms of earnings and employability (Sanders, Field, and Diego 2001). Studies have shown that higher levels of educational aspirations significantly correlate with higher academic achievement (Gil-Flores et al. 2011; Saha, 1994; Sanders et al. 2001; Marjoribanks 2002, 2003a, b). For instance, Marjoribanks (2002, 2003a) found that students' academic aspirations significantly contribute to their educational achievement and affect (e.g., self-concept). There are mixed findings in the literature about any gender differences in students' educational aspirations: with no gender difference reported (e.g., Garg et al. 2002; Watt et al. 2012); some studies have reported higher educational aspirations for females (e.g., Mau 2000); and in others, males were found to have higher educational aspirations (e.g., Mendez and Crawford 2002; Wilson and Wilson 1992).

Moreover, students' expectancies and goals are directly related to their mathematics self-concepts and the perceptions of their parents' and teachers' attitudes towards their mathematics competence (Eccles[Parsons] et al. 1983; Gunderson et al. 2012). For instance, studies have shown that even when there was no evidence of achievement differences between males and females in mathematics, parents' perceptions of their children's mathematics achievement were influenced by the child's gender (e.g., Eccles et al. 1990; Eccles[Parsons] et al. 1983), with the parents of boys believing that their children have higher mathematics abilities than the parents of girls. This gender-stereotyping of children's mathematics abilities can influence children's self-concept because they in turn internalize gender stereotypes directly or indirectly (Eccles et al. 1990; Gunderson et al. 2012). From the above, we can see that gender-role stereotyping and societal influence (e.g. parents and teacher) have a significant influence on students' mathematics motivation. We will explore gender and educational aspirations differences across the five countries in an attempt to determine how they are interrelated with motivational beliefs and achievement.

A priori predictions and the research question

In the present study, we will investigate the psychometric properties (factor structure, reliabilities, method effect, and measurement invariance—country and gender) of the TIMSS 2011 motivational constructs (e.g. mathematics confidence, students like

learning mathematics, students' value for mathematics, teacher responsiveness, and parental involvement) across the five educational systems/nations/cultures. We tested if the factor structure of the constructs supports the a priori structure they were designed to measure. Specifically, we tested the reliability of the constructs, measurement invariance of the constructs across the five nations, country differences and their relationship to background variables such as students' parental education, aspirations, mathematics achievement, and gender differences. Specific hypotheses with respect to the aims are discussed below.

Factor structure

First, the motivational constructs have been developed in western contexts. Second, TIMSS 2011 included new items in the motivational constructs that were not present in previous years. Therefore, we leave open the research question whether the 28 motivational items support a five-factor a priori structure, i.e., the student confidence in mathematics (SCM) scale, the students like learning mathematics (SLM) scale, the students value mathematics (SVM) scale, the teacher responsiveness scale (TRES), and the parental involvement scale-positive parenting (PIV). This a priori factor structure is derived from the design of the TIMSS 2011 scales. However, we did expect a high correlation between the students confident in mathematics scale and the students like learning mathematics scale.

Method effects: the negative item effect

On balance, researchers include negatively worded items or reverse-coded items in survey instruments to reduce the effect of response pattern biases (e.g., acquiescence) (Numally 1978). The notion is that negatively worded items may act like cognitive "speed bumps" that require respondents to engage in more controlled, as opposed to automatic cognitive processing (Podsakoff et al. 2003, p. 884). Unfortunately, a review of the literature does not support these assertions. However, responses to negatively worded items have been found to yield systematic variance that is irrelevant to the content under study, this is irrespective of the age group but it is more commonly with young children (e.g., Benson and Hocevar 1985; Hooper et al. 2013; Urbán et al. 2014; Marsh 1986). Marsh and his colleagues define method effects as "nontrait effects associated with idiosyncratic aspects of particular items or methods of data collection" (Marsh et al. 2013, p. 112).

One common source of method effect is the use of negatively worded items in a survey instrument. Studies have indicated that the method effect can be seen if at least about 10 % of respondents fail to acknowledge the presence of negatively worded items in a survey instrument and respond appropriately (Woods 2006; Schmitt and Stults 1985). Moreover, many studies on constructs validation (e.g., even for older students and adults) have shown that positively and negatively worded items in a survey often form two distinct factors structures in factor analysis (e.g., Marsh 1986, 1996; Metsämuuronen 2012a; Roszkowski and Soven 2010; Chiu 2008, 2012). However, the split in factors is very much associated with uni-dimensional construct.

Studies have indicated that response to negatively worded items differ cross-culturally, (e.g., Schmitt and Allik 2005; Benson and Hocevar 1985; Marsh 1986; Hooper et al. 2013;

Metsämuuronen 2012a). Using self-esteem measures (i.e., the Rosenberg Self-Esteem Scale) in 53 countries, D. P. Schmitt and his colleagues found that "... in many cultures the answers to negatively worded items are systematically different from the answers to positively worded items" (Schmitt and Allik 2005, p. 638; see also Hooper et al. 2013; Metsämuuronen 2012a). However, other studies have found that the method effect associated with negatively worded items are related to age and verbal ability (Marsh 1986, 1996; Corwyn 2000; Metsämuuronen 2012a), insufficient cognitive ability in mathematics and science (Hooper et al. 2013; Metsämuuronen 2012a), careless response patterns (Schmitt and Stults 1985; Woods 2006), and substantively irrelevant artefacts (Marsh 1996). Moreover, a direct relationship between negatively worded items and achievement has been established. The results hint at a relationship between country-level item fit and country-level performance on assessment (Hooper et al. 2013; Metsämuuronen 2012a).

When method effects are not accounted for, they have been shown to affect the goodness of fit statistics, lead to biased parameter estimates (e.g., Chiu 2012, 2008; Bagozzi 1993; Quilty et al. 2009; Marsh 1994, 1996; Distefano and Motl 2009; Marsh et al. 2013; Lindwall et al. 2012; Tomõs et al. 2013; Horan et al. 2003; Magazine et al. 1996) and influence the validity and reliability of the scale (e.g., Roszkowski and Soven 2010; Woods 2006; Raykov 2001; Brown 2015). Moreover, the method effect can lead to wrong inferences by suppressing or inflating the relationships between constructs by contributing to Type I or Type II errors if not incorporated into the measurement model (e.g., Bagozzi 1993; Magazine et al. 1996).

The methodological strategy commonly used to counteract the method effect fall within the framework of multitrait-multimethod (MTMM) under structural equation modeling. The most common approach is the correlated traits, correlated uniquenesses (CTCU) framework (e.g., Kenny and Kashy 1992; Marsh 1986, 1996; Distefano and Motl 2009; Marsh et al. 2013; Horan et al. 2003). The CTCU framework treats the negative wording effect as a methodological artefact rather than a distinct factor by allowing the item uniqueness related to the negatively worded items to be correlated (Ye and Wallace 2013). Thus, the CTCU framework allows the method wording effect to be removed from the construct by not allowing the wording effect to be examined as a unique factor (Distefano and Motl 2009). This helps eliminate any irrelevant empirical relationship with the constructs or variables in the study. Method effects related to negatively worded items using the CTCU method has been reported in both TIMSS 2003 (e.g., Chiu 2008, 2012) and TIMSS 2007 (e.g., Marsh et al. 2013, 2014), and the National Educational Longitudinal Survey (NELS:88) (Marsh 1994) for the self-concept variate.

In line with the findings presented above we hypothesized that there should be method effects associated with negatively worded items. Thus, achieving a model with an acceptable goodness of fit, unbiased standard errors and parameter estimates would require controlling for the method effects associated with the use of negatively worded items in our data set. We also expect to find a relationship between the country-level model fit due to method effect associated with the negatively worded items and country-level performance on assessment.

Reliability

Although it is subject to much misuse, misunderstanding and confusion (Sijtsma 2009; Yang and Green 2011), Cronbach's alpha (α) has for a long time been the most widely used quality indicator test statistics measure (e.g., Novick and Lewis 1967). Moreover, α is known to either underestimate or overestimate reliability (Bofah and Hannula 2014; Brown 2015; Geldhof et al. 2014; Novick and Lewis 1967), and studies have shown that imported constructs regardless of their high reliabilities in the original settings, often show lower reliabilities when imported to different cultural settings (see Bofah and Hannula 2014, 2015; Tuohilampi et al. 2013). Furthermore, if the underlying construct includes correlated measurement errors, α can underestimate or overestimate the reliability construct depending on the parameters of the measures (e.g., Brown 2015; Raykov 2001; Peterson and Kim 2013; Bentler 2009; Green and Yang 2009; Raykov 2001; Yang and Green 2011). The composite reliability measure (ω) (Raykov 2012), which is usually associated with CFA models was estimated to complement the α estimates. The use of ω provides reliability estimates, which are in the direct context of the estimated CFA model. Composite reliability incorporates the computed factor loadings, error variance, error covariances (method effect if any, e.g., correlated uniquenesses associated with negatively worded items) and produces more precise estimates of reliability than the reliabilities estimates provided by α (Brown 2015; Geldhof et al. 2014; Raykov 2012; Yang and Green 2011). It is interpreted in the same way as α ; ω values of 0.600–0.700 are acceptable in exploratory research (Hair et al. 2010). Because the constructs come largely from western research and empirical evidence based on previous TIMSS research (e.g. Rutkowski and Rutkowski 2010; Metsämuuronen 2012a, b), we hypothesized that the reliability estimates would be lower in our sub-sample. Moreover, we expected the reliability of the confidence and like mathematics constructs to be affected by the method effect associated with the negatively worded item (e.g., Roszkowski and Soven 2010; Woods 2006; Brown 2015).

Using robust maximum likelihood (MLR) to account for the normality in the scales, the MLR estimator for the Composite reliability (ω) was computed as:

$$\omega = \frac{(b_1 + \dots + b_i)^2}{(b_1 + \dots + b_i)^2 + \theta_1 \dots + \theta_i},$$

where b_i represents the i th factor loading onto a single common factor and θ_i represents the unique variance of item i (with factor variance fixed at 1). In the event that the model contains correlated error covariance, the denominator is extended by twice the sum of the estimated error covariances

$$\omega = \frac{(b_1 + \dots + b_i)^2}{(b_1 + \dots + b_i)^2 + \theta_1 \dots + \theta_i + 2 \sum_{j < k} Cov(E_j, E_k)},$$

(Raykov 2012).

Measurement invariance

We tested for measurement invariance, i.e., if the measurement instruments are equal across the different educational/cultural systems. Measurement invariance is

an important aspect of cross-cultural research and constructs validation. Moreover, because of the diverse cultural backgrounds of students from the participating countries in this study, we leave open the hypothesis whether there is support for an invariance of factor loading (metric invariance), and item intercept (scalar invariance) across the countries or educational systems.

Country differences in achievement and the motivational constructs

The results from TIMSS 2011 (Mullis et al. 2012) indicated that the performance of some countries within the present study were among the lowest. We expect lower student motivation to lead to lower student mathematics achievements but as Marsh et al. (2013, p. 7) argue “the process of forming self-beliefs is complex and is not a simple function of achievement levels”. Moreover, Marsh and his colleagues argue that there is a strong frame of reference effect associated with students’ self-belief (Marsh 2007; Marsh, et al. 2013). In most cases comparisons are made in reference to other students in the same school or class, so country level differences may not be reflected in individual students’ mathematics self-beliefs. For instance, Tunisia and Morocco are Arab countries with educational systems based mainly on single-sex schools, therefore their students’ frames of reference are specific for each gender (Marsh et al. 2013) whereas in Ghana, students study in both single-sex and co-educational schools consequently the frames of reference of co-educational students and single-sex students may differ. Because of the diverse school systems in the present study, and since the literature on the relationship across different cultures is diverse coupled with no clear consistent findings in respect to the relationship between mathematics self-belief and achievement at the country-level (e.g., Lee 2009), we leave open the question of how achievement relates to these constructs. However, we expected to find a positive relationship between the students confident in mathematics, the students values for mathematics and students like mathematics scales.

Gender and educational aspiration difference

Studies have indicated that differences in gender achievement levels have been reduced (e.g., Ma 2008; Else-Quest et al. 2010), are non-existent (e.g., Cheema and Galluzzo 2013; Hyde et al. 1990; Ma 2008; Mullis et al. 2012) but still persist in some countries (e.g., Ma 2008; Mullis et al. 2012; Cheema and Galluzzo 2013). For instance, in Eastern Arab countries girls have been consistently performing better than boys in TIMSS (e.g., Mullis et al. 2012). With this inconsistency, it is difficult to predict which gender in our present sample would have better achievement performance for TIMSS. We have also left open the research questions concerning the relationship between the motivational constructs and student gender. However, we did expect higher parental involvement and aspiration for girls, and higher mathematics confidence for boys.

For the students’ long term educational aspirations (thereafter aspirations)—aspiring to higher levels of education—have been found to significantly correlate with higher performance (e.g., Gil-Flores et al. 2011; Marjoribanks 2002, 2003a, b). Moreover, students’ aspirations have been found to influence their affective (e.g., self-concept) dispositions (Eccles [Parsons] et al. 1983; Gunderson et al. 2012). However, because of the diverse cultural background of the students’ in our study we leave open the question of how

students long term educational aspirations relate to achievement and the motivational constructs.

Methods

Sample and measures

TIMSS is an international assessment of mathematics and science for fourth and eighth graders that has been conducted every 4 years since 1995 by the International Association for the Evaluation of Educational Achievement (IEA) (Mullis et al. 2012). In TIMSS 2011 45 countries/school systems participated in the eighth grade assessment. All students enrolled in grade eight were eligible to take part in the survey. The mean age for the present Africa participants was 15.58, with a median age of 15.42 and a standard deviation of 1.40. The sampling design is a two-stage cluster design consisting of sampling of schools and sampling of intact classrooms from the target grade in the schools. There were 989 intact classrooms (clusters) with an average of 39 students per class. For the present study, the participants were eighth-grade students from Ghana (7323 students, 48 % female, from 173 intact classrooms with an average cluster size of 41), Morocco (8986 students, 48 % female, from 279 intact classrooms with an average cluster size of 32), and Tunisia (5128 students, 51 % female, from 207 intact classrooms with an average cluster size of 25) and the ninth grade in Botswana (5400 students, 49 % female, from 151 intact classrooms with an average cluster size of 35), and South Africa (11,969 students, 49 % female, from 317 intact classrooms with an average cluster size of 37). Countries have the options of testing a higher grade if the questions are too difficult for their eighth graders. As such, Botswana and South Africa tested grade 9.

The TIMSS 2011 students' affect questionnaire was designed to measure students' self-perception and attitudes toward mathematics. As is usual for TIMSS, in TIMSS-2011 new items were included and others dropped from the TIMSS 2007 items. As shown in Additional file 1: Table S1, the students like learning mathematics (SLM) scale was created based on the students' degree of agreement with five statements. The students' value mathematics (SVM) scale was created based on the students' degree of agreement with six statements. The students confident in mathematics (SCM) scale was created based on students' degree of agreement with nine statements; whereas both the parental involvement/positive parenting (PIV) scale and the teacher responsiveness (TRES) scale consist of four items each. The teacher responsiveness scale is a measure of the extent to which students perceive their teachers are caring, helpful and responsive to their learning needs. Responses for the students like learning mathematics scale, the students value mathematics scale, the students confident in mathematics scale, and teacher responsiveness were based on a four-point Likert scale: Agree a lot (1), Agree a little (2), Disagree a little (3), Disagree a lot (4). On the other hand, responses to the parental involvement scale range from: Every day or almost every day (1), Once or twice a week (2), Never or twice a month (3), Never or almost never (4). In the present study, the scores were reverse-coded so that a higher value corresponds to a higher response. In this study, we use the terms confidence, value of mathematics, like mathematics and achievement, to mean the Students Confident in Mathematics scale (SCM), the Students Value Mathematics scale (SVM), the Students Like Learning Mathematics scale (SLM), and achievement in mathematics respectively. In particular, for readability purposes the

terms self-concept, confidence, self-confidence have been used interchangeably in the present study.

Three demographic variables were employed as indicators of students' background variables:

- a. Gender (coded 1 for girls, 2 for boys);
- b. Long-term educational aspirations—participants were required to indicate the highest level of education they expected to attain on a scale ranging from (1) “Lower secondary education” to (6) “University program- Master/Doctorate” (See Foy et al. 2013 for specific nationally definition classifications).
- c. Parental education—respondents were asked to record the highest level of education completed by their male, and female guardian on a seven-point scale; (1) “Some Primary or [some] Junior Secondary or did not go to school” to (7) “University program- Master/Doctorate”. The parental education variable was further collapsed into 5 categories: if both parents (5) “Finished University or higher”, (4) “Finished Post-Secondary Education but not University”, (3) “Finished Upper-secondary”, (2) “Finished Lower-Secondary”, (1) “Finished some primary or lower-secondary schooling or did not go to school”. High scores represent high levels of education and low scores represent low levels of education (for specific nationally definition classifications, see Foy et al. 2013).

Achievement scores

TIMSS 2011 reported students' achievements scores in five plausible values—they are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual. Following the advice from the TIMSS 2011 user guide and the TIMSS technical report we used all five plausible values in our analyses (Foy et al. 2013). The use of plausible values has been discussed in detail in the TIMSS2011 user guide (see Foy et al. 2013). The reliability for the achievement scores was (α : 0.966; ω : 0.962) for the whole sample (Botswana α = 0.969, ω = 0.966; Ghana α = 0.946, ω = 0.942; Morocco α = 0.957, ω = 0.953; South Africa α = 0.966, ω = 0.963; Tunisia α = 0.966, ω = 0.963). We performed all data analyses five times, once with each plausible value, the results were aggregated to obtain parameter estimates and standard errors.

In the present sample, Morocco and Ghana had significant percentages of very low eighth-grade performing students (percentage of students with achievement levels too low for estimation exceeded 25 %) with South Africa and Botswana having significantly low performing ninth-grade students (percentage of students with achievement levels too low for estimation exceeded 15 % but did not exceed 25 %) (Mullis et al. 2012). Therefore, any differences in achievement when comparing these countries must be interpreted with caution.

Analysis

Estimations

All analyses in the study were done using Mplus 7.31. We used the IEA International Database (IDB) Analyzer (ver. 3.1) (IEA IDB 2014) incorporated into IBM SPSS (ver. 21) for combining the data sets from the different countries and to select the background

variables. To control for non-normality (skewedness or kurtoses) in the data set, we used the robust maximum likelihood estimator (MLR) with standard errors and a Chi square test statistic that is robust to non-normality and non-independence of observations (to account for the dependency issue of students nested within classrooms) (*Muthén and Muthén 1998-2012).

Skewness (−2.848–0.276) and kurtoses (−1.438–8.053) were high above the recommended range (Kline 2005) for the motivational items. Missing data was treated using the Mplus multiple imputation procedure (Asparouhov and Muthén 2010; Schafer and Graham 2002). As is common, the input of 5 datasets was required to remove any uncertainty in the imputed process (Allison 2003; Schafer and Olson 1998). All variables used in the analyses were included in the imputation process. Because of the design of the TIMSS survey, with students clustered within schools, we used the Mplus complex survey design option to control the clustered design and adjust standard errors. Class was used as the clustering variable as it is used to uniquely identify the sampled classrooms within a country. Sampling weights were also taken into account in the analyses (*weighting variable supplied with the data*).

Model evaluation criteria

Evaluation of a model fit was based on multiple criteria. Since the Chi square test is sensitive to large sample size, we considered the Root Mean Square Error of Approximation (RMSEA), the comparative fit index (CFI) and the change in fit between nested measurement invariance models (e.g. Δ CFI) (Chen 2007). The value of CFI can vary between 0-to-1. Fit is considered adequate if the CFI values are >0.900 and better if they are ≥ 0.950 , and RMSEA ≤ 0.060 for models with a better fit, an RMSEA ≤ 0.080 indicates an adequate model fit (Bentler and Bonett 1980; Brown and Cudeck 1993; Hu and Bentler 1999). We examined the change in CFI. If the decrease in model fit for the more restrictive model is less than or equal to 0.010 for CFI, then there is reasonable support for the more restrictive model (Chen 2007). These thresholds were used as mere guidelines because they can vary across studies (Kline 2005). We also took into account our a priori prediction, common sense, comparison of viable alternative models and detail evaluation of parameter estimates (see Marsh et al. 2013).

Test for measurement invariance

To make any meaningful comparison, researchers test if a measured construct across groups in any population measures “the same thing”. The process is known as the measurement invariance. Invariance measurements allow the comparability of scores between groups. They involve the process of investigating whether items contribute equally (often based on the factor loadings, intercepts or thresholds) across these variables, “thus ensuring that the constructs are operationalized similarly (Sass 2011, p. 348)”. A multigroup invariance test begins with a *configural invariance* model, (a model with all the parameters freely estimated) which is the baseline model for comparing other models. Cultural differences may lead to a lack of *configural invariance* when a construct is imported from one cultural setting to another (Chen 2008). A good fit should be achieved to pave the way for the estimation of later models. The *configural* model is followed by a *metric invariance* model, which requires that the factor loadings should be

equal (invariant) across the groups. If *metric invariance* holds, we could conclude that the constructs are manifested the same way in each of the groups (Millsap and Olivera-Aguilar 2012). The most probable cause for non-metric invariance is the importation of a scale from one culture to different cultures when the definitions and meanings of concepts do not overlap across the cultures, because of inappropriate translation, or/and the respondents' tendencies to use or avoid extreme responses (Chen 2008).

Scalar invariance requires the indicator intercept and factor loadings to be invariant across the groups. *Scalar invariance* is important because it implies that population differences in the means of the measured variables must be due to the influence of common factors (Millsap and Olivera-Aguilar 2012). Support for *scalar measurement invariance* indicates that mean differences between groups at the item level can be explained in terms of differences at the latent factor mean level. Support for *scalar invariance*, helps eliminate ambiguity in the explanation of group mean differences (Chen 2008). Possible causes of *scalar invariance* are social desirability, the frame of reference effect (see Marsh 2007) and when a group within the study is preoccupied with its own defects or deficiencies, which may produce a significant effect on their responses (Chen 2008, p. 1007). Finally, there is *strict measurement invariance*, which requires invariance of item uniqueness as well as factor loadings and intercepts. The most severe possible cause is the method effect. In the present study, the five countries were treated as grouping variable for the multi-group analysis.

We also evaluated gender invariance across the five countries. We created 10 groupings (10 groups: 2 gender \times 5 countries) and performed measurement invariance with the grouping variable.

Results

The factor structure and method effects

We posit that the 28 motivational items can be explained by 5 factors namely: students' confidence in mathematics, like mathematics, value of mathematics, teacher responsiveness, and parental involvement. These factor structures are a result of how TIMSS has operationalized the motivational constructs in their study. The 6 negatively worded items (see Additional file 1: Table S1), 4 in the confidence in mathematics scale, 2 in the like mathematics scale were set correlated in order to control the method effect associated with including negatively worded items in the questionnaire. The rationale for this approach has been outlined and illustrated by Chiu (2008, 2012); Marsh (1994, 1996); Quilty et al. (2009); Distefano and Motl (2009) and Marsh et al. (2013) and there are appeals for extensively in the section discussing the method effect associated with negatively worded items. As can be seen in Additional file 2: Table S2, the fit indices for section A (part without the correlated uniqueness of the negative items) tend to be outside the range of what is generally considered to be an adequate fit (i.e., CFI, >0.90) for each country. However, as indicated in section B, when the error terms of the negatively worded items were allowed to covary, the model fit improves dramatically for all the cases except Tunisia. The variability of the improvement in the CFI fit from 0.053 in Tunisia to 0.155 in Ghana indicates that the method effect associated with the negatively worded items appears to have an effect in all the countries, but the effect is less

pronounced in some countries than others. Moreover, the correlation between the negatively worded items was statistically significant in all the countries.

The total sample Model M0 (Additional file 2: Table S2, Section A) is the theoretical model; this did not fit our data set due to the method effect. Including the correlated uniquenesses between the negatively worded items in the model improved the model fit to the acceptable level seen in Model M1 (Additional file 2: Table S2, Section B). This finding supports our hypothesis that achieving an acceptable model fit requires that steps should be taken to control for the method effects associated with the combined use of negatively and positively worded items in our data set. Moreover, it supports the literature (e.g., Chiu 2012; Marsh 1996; Marsh et al. 2013). Model M1 depicted in Fig. 1 was the baseline model for subsequent analyses. Although, the factor loadings (Additional file 1: Table S1) were all statistically significant, the loadings for the negatively worded items were very low.

Invariance across nations over the motivational constructs

In order to determine to what extent the factor structures for mathematics confidence, like mathematics, value of mathematics, teacher responsiveness, and parental involvement can be generalized to each of the groups and that the country differences have not been influenced by the characteristics of the underlying constructs, multi-group confirmatory analysis was implemented. Indicated in both Additional files 2, 3: Tables S2, S3 as Model 1 (M1), and depicted in Fig. 1 is the hypothesized model (the baseline model) for the multi-group analysis. As indicated earlier, multiple group (MG) modeling starts with a *configural model*. As Additional file 3: Table S3 indicates, the *configural model* (MG2) provided a good fit to the data, indicating support for *configural* validity across nations, thus paving the way for further model testing.

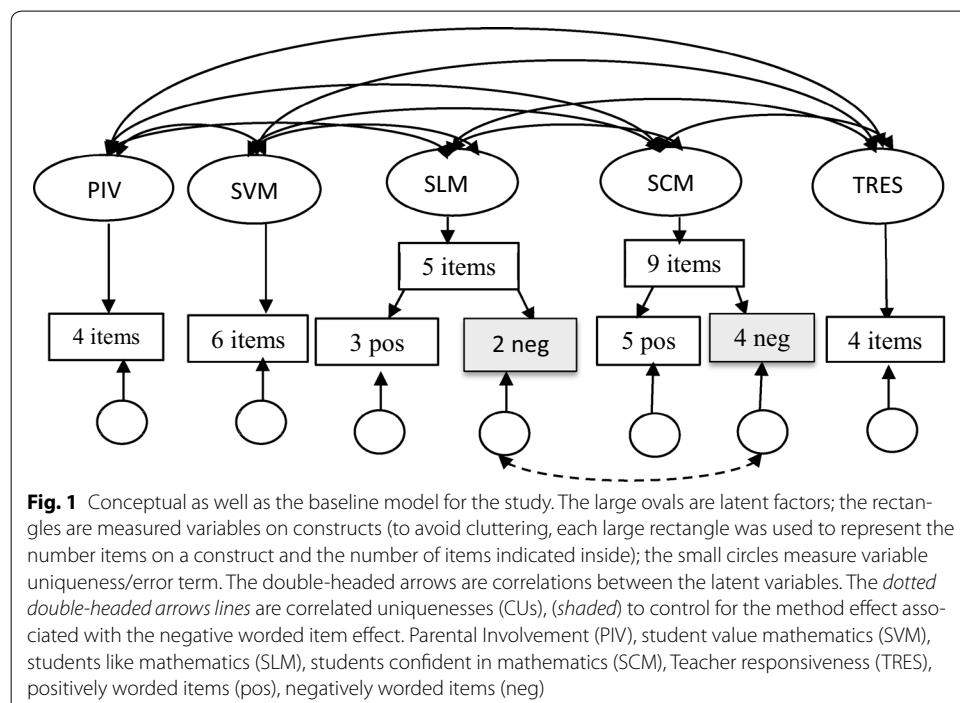


Fig. 1 Conceptual as well as the baseline model for the study. The large ovals are latent factors; the rectangles are measured variables on constructs (to avoid cluttering, each large rectangle was used to represent the number items on a construct and the number of items indicated inside); the small circles measure variable uniqueness/error term. The double-headed arrows are correlations between the latent variables. The *dotted double-headed arrows lines* are correlated uniquenesses (CUs), (*shaded*) to control for the method effect associated with the negative worded item effect. Parental Involvement (PIV), student value mathematics (SVM), students like mathematics (SLM), students confident in mathematics (SCM), Teacher responsiveness (TRES), positively worded items (pos), negatively worded items (neg)

With regard to the *metric* invariance (MG3), the results revealed that on comparing the *metric* invariance to the *configural* invariance (MG2), the *metric* invariance model fits the data adequately. In addition, the drop in fit (i.e., ΔCFI) is within the acceptable cut-off limits. Metric invariance is retained. Thus, confidence in the mathematics scale, the like mathematics scale, the value mathematics scale, the teacher responsiveness scale, and the parent involvement scales are invariant across the five African countries. In other words, the constructs can be generalized to have been measuring the same dimensions across the five African countries. We then proceeded to testing *scalar invariance* (Model 4 MG4). Model 4 (MG4) the *scalar invariance* fit is not as good a fit as the *metric invariance* model (MG3). Although RMSEA supports the model fit, CFI did not (including $\Delta\text{CFI} > 0.010$). Thus, *scalar invariance* was not supported. This is in line with Chen's (2008) argument that measurement invariance especially *scalar invariance* is rarely seen in cross-cultural research. This is a possible sign of social desirability, and the frame of reference effect (Chen 2008) in the students' responses.

Since the invariance of item intercepts is important for the interpretation of latent mean differences, *partial invariance* (e.g., Byrne et al. 1989; Steenkamp and Baumgartner 1998) for the item intercepts is necessary. On the basis of the modification indexes guidelines proposed by Byrne et al. (1989), and Cheung and Rensvold (1998), we explored the partial measurement invariance model. A *partial scalar* invariance model is one in which some measurement intercepts are invariant and others are not. This resulted in allowing 4 out of 9 'mathematics confidence items', 2 out of 5 'like mathematics' items, 1 out of 4 'teacher responsiveness' items and 2 out of 6 'value of mathematics' items to be freely estimated across countries. The resulting Model 5 (M5)¹ supported *partial* invariance for item intercepts, which allows the comparison of latent mean differences between the countries. Since there was support for *metric* invariance and the majority of the item intercepts were invariant across countries, we are confident that any estimated latent mean differences across the countries are reliably (i.e., they are based on many cross-culturally comparable items) (Chen 2008; Steenkamp and Baumgartner 1998). That is, the pattern of latent mean differences is a terse summary of the observed means across the countries. Nevertheless, interpretation of the latent mean comparison should be done with caution. Using Model 5 (MG5) as the baseline we tested gender measurement invariance over the five countries (10 groups: dividing the five country samples into separate boy and girl groups (i.e., 2 genders \times 5 countries). The results show support for *configural* (MG7), *metric* (MG8) and *scalar* (MG9) invariance (See Additional file 3: Table S3). The support for the full measurement invariance provides a good argument for comparison of means and other scores for gender.

Reliabilities of the constructs

In response to our third research question, TIMSS 2011 reliability estimates were computed for each country. We incorporated the correlated uniquenesses (method effect) associated with the negatively worded items in the computation of the composite reliabilities (ω). Overall sample reliabilities (α : 0.620–0.780; ω : 0.605–0.786) were generally

¹ Intercepts of Items [BSBM16A BSBM16F BSBM16D BSBM16E] on the students' confidence in mathematics variate, [BSBM14C BSBM14B] on the students like mathematics variate, [BSBM15D] on the teacher responsiveness variate, and items [BSBM16K BSBM16N] on the students' value mathematics variate were freely estimated.

acceptable for α estimates but some of the constructs were far below the acceptable limit for the ω estimates. Individual nation reliabilities for some scales reached the desirable standard of 0.800, but some also fell below the acceptable value of 0.600 (Hair et al. 2010). The teacher responsiveness constructs for Ghana and Morocco were far below the acceptable limit. The mathematics confidence scale was particularly worrisome because here ω was unacceptably low for all the countries. Moreover, ω was unacceptably low in Ghana for the like mathematics construct. This may indicate a possible substantial error of measurement and/or limited true individual differences and problems of translation associated with the definition of the constructs.

Moreover, there is the possibility that the lower reliabilities may attenuate the validity of any interpretations based on manifest scale scores, and weaken statistical power, as well as the effect sizes (Raykov 2012; Schmitt 1996; Marsh et al. 2013). To deal with this situation Cole and Preacher (2013) advise researchers to base any comparisons on latent-variable models that account for unreliability, biasedness, and measurement errors.

To further ascertain whether the model without the method effect due to the negatively corrected items was poorly specified, composite reliabilities with the correlated uniquenesses (see Additional file 4: Table S4) and without the correlated uniquenesses (not reported) were compared. The results showed that the reliability *seems* to be high initially as is the case for α , but once the model was properly specified its poor reliability became evident (the ω estimates). Without accounting for the correlated uniquenesses, the common factors would have to “pick up the slack” and the reliability was larger than it really was (K. J. Preacher, personal communication, February 18, 2014).

Latent mean differences in achievement and the motivational constructs

After imposing invariance constraints on the measurement intercepts, the entire mean structure for the factor model across groups can be identified by fixing the factor mean to zero in one group and freely estimating the factor means in all the other groups. If these values are positive, we interpret them as indicating that the compared groups have higher latent mean values than the reference group and vice versa. For the purposes of these analyses, the Ghanaian sample which seems to be the “most deviant” of the countries in terms of the psychometric properties was chosen as the reference group; as such, the latent means for the four constructs were fixed at zero in the Ghanaian sample so that the size and direction of differences in all the remaining four countries could be evaluated in relationship to the Ghanaian sample (see Additional file 5: Table S5).

To illustrate the reasoning for the mean difference comparison already discussed (Additional file 5: Table S5), when the latent mean of student mathematics confidence was fixed at zero in the Ghanaian sample a standardized latent mean value of -0.461 was found in the Botswana sample. Thus, students are less confident of their mathematics skills in Botswana than in Ghana. These are standardized effects sizes and the differences between the countries are in standard deviation units. All the other countries' estimates for mathematics confidence were statistically significant and negative, an indication that confidence in mathematics in all the four countries are statistically lower than in Ghana. For the like mathematics, value mathematics, and teacher responsiveness constructs, all the countries had lower statistically significant values. Thus, students in all other countries place less value on mathematics, a lower liking for mathematics and

think their teachers are less responsive than Ghanaian students. With parental involvement, there was statistically significantly higher parental involvement in South African than in Ghana and it was lower in Botswana, Morocco, and Tunisia than in Ghana.

All the achievement scores were standardized (mean = 0, $SD = 1$), as such a positive achievement score indicates better than average mathematics achievement across the five countries and negative scores reflect lower than average achievement. Mathematics achievement scores were significantly positive for Botswana (+0.293 SD above the mean) and Tunisia (+0.677 SD above the mean) and significantly negative for Ghana (−0.497 SD below the mean) and South Africa (−0.254 SD below the mean). Indeed, it is only in Morocco that mathematics achievement is not statistically significantly below the five country mean, however the score is just below the mean achievement of the five countries.

Students' long-term educational aspirations (LEA in Additional file 4: Table S4) were higher in Ghana (+0.050 SD below the mean), and lower in Botswana (−0.081 SD below the mean) and Morocco (−0.235 SD below the mean). There were no differences in students' reported long-term educational aspirations in South Africa and Tunisia.

Motivational constructs, achievement, and background variables

Model 6 (M6) was used as the basis for the analysis of how the three background variables (gender, parent education, and student's long term educational aspirations) and students' mathematics achievements were related to the motivational constructs. M6 contained the 3 background variables and students' achievements variate as *correlates* (i.e., treated as covariates in Mplus). This approach is called the multiple-indicator-multiple-indicator cause (MIMIC) (Kaplan 2000) and is very similar to regression analysis but in the CFA framework measurement errors are controlled for bias using latent constructs (Mayer et al. 2014).

The range and direction of the correlations amongst the motivational constructs were within the expected ranges and directions seen in previous studies (e.g. Marsh et al. 2013). As with our a priori prediction, the highest correlations were recorded for the mathematics confidence and like mathematics scales in all countries ($r_s = 0.633$ – 0.787 , $p < 0.001$: Additional file 5: Table S5). The correlations amongst the motivational constructs were higher in Ghana than in the other countries. There was a systematic cross-cultural universal pattern for the relationship between mathematics confidence, like mathematics and value of mathematics ($r_s = 0.233$ – 0.787 , $p < 0.001$). The relationship between teacher responsiveness and students' motivational belief was systematically from low to high and universal across the countries ($r_s = 0.246$ – 0.753 , $p < 0.001$). The relationship between parental involvement and the other motivational construct were positive and universal across the countries; however, the magnitude of the relationships was culturally specific. For instance, in Tunisia, Botswana and South Africa the relationship between parental involvement and teacher responsiveness was the strongest, whereas in Ghana and Morocco the relationship between parental involvement and confidence in mathematics was the strongest.

On the relationship between the motivation construct and achievement, there was a positive relationship between parental involvement and achievement in Ghana and Morocco but a negative correlation in Botswana and South Africa. No statistically

significant relationship was found between parental involvement and achievement in Tunisia. Achievement correlated strongly with the value of mathematics in Botswana to very low in Tunisia ($Mdn r = 0.217$), but correlations with liking for mathematics were smaller ($Mdn r = 0.235$). Students' confidence in mathematics correlated strongly with achievement in Tunisia to very low in South Africa ($Mdn r = 0.211$), but the relationships between achievement and teacher responsiveness ($Mdn r = 0.110$) were very small and was not statistically significant in South Africa and Tunisia. For Ghana, Botswana and South Africa, in relation to achievement, the highest relationship was between the value for mathematics and achievement whereas in the two Arab countries (Morocco and Tunisia) the highest relationships were for confidence in mathematics. There was a statistically significantly negative correlation between parental involvement and mathematics achievement in South Africa and Botswana, whereas a positive relationship was seen in Morocco and Ghana. No relationship was seen in Tunisia. The pattern of relationship between the motivational constructs and achievement was universal for value of mathematics and confidence in mathematics whereas it was cross-culturally specific for the others.

Students' long-term educational aspirations (LEA in Additional file 5: Table S5) were positively related to all the motivational constructs in Ghana Morocco, and Tunisia. In Botswana, there was a statistically significantly positive relationship between students' long-term educational aspirations and all the motivational constructs except between parental involvement. In South Africa students' long-term educational aspirations was not statistically related to their confidence in mathematics and parental involvement. The relationship between students' long term educational aspiration and achievement ($Mdn r = 0.444$) was stronger than any other measure in all the countries.

Parental education correlated significantly with mathematics achievement ($Mdn r = 0.287$) in all the countries. The relationship between parental education and the motivational measures were culturally varying. For instance, there was a negative relationship between parental education and student liking for mathematics in Botswana, whereas a positive relationship was found in Morocco and no relationship found in Ghana, South Africa and Tunisia. There was a cross-culturally universal finding in respect to the relationship between parental education and performance teacher responsiveness, as well as between parental education and teacher responsiveness although the latter relationship was not statistically significant in all the countries.

Gender difference

As discussed above, two of the countries (Morocco and Tunisia) within the present sample have single-sex schools as their main school type. Because of this, gender differences between the motivational construct and achievement were analyzed and shown as correlational measures in Additional file 5: Table S5. A positive correlation indicates that males have higher positive scores than females. Across all the countries, gender differences were small to moderate and favored boys in the motivational constructs and girls in the background variables. For the five motivational constructs across the five countries, there were seven statistically significant differences favoring girls and ten statistically significant differences favoring boys. The largest differences are for the mathematics confidence construct ($Mdn r = +0.134$). Boys outperform girls in mathematics

achievement in Ghana and Tunisia, whereas girls outperformed boys in Botswana. There was no difference in mathematics achievement between boys and girls in Morocco and South Africa. Girls aspire to higher education ($Mdn r = -0.084$) more than boys. These differences are consistent across all countries except for Ghana where boys tended to aspire to higher education more than girls. Long-term educational aspiration levels were higher for girls in Morocco, although not statically significant. There was no statistically significant difference between parental education (educational levels of male and female guardian) in all the countries except in Ghana where girls' parents tended to have greater levels of higher education than did the parents of boys.

Parental involvement favors girls in South Africa and Botswana, but there were no differences in Ghana, Morocco and Tunisia. The most distinct country was Ghana where three out of the five motivational constructs favored boys. This is clear indication of gender stereotypic differences in the motivational constructs favoring boys in Ghana and in the Arab countries, whereas they favor girls in South Africa and Botswana.

In summary, gender differences on motivation and achievement favor girls in the two South Africa countries (Botswana, South Africa), and boys in Ghana and the Arab countries. Moreover, girls had higher educational aspirations in all the countries except for Ghana where boys were found to have higher educational aspirations. Mathematics confidence was higher for boys in all countries whereas girls place much value on mathematics, although there were no gender differences regarding the value of mathematics in Ghana and Tunisia.

Discussion

Our study is a substantive-methodological synergy (Marsh and Hau 2007) of the potential importance for theory and practice in mathematics-related affect. Specifically, we applied structural equation modeling to examine the factorial structure of the TIMSS 2011 motivational constructs, using one of the strongest datasets in educational research to address the psychometric properties, as well as a more applied investigation into the relationships between the TIMSS 2011 motivational constructs, mathematics achievement and other background data in an African context. We substantively tailored our investigation to a less research population using the strongest available nationally representative data set and applied substantively evolving models that are methodologically robust. Importantly, the methodological focus is to address an important limitation associated with TIMSS 2011 motivational constructs for countries from less developed economics (the sample in this study) where the reliability and the validity of the constructs are often problematic. Ironically, the current study was unique in that it formally evaluated measurement invariance across five less researched educational systems/nations/cultures and also addressed other psychometric issues normally not discussed in large scale assessments.

After establishing a substantive multidimensional motivational measure, differences between the countries for these motivational constructs were investigated across the five countries. This also stems from the fact that research and theory that integrate cross-cultural views are crucial to the establishment of more useful and universal theories (e.g., Van de Vijver 2000). Furthermore, there is a paucity of cross-cultural studies on

TIMSS motivational constructs and achievement in the African context. In the following, we summarize and discuss our results in more detail.

Method effect, reliability and measurement invariance of TIMSS motivational constructs

We first identified the best model to account for the multidimensional structure of the TIMSS 2011 measures. Consistent with our a priori prediction, there was the need to explicitly model the method effects associated with the use of negatively worded items in the TIMSS 2011 motivation measures as substantial method effects were found to be associated with the negatively worded items in the mathematics confidence and like mathematics constructs. The CFA model failed when method effects (correlated uniquenesses) associated with the negatively worded items were not accounted for. The need to include the correlated uniquenesses associated with negatively worded items was also justified by the composite reliabilities estimates. Without accounting for the correlated uniquenesses, the common factors “pick up the slack” and the reliability estimates were larger than they really were. Thus, the reliability estimates were higher when the method effect associated with the negatively worded items was not taken into account. Results for the composite (ω) and Cronbach’s alpha (α) reliabilities estimates were roughly the same for the construct without negatively worded items but yet the α values still slightly underestimated the parental involvement, teacher responsiveness and the students value mathematics constructs. The α overestimated the construct with the negatively worded effect (i.e., the mathematics confidence and the like mathematics constructs). Thus this and other research (e.g., Brown 2015; Geldhof et al. 2014) has shown that Cronbach’s alpha is not a dependable estimate of scale reliability and under certain conditions, such as when there are correlated errors among the items in the construct, Cronbach’s alpha estimates may exceed composite reliability estimates (Bentler 2009; Green and Yang 2009; Peterson and Kim 2013; Raykov 2001; Yang and Green 2011). Comparing the change in the composite reliabilities with and without the method effect, it was evident that the method effect was stronger in the responses to the mathematics confidence variate than the like mathematics variate. The reliabilities were below the acceptable limit for the teacher responsiveness construct for all the countries. However, the like mathematics and *confidence* in mathematics constructs were remarkably low in all the countries.

The findings from the model fit and reliability estimates support the claim in the literature that including negatively worded items in a construct can influence the validity and reliability of the scale (Roszkowski and Soven 2010; Woods 2006; Raykov 2001). The results indicate that method effects may not only obscure the underlying structure of these scales but also can possibly bias the outcomes. These findings support the need for control as discussed in the literature. A further look at individual country factor loadings (not reported), individual country method effect indices (Additional file 2: Table S2: exhibits A and B), and the reliabilities (composite reliability) indicated a pattern consistent with individual country achievement in TIMSS 2011. Moreover, it was realized that in the countries where the method effect associated with the negatively worded items was strongest, the achievements of the students were amongst the lowest performing countries. Notably, countries where the method effects were the strongest tend to be lower achieving countries. This study, as well as that of Hooper et al. (2013), seems to indicate that the method effect associated with negatively worded items appears to

diminish with increased mathematics ability (see also Metsämuuronen 2012a). One plausible reason for the link between low performance nations and the method effect associated with negatively worded items is that the reading ability of the students in these low performing countries may be too inadequate for them to fully comprehend and understand the negative worded items. However, the method effect associated with the negatively worded items appears to have an effect in all the countries, but the effect differs across nations. The current study, as well as that of Schmitt and Allik (2005), supports the findings that the method effect associated with negatively worded items are cross culturally specific. However, the outcome of the method effect indicates that differences in responses to negatively worded items are worth studying in their own right, and not just as substantively irrelevant artefact/noise that are theoretically uninteresting and need to be eliminated (cf. Marsh 1996).

The second step of analyses utilized multiple group CFAs to evaluate invariance of model parameters across the five educational systems/nations/cultures. In that respect our focus will be on supporting evidence for Model 2—testing if the factor loadings were equal. The findings indicated that responses to the instruments were the same across the five educational systems/nations/cultures and gender. The support for *configural*, *metric* and *scalar (partial)* invariance across the five educational systems/cultures suggests that further mean comparisons within the motivational constructs can be interpreted as representing the underlying mean differences in the data set. Although some item intercept measurement differences across groups were found, these non-invariant parameters were relatively too small to influence any latent mean comparison across the five educational systems/nations/cultures (Steenkamp and Baumgartner 1998; Chen 2008).

Ironically, there was strong support for *metric invariance*, although there were substantial differences in the construct reliabilities for some of the measures. Additionally, although they were significant, all the factor loadings of the negatively worded items were much lower on each scale (Additional file 1: Table S1). These findings support Chiu's (2008) argument that "items that are negatively worded appear to be unreliable in cross-cultural studies" (p. 251). The present findings are possible indications of substantial errors in measurement, limited true individual differences, participants responding to the negatively worded items differently in each country and problems of translation associated with the definition of the constructs. They also indicate a lack of support for strict measurement invariance (i.e., invariance of item uniqueness as well as factor loadings and intercepts). Strict measurement invariance was not a focus of this study so it was not formally tested.

Differences in achievement and the motivational constructs

The finding of the latent mean differences on the motivation constructs and mathematics achievement was similar to what Marsh et al. (2013) describe as *perplexing* and Shen and Tam (2008) as *paradoxical*. In particular, Ghana was the lowest performing country but its students had the highest values for all the five motivational constructs. The results although *perplexing* and *paradoxical* support other large scale surveys studies where students motivational constructs (e.g., self-concept) and achievement correlated positively at the individual level but negatively at the country level (Marsh and Hau 2004; Marsh et al. 2013; Shen and Tam 2008; Zhu and Leung 2011).

Several explanations have been put forward to explain this discrepancy. The first is based on the frame of reference effect based on the self-concept theory as discussed by Marsh and his colleagues (e.g., Marsh and Hau 2004; Marsh 2007). Self-concept beliefs have been found to be highly influenced by a *frame of reference effect*—the context or standards against which people judge their own accomplishments and failures: for instance, students' responses to a self-belief construct (e.g., self-confidence) will be based on comparisons with other students in the same school or class rather than students in another country (*ibid*). When students report their *self-beliefs*, they normally "... use normative judgments about their ability and social comparison processes with reference to their peers, but also internal comparisons of their performance in one academic domain relative to other academic domains" (Parker et al. 2014, p. 32; see also Marsh 2007). Therefore, the Ghanaian students' self-belief responses were based on comparisons with their peers in the same school or class or different schools in Ghana rather than students from other African countries.

Our findings in relation to gender were very similar within the two Arab countries with similar school systems (single-sex schools) and cultures, as well as between the two South Africa countries, thus partly supporting the argument of the *frame of reference effect*. This is quite understandable since the survey asks students to make subjective assessments about issues such as how enjoyable, difficult or boring mathematics is in relation to their classmates, while at the same time students perceive their attitudes and behaviors within a frame of reference shaped by their schools and cultures. The cultural factors can profoundly influence the way in which these responses are decided. Thus, the relationship between these constructs can change dramatically depending on how the comparisons (for students with different social, educational/academic or cultural background) are made.

The second argument is that such findings may be artefacts of "low academic expectations and standards in low performing countries and high academic expectations and standards in high performing countries" (Shen and Pedulla 2000, p. 237). This proposal is further supported by other studies (e.g., Chiu 2012; Marsh et al. 2013; Shen and Tam 2008). These studies assume that the curricula countries use have an impact on students' achievements. That is, nations with strong mathematics curricula or high-level mathematics classes tend to have high mathematics achievers while at the same time their students perceive mathematics as difficult, have low confidence, and dislike mathematics, therefore reporting more negative mathematics attitudes (see also Papanastasiou 2002). Thus, it is possible that students in Ghana have a not very challenging mathematics curriculum, resulting in lower mathematics performance compared to the other countries.

The analyses indicated significant country differences in the relationship between motivation and achievement. Whereas findings from other studies indicate a strong relationship between self-concept and achievement (e.g., Marsh et al. 2013), this study supported as well as refuted those studies. The relationship between achievement and value was the strongest in three out of the five countries supporting the expectancy-value model's expectations that students' competency beliefs and task-value are positively related (e.g., Eccles and Wigfield 2002; Eccles[Parsons] et al. 1983; Jacobs and Eccles 2000; Wigfield and Eccles 2000). In summary, nearly all of the motivational constructs were significantly related to the students' mathematics scores for all five countries. There

were a few consistent cross-cultural differences in the pattern of the relations. As is evident in Additional file 5: Table S5, the South African students (i.e., South Africa and Botswana) and the Ghanaian students appeared to be more aware of the importance of mathematics than students in the Arab countries, but this did not translate into achievement. On the relationship between parental involvement and achievement, the findings support the mixed conclusions in the literature. In Ghana and Morocco, the relationship was positive (i.e., Campbell and Mandel 1990; Abu-Hilal 2001; Epstein 2008; Epstein 2010; Fan and Chen 2001; Ice and Hoover-Dempsey 2011; Jeynes 2003, 2007), whereas in Botswana and South Africa the relationship was negative (i.e., Desimone 1999; Fan and Chen 2001), and no relationship was found in Tunisia. (i.e., Topor et al. 2010; Reynolds 1992). This pattern of mixed findings support the argument put forward by Ice and Hoover-Dempsey (2011, p. 346) that student achievement may increase as a result of parental involvement (i.e., positive correlation)—in the case of Ghana and Morocco—grade eight students, *and* that parental involvement may increase as a results of students' poor performance (i.e. negative correlations as parents become more involved in order to support lagging students)—in the case of Botswana and South Africa—ninth grade students (see also, Ames et al. 1993; Gonzalez-DeHass et al. 2005). These mixed findings also indicate that the relationship between parental involvement and achievement is culturally specific. Is it also possible that social and cultural norms for education in these countries may influence these outcomes?

A thorough look at the correlations between the constructs indicates a universal as well as cultural specific patterns. For instance, the relationships between parental involvement and all the constructs were country specific whereas the relationships between the other constructs showed a universal pattern. This is perhaps not surprising; in light of the fact that parental involvement in children's education is a complex phenomenon that often transcends the geographical boundaries of home and school (Ice and Hoover-Dempsey 2011). Another important finding was the highly positive correlations between the teacher responsiveness variate and the students' mathematics confidence, value of mathematics and like mathematics variates. These findings support other studies, which have indicated the significant role of the teacher in students' self-beliefs (e.g., Ice and Hoover-Dempsey 2011).

As far as student gender and mathematics achievement is concerned, the findings support the varying conclusions in the literature; for instance in Ghana and Tunisia boys outperform girls (e.g., Ma 2008; Marks 2008; Cheema and Galluzzo 2013; Guiso et al. 2008; Fryer and Levitt 2010), in Botswana girls outperform boys (e.g., Ma 2008; Vermeer et al. 2000; Hyde et al. 1990), and in Morocco and South Africa there are no gender differences (e.g., Cheema and Galluzzo 2013; Else-Quest et al. 2010; Chen 2002). In respect to the Arabic countries (Morocco and Tunisia), the findings in this study are contrary to those that have been found for girls in the Middle Eastern Arab countries. One plausible explanation as to why girls did not perform better in Arab Africa than their counterparts may be that there are less strict rules controlling girls than those found in the Middle Eastern Arab countries whereby boys have more freedom of movement than girls and therefore girls spend more time focusing on schoolwork than boys (Abu-Hilal 2001).

On parental involvement and teacher responsiveness, there were mixed findings. Whereas girls' in Botswana and South Africa enjoy higher parental involvement, in the

other three countries parental involvement was not a function of gender. On aspirations, Ghanaian boys claimed they had higher aspirations, whereas girls indicated higher aspirations in Botswana, Morocco, South Africa and Tunisia. There were varying outcomes on the relationship between affect and gender. The varying gender differences across nations under consideration support the cultural universality as well as the cultural specificity of gender as a moderator in these relations (Hyde and Mertz 2009; Forgasz et al. 2015). For instance, the relationship between gender and confidence in mathematics was culturally universal with boys reporting higher mathematics confidence in all the countries, whereas for mathematics achievement, teacher responsiveness, value for mathematics, parental involvement and like mathematics constructs the relationship were culturally specific. The relationship between gender and students' aspirations was culturally specific with boys reporting they would like to pursue higher education in Ghana whereas more girls reporting they had plans to pursue higher education in Botswana, Morocco, South Africa and Tunisia.

Students' educational aspirations—that is, their plans to pursue further studies were highly significantly associated with student achievement than all the variables in the study. This supported the literature (Gil-Flores et al. 2011; Saha, 1994; Sanders et al. 2001; Marjoribanks 2002, 2003a, 2003b) and contradicted studies that listed self-belief constructs as the strongest predictor of academic achievement (Hattie 2009; Stankov and Lee 2014).

However, the relationship between students' plans for future studies and students' achievement was strongest in Ghana, Botswana and South Africa. Plans to pursue further studies also correlated strongly with students' value of mathematics than with the other motivational measures.

There was a statistically significant relationship between parental education and mathematics achievement in all the countries, which supported the literature findings (e.g., Davis-Kean 2005; Haveman and Wolfe 1995; Klebanov et al. 1994). These relationships were cross-culturally universal. The relationship between parental education and students' motivational belief was culturally specific being negative or positive in some countries and no relationship in others. The relationship between parental education and teacher responsiveness was cross-culturally universal although not statistically significant in all the countries.

Limitations of the present study

Measurement invariance, especially scalar invariance is difficult to achieve in cross-cultural studies (Chiu 2012; Van de Vijver 2000), which was also the case in this study. This shows that using manifest scale scores (e.g., aggregate data) from different countries or cultures for any statistical analysis in cross-cultural research should be done with caution (Chiu 2012). Parental involvement has a multidimensional construct (Epstein 2010) but in this study it was measured using a uni-dimensional measure of parental involvement. Researchers should separate the various aspect of parental involvement variables in order to determine their unique influence on students' motivation and achievement. Another limitation is that all the data were from self-reports and thus subject to social desirability biases. Moreover, the present study failed to address the issue of reciprocal/causal relations between most of the motivational constructs (e.g., self-confidence) and

achievements because TIMSS studies are snapshots of ongoing dynamic processes. Such causal relationships are best investigated using longitudinal data sets (see, XXXPME 2015). Our last limitation was sample heterogeneity, which had a significant impact on the construct reliability.

Conclusions

It needs to be recognized that the reality of comparing countries is a complex multi-dimensional issue well beyond TIMSS and other large scale assessment (Goldstein, 2004). Based on TIMSS 2011, we first investigated the psychometric properties of the TIMSS 2011 motivational constructs. The multidimensionality of the motivation measure was validated. However, there were systematic psychometric problems with some of the constructs. For instance, there was a substantive method effect associated with the negatively worded items in the mathematics self-confidence and like mathematics constructs. The findings in these studies hint to a strong relationship between country-level construct measurement fit due to the method effect attributable to negatively worded items and country-level achievement. Moreover, there was some evidence that negatively worded items can attenuate the reliability and validity of a measure (Roszkowski and Soven 2010; Raykov 2001). The factor loadings of the negatively worded items and reliability estimates of the construct with negatively worded items were considerably low.

We advocate the use of composite reliabilities in estimating construct reliability in cross-cultural research because it is able to give the “true” reliabilities by compensating for any method effect associated with the constructs. Moreover, we support the use of robust methods like structural equation modeling in cross-cultural research to account for bias parameter estimates, measurement error and possible method effects. However, these strategies may not be realistic for many researchers if they are not familiar or conversant with SEM approaches and in particular the models discussed in this article.

These results may have a broad influence on researchers using both negatively and positively worded items in mathematics-related affects research and surveys in general. The authors believe that negatively worded items are a “cognitive nuisance” for low ability students cross-culturally and worth studying in their own right, and not just irrelevant noise that needs to be eliminated. It is recommended that researchers should be aware of the potential bias associated with survey measures involving both negatively worded items so as to take the necessary steps to address this bias appropriately. The authors think TIMSS should give more emphasis to positively worded items as proposed by Marsh (1996) and Corwyn (2000). However, this will mean discarding control for the effect of response pattern biases (e.g., acquiescence).

Concerning the association between students’ motivation, achievement, parental education, educational aspiration and gender, the results indicate that the achievement levels and motivation factors such as parental involvement differ across these countries. Whereas the relationship between parental involvement and the other constructs were cultural specific the relationship between like mathematics, value mathematics, confidence in mathematics, and teacher responsiveness were universal. Moreover, parental education, gender and long-term educational aspiration also influence

the students' achievement and motivation. As with previous studies (e.g., Gil-Flores et al. 2011), higher educational aspirations and higher parental education were associated with higher achievement. Moreover, the study found that parental education and students' aspirations were more associated with students' achievement than students' motivation (cf. Marks et al. 2001; Schoon and Parson 2002; Schoon et al. 2004). These findings contradict studies that indicate that self-concept (i.e., mathematics confidence) is the strongest motivational predictor of students' achievement (e.g., Marsh et al. 2008; Marks et al. 2001). The strongest association was between teacher responsiveness and students' self-beliefs. Interestingly, Ghana was the only nation where boys reported higher educational aspirations than girls as well as showing higher motivation levels for all the constructs. Even so achievement levels in Ghana were the lowest of all the countries.

The countries in this study can use our results to fine-tune their education policies on affect and performance. This is because the relationship between affect and performance is of practical importance in many affective enhancement programs as well as educational policy statements throughout the world. This is based on the prelude that an improvement in one affect (e.g., self-concept) will lead to improvements in the other. The findings from this study especially the gender gaps, could help educators and policy makers design curricula to help students of both genders apply their talents and deal with their weaknesses. This study underlines how "cultural differences challenge mainstream theoretical notions about the nature of people and force us to rethink our basic theories of personality, perception, cognition, emotion, development, social psychology, and the like, in fundamental and profound ways" (Matsumoto 2001, p. 107–108). Therefore, the effect of cultural norms, values, and practices should be factored into cross-cultural studies of academic motivation (e.g., Ng 2003). Failing to do this, may lead to inferences that are not valid because too much reliance is being placed on studies that are completely based on western theories or concepts (Zhu and Leung 2011). However, the differences between the latent means of the motivational constructs, and achievement should be interpreted with caution, because of the low reliabilities of some of the construct, the method effect, and low performance of students in the TIMSS 2011 mathematics achievement in the five participating countries. Finally, we examined the less-studied population of 8–9th-grade students using a nationally representative sample that allows us to generalise the results to a larger population.

Additional files

Additional file 1: Table S1. A Priori Factor Structure relating to TIMSS2011 Motivational Items to Latent Factors.

Additional file 2: Table S2. Goodness-of-Fit Summary: Country Influence by Method effect.

Additional file 3: Table S3. Summary of Goodness of Fit Statistics for Group Models.

Additional file 4: Table S4. Reliability of the TIMSS motivational constructs used in this study.

Additional file 5: Table S5. Estimated latent means and correlations.

Authors' contributions

EAB drafted the manuscript. MSH as a doctoral supervisor shared his expertise during the preparation and the development of the manuscript. The work as a whole is an extensive collaboration and discussion between EAB and MSH. Both authors read and approved the final manuscript.

Authors' information

Emmanuel Adu-tutu Bofah is a PhD candidate at the University of Helsinki, Finland, under the supervision of Professor Markku S. Hannula. He obtained an MA in Educational Science from the University of Turku and Master of Social Sciences from University of Helsinki; both in Finland. His research interest is mathematics affect and its relationships to students' achievement. He has also written methodology papers on cross-cultural research on affect and achievement. With Markku S. Hannula, he has published on methodological aspects of cross-cultural studies for both Sage and Springer.

Markku S. Hannula is a professor of mathematics education and the Director of the Research Centre for Mathematics and Science Education (RCMSE) in the Department of Teacher Education at the University of Helsinki. His main research interests are the affective domain and problem solving in mathematics and he uses both qualitative and quantitative methods. He is an editor for the international journal *Nomad—Nordic Studies in Mathematics Education* and serves currently as a board member for the European Society for Research in Mathematics Education and for the Nordic Society for Research in Mathematics Education.

Competing interests

The author(s) declare that they have no competing interests.

Received: 11 June 2014 Accepted: 26 November 2015

Published online: 21 December 2015

References

- Abu-Hilal, M. M. (2001). Correlates of achievement in the United Arab Emirates: A sociocultural study. In D. M. McInerney & S. Van Etten (Eds.), *Research on Sociocultural Influences on Motivation and Learning* (Vol. 1, pp. 205–230). Greenwich, CT: Information Age.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545–557. doi:10.1037/0021-843X.112.4.545.
- Ames, C., Khoju, M., & Watkins, T. (1993). *Parent involvement: The relationship between school-to-home communication and parents' perceptions and beliefs* (Report No. 15). Urbana, IL: ERIC Document Service No. ED362271, Center on Families, Communities, Schools, and Children's Learning, Illinois University.
- Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching, 36*(6), 719–747.
- Areepattamannil, S., & Freeman, J. G. (2008). Academic achievement, academic self-concept, and academic motivation of immigrant adolescents in the greater Toronto area secondary schools. *Journal of Advanced Academics, 19*(4), 700–743.
- Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus (Version 2). Mplus Technical appendices*. Los Angeles, CA: Muthén & Muthén.
- Ayalon, H., & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social science research, 42*, 432–445.
- Bagozzi, R. P. (1993). Assessing construct validity in personality research: Applications to measures of self-esteem. *Journal of Research in Personality, 27*(1), 49–87. doi:10.1006/jrpe.1993.1005.
- Belcher, C., Frey, A., & Yankeelov, P. (2006). The effects of single-sex classrooms on classroom environment, self-esteem, and standardized test scores. *School Social Work Journal, 31*(1), 61–75.
- Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement, 22*(3), 231–240. doi:10.1111/j.1745-3984.1985.tb01061.x.
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137–143. doi:10.1007/s11336-008-9100-1.
- Bentler, P. M., & Bonnett, D. G. (1980). Significance test and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.
- Bhanot, R. T., & Jovanovic, J. (2009). The links between parent behaviours and boys' and girls' science achievement Beliefs. *Applied Development Science, 13*(1), 42–59.
- Bofah, E. A. (2015). Reciprocal determinism between students' maths self-concept and achievement in an African context. In *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education*. Prague, Czech Republic. <https://hal.archives-ouvertes.fr/>
- Bofah, E. A., & Hannula, M. S. (2014). Structural equation modelling: Testing for the factorial validity, replication and measurement invariance of students' views on mathematics. In *SAGE Research Method Cases*. London: SAGE Publications, Ltd. doi:<http://dx.doi.org/10.4135/978144627305014529518>
- Bofah, E. A., & Hannula, M. S. (2015). Studying the factorial structure of Ghanaian twelfth-grade students' views on mathematics. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education: Exploring a mosaic of relationships and interactions* (pp. 355–381). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-06808-4_18
- Brady, K. L., & Eisler, R. M. (1999). Sex and gender in the college classroom: A quantitative analysis of faculty-student interactions and perceptions. *Journal of Educational Psychology, 91*(1), 127–145. doi:10.1037/0022-0663.91.1.127.
- Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Testing structural equation models* (pp. 136–163). Newbury Park CA: Sage.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, NY: Guilford Press.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

- Campbell, J. R., & Mandel, F. (1990). Connecting math achievement to parental influences. *Contemporary Educational Psychology*, 15(1), 64–74.
- Canada, K., & Pringle, R. (1995). The role of gender in college classroom interactions: A social context approach. *Sociology of Education*, 68(3), 161–186.
- Cheema, J. R., & Galluzzo, G. (2013). Analyzing the gender gap in Math achievement: Evidence from a large-scale US sample. *Research in Education*, 90(1), 98–112.
- Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development*, 66(4), 1215–1234.
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling*, 14(3), 464–504.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of personality and social psychology*, 95(5), 1005.
- Chen, P. P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, 14(1), 79–92. doi:10.1016/j.lindif.2003.08.003.
- Cheung, G. W., & Rensvold, R. B. (1998). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science Review*, 6, 93–110.
- Chiu, M.-S. (2008). Achievements and self-concepts in a comparison of math and science: exploring the internal/external frame of reference model across 28 countries. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14(3), 235–254. doi:10.1080/13803610802048858.
- 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. doi:10.1007/s10763-011-9309-9.
- Cokley, K. O., Bernard, N., Cunningham, D., & Motoike, J. (2001). A psychometric investigation of the academic motivation scale using a United States sample. *Measurement & Evaluation in Counseling & Development*, 34(2), 109–119.
- Cole, D. A., & Preacher, K. J. (2013). *Manifest variable path analysis: potentially serious and misleading consequences due to uncorrected measurement error*. Psychological Methods: Advance online publication. doi:10.1037/a0033805.
- Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology*, 98(4), 665–689. doi:10.1037/0022-0663.98.4.665.
- Connor, C. M., Son, S.-H., Hindman, A. H., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*, 43(4), 343–375. doi:10.1016/j.jsp.2005.06.001.
- Cornbleth, C., & Korth, W. (1980). Teacher perceptions and teacher-student interaction in integrated classrooms. *The Journal of Experimental Educational*, 48(4), 259–263.
- Corwyn, R. F. (2000). The Factor Structure of Global Self-Esteem among Adolescents and Adults. *Journal of Research in Personality*, 34, 357–379. doi:10.1006/jrpe.2000.2291.
- Davis-Kean, P. E., Jacobs, J., Bleeker, M., Eccles, J. S., Melanchuk, O. (2007). How dads influence their daughters' interest in mathematics [University of Michigan]. *ScienceDaily*. Retrieved February 2, 2014 from www.sciencedaily.com/releases/2007/06/070624143002.htm.
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19(2), 294. doi:10.1037/0893-3200.19.2.294.
- Desimone, L. (1999). Linking parent involvement with student achievement: Do race and income matter? *Journal of Educational Research*, 93(1), 11–30.
- Distefano, C., & Motl, R. W. (2009). Methodological artifact or substance? Examinations of wording effects associated with negatively worded items. In T. Teo & M. S. Khine (Eds.), *Structural Equation Modeling in Educational Research* (pp. 59–77). Boston: Sense Publishers.
- Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, 46, 183–201. doi:10.1111/j.1540-4560.1990.tb01929.x.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the achiever: The structure of adolescents' academic achievement related-beliefs and self-perceptions. *Personality and Social Psychology Bulletin*, 21(3), 215–225.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, 51(3), 109–132.
- Eccles [Parsons], J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J., Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: W. H. Freeman.
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A Meta-Analysis. *Psychological Bulletin*, 136(1), 103–127. doi:10.1037/a0018053.
- Epstein, J. L. (2008). Improving family and community involvement in secondary schools. *Education Digest*, 73(6), 9–12.
- Epstein, J. L. (2010). School/ family/ community partnerships: Caring for the children we share. *Phi Delta Kappan*, 92(3), 81–96.
- Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, 13(1), 1–22.
- Foy, P., Arora, A., & Stanco, G. M. (2013). *TIMSS 2011 User Guide for the International Database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Forgasz, H., Leder, G., Mittelberg, D., Tan, H., Murimo, A. (2015). Affect and gender. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education* (pp. 245–268). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-06808-4_12.
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507–537. doi:10.1111/j.1532-7795.2010.00645.x.
- Fryer, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, American Economic Association, 2(2), 210–240.

- Furinghetti, F., & Pehkonen, E. (2002). Rethinking characterizations of beliefs. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *In beliefs: A hidden variable in mathematics education?* (pp. 39–58). Dordrecht, The Netherlands: Kluwer.
- Garg, R., Kauppi, C., Lewko, J., & Urajnik, D. (2002). A structural model of educational aspirations. *Journal of Career Development, 29*(2), 87–108.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91. doi:10.1037/a0032138.
- Gil-Flores, J., Padilla-Carmona, T. M., & Suárez-Ortega, M. (2011). Influence of gender, educational attainment and family environment on the educational aspirations of secondary school students. *Educational Review, 63*(3), 345–363. doi:10.1080/00131911.2011.571763.
- Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice, 11*(3), 319–330.
- Gonzalez-DeHass, A. R., Willems, P. P., & Holbein, M. F. D. (2005). Examining the relationship between parental involvement and student motivation. *Educational Psychology Review, 17*(2), 99–123. doi:10.1007/s10648-005-3949-7.
- Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*(1), 121–135. doi:10.1007/s11336-008-9098-4.
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*(5880), 1164–1165. doi:10.1126/science.1154094.
- Gunderson, E., Ramirez, G., Levine, S., & Beilock, S. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*(3/4), 153–166.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hannula, M. S. (2012). Exploring new dimensions of mathematics-related affect: embodied and social theories. *Research in Mathematics Education, 14*(2), 137–161. doi:10.1080/14794802.2012.694281.
- Hannula, M. S., Bofah, E. A., Tuohilampi, L., & Metsämuuronen, J. (2014). A longitudinal analysis of the relationship between mathematics-related affect and achievement in Finland. In S. Oesterle, P. Liljedahl, C. Nicol, & D. Allan (Eds.), *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (pp. 249–256). Vancouver, Canada, Canada: PME.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child development, 72*(2), 625–638.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses related to achievement*. London: Routledge Taylor and Francis Group.
- Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature, 33*(4), 1829–1878.
- Heine, S. J. (2001). Self as cultural product: an examination of East Asian and North American selves. *Journal of Personality, 69*(6), 881–906. doi:10.1111/1467-6494.696168.
- Herring, M., & Wahler, R. G. (2003). Children's cooperation at school: The comparative influences of teacher responsiveness and the children's home-based behavior. *Journal of Behavioral Education, 12*(2), 119–130.
- Ho, H. Z., Senturk, D., Amy, G. L., Zimmer, J. M., Hong, S., Okamoto, Y., et al. (2000). The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal of Research in Mathematics Education, 31*(3), 362–379.
- Hooper, M., Arora, A., Martin, M. O., and Mullis, I. V. S. (2013). Examining the behavior of "reverse directional" items in the TIMSS 2011 Context Questionnaire Scales. In *IEA International Research Conference*. Singapore: Nanyang.
- Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental involvement in children's education: Why does it make a difference? *Teachers College Record, 97*, 310–331.
- Hoover-Dempsey, K. V., Walker, J. M., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., & Closson, K. (2005). Why do parents become involved? Research findings and implications. *Elementary School Journal, 106*(2), 105–130.
- Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling, 10*(3), 435–455. doi:10.1207/S15328007SEM1003_6.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of school psychology, 43*(4), 303–320.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological bulletin, 107*(2), 139.
- Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences of the United States of America, 106*(22), 8801–8807. doi:10.1073/pnas.0901265106.
- Ice, C. L., & Hoover-Dempsey, K. V. (2011). Linking parental motivations for involvement and student proximal achievement outcomes in home schooling and public schooling settings. *Education and Urban Society, 43*(3), 339–369.
- International Association for the Evaluation of Educational Achievement (IEA). (2014). About TIMSS and PIRLS. Retrieved from <http://timssandpirls.bc.edu/>
- IEA IDB. (2014). *IEA International Database (IDB) Analyzer (version 3.1) [Software]*. Herengracht, Amsterdam. Available at <http://www.iea.nl/data.html>.
- Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter? *New directions for child and adolescent development, 106*, 5–21.
- Jacobs, J. E., & Eccles, J. S. (2000). Parents, task values, and real-life achievement-related choices. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 405–439). San Diego, CA: Academic Press.
- Jacobs, J. E., Lanza, S., Osgood, W. D., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509–527.
- Jeynes, W. H. (2003). A meta-analysis: The effects of parental involvement of minority children's academic achievement. *Education and Urban Society, 35*(2), 202–218.

- Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement. *A meta-analysis. Urban Education, 42*(1), 82–110.
- Jones, S. M., & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research, 74*(4), 443–471. doi:10.3102/00346543074004443.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Newbury Park, CA: SAGE.
- Keith, T. Z., Keith, P. B., Troutman, G. C., Bickley, P. G., Trivette, P. S., & Singh, K. (1993). Does parental involvement affect eighth-grade student achievement? Structural analysis of national data. *School Psychology Review, 22*(3), 474–496.
- Kelly, A. (1988). Gender differences in teacher-pupil interactions: A meta-analytic review. *Research in Education, 39*(1), 1–23.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*(1), 165–172. doi:10.1037/0033-2909.112.1.165.
- Kifer, E. (2002). Students' attitudes and perceptions. In D. Robataille & A. Beaton (Eds.), *Secondary analysis of the TIMSS results: A synthesis of current research* (pp. 251–275). Boston, MA: Kluwer Academic.
- Klebanov, P. K., Brooks-Gunn, J., Duncan, G. J. (1994). Does neighborhood and family poverty affect mothers' parenting, mental health, and social support? *Journal of Marriage and the Family, 441*–455.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.
- Leaper, C., Farkas, T., & Brown, C. S. (2012). Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and English. *Journal of youth and adolescence, 41*(3), 268–282.
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*, 355–365.
- Lerkkänen, M.-K., Kiuru, N., Pakarinen, E., Viljaranta, J., Poikkeus, A.-M., Rasku-Puttonen, H., et al. (2012). The role of teaching practices in the development of children's interest in reading and mathematics in kindergarten. *Contemporary Educational Psychology, 37*(4), 266–279. doi:10.1016/j.cedpsych.2011.03.004.
- Leung, K., & Zhang, J. X. (1995). Systemic considerations: Factors facilitating and impeding the development of psychology in developing countries. *International Journal of Psychology, 30*, 693–706.
- Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment, 94*(2), 196–204. doi:10.1080/00223891.2011.645936.
- Liu, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept. *Educational Psychology: An International Journal of Experimental Educational Psychology, 30*(6), 699–712. doi:10.1080/01443410.2010.501102.
- Ma, X. (2008). Within-School Gender gaps in reading, mathematics, and science literacy. *Comparative Education Review, 52*(3), 437–460.
- Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic review. *Educational Psychology Review, 9*(2), 89–120.
- Magazine, S. L., Williams, L. J., & Williams, M. L. (1996). A confirmatory factor analysis examination of reverse coding effects in MEYER and ALLEN'S affective and continuance commitment scales. *Educational and Psychological Measurement, 56*(2), 241–250.
- Marchant, G. J., Paulson, S. E., & Rothlisberg, B. A. (2001). Relations of middle school students' perceptions of family and school contexts with academic achievement. *Psychology in the Schools, 38*(6), 505–519. doi:10.1002/pits.1039.
- Marjoribanks, K. (2002). Family contexts, individual characteristics, proximal settings, and adolescents' aspirations. *Psychological Reports, 91*(3), 769–779. doi:10.2466/pr0.2002.91.3.769.
- Marjoribanks, K. (2003a). Family background, individual and environmental influences, aspirations and young adults' educational attainment: A follow-up study. *Educational Studies, 29*(2–3), 233–242. doi:10.1080/03055690303283.
- Marjoribanks, K. (2003b). Learning environments, family contexts, educational aspirations and attainment: A moderation-mediation model extended. *Learning Environments Research, 6*(3), 247–265. doi:10.1023/A:1027327707647.
- Marks, G., McMillan, J., Hillman, K. (2001). Tertiary entrance performance: The role of student background and school factors. *LSAY Research Reports* (Vol. 22). Longitudinal surveys of Australian youth research report. Retrieved from http://research.acer.edu.au/lsay_research/24.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. doi:10.1037/0033-295X.98.2.224.
- Marks, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education, 34*(1), 89–109.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., et al. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology, 105*(1), 108–128.
- Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of U.S. and Saudi Arabian TIMSS Responses. *Journal of Cross-Cultural Psychology, 45*(5), 777–804. doi:10.1177/0022022113519858.
- Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*(1), 37.
- Marsh, H. W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The Self-Description Questionnaire. *Journal of Educational Psychology, 86*(3), 439–456. doi:10.1037/0022-0663.86.3.439.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of personality and social psychology, 70*(4), 810.
- Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, England: Britain Psychological Society.
- Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56–67.

- Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 32, 151–170. doi:[10.1016/j.cedpsych.2006.10.008](https://doi.org/10.1016/j.cedpsych.2006.10.008).
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, 16(2), 191–225. doi:[10.1080/10705510902751010](https://doi.org/10.1080/10705510902751010).
- Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and bigfish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, 100, 510–524.
- Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011: Sample Design and Implementation*. Chestnut Hill, MA: TIMSS & PIRLS International Study, Boston College.
- Matsumoto, D. (2001). Cross-cultural psychology in the 21st century. In J. S. Halonen & S. F. Davis (Eds.). *The many faces of psychological research in the 21st century* (chap. 5). Retrieved from <http://teachpsych.org/ebooks/faces/index.php>.
- Mau, W. (2000). Educational and vocational aspirations of minority and female students: A longitudinal study. *Journal of Counseling & Development*, 78(2), 186–194.
- Mayer, A., Nagengast, B., Fletcher, J., & Steyer, R. (2014). Analyzing average and conditional effects with multigroup multi-level structural equation models. *Frontiers in Psychology*, 5, 304. doi:[10.3389/fpsyg.2014.00304](https://doi.org/10.3389/fpsyg.2014.00304).
- Mendez, L. M. R., & Crawford, K. M. (2002a). Gender-role stereotyping and career aspirations. *Journal of Secondary Gifted Education*, 13(3), 96–107.
- Mendez, L. M. R., & Crawford, K. M. (2002b). Gender-role stereotyping and career aspirations: A comparison of gifted early adolescent boys and girls. *Journal of Secondary Gifted Education*, 13(3), 96–107.
- Metsämuuronen, J. (2012a). Challenges of the Fennema-Sherman test in the international comparisons. *International Journal of Psychological Studies*, 4(3), 1–22.
- Metsämuuronen, J. (2012b). Comparison of mental structures of eighth-graders in different countries on the basis of Fennema-Sherman Test. *International Journal of Psychological Studies*, 4(4), 1–17.
- Middleton, J. A., & Spanias, P. A. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education*, 30, 65–88. doi:[10.2307/749630](https://doi.org/10.2307/749630).
- Millsap R. E., & Olivera-Aguilar M. (2012). Investigating measurement invariance using confirmatory factor analysis In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380–392). New York: the Guilford Press.
- Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/timss2011/international-results-mathematics.html>.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence*, 20(2), 482–506. doi:[10.1111/j.1532-7795.2010.00644.x](https://doi.org/10.1111/j.1532-7795.2010.00644.x).
- Ndlovu, M., & Mji, A. (2012). Alignment between South African mathematics assessment standards and the TIMSS assessment frameworks. *Pythagoras*, 33(3), doi:[10.4102/pythagoras.v33i3.182](https://doi.org/10.4102/pythagoras.v33i3.182).
- Ng, C. H. (2003). *Re-conceptualizing achievement goals from a cultural perspective*. Auckland, New Zealand: In joint conference of NZARE & AARE.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1–13. doi:[10.1007/BF02289400](https://doi.org/10.1007/BF02289400).
- Numally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- OECD. (2015). PISA 2012 results: The ABC of gender equality in education: Aptitude, behaviour, confidence. Paris: PISA, OECD Publishing. Retrieved March 13, 2015, from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-gender.htm>
- Ormrod, J. E. (2011). *Educational psychology* (7th ed.). New Jersey: Merrill Prentice-Hall.
- Page, R. (1987). Teachers' perceptions of students: A link between classrooms, school cultures, and the social order. *Anthropology & Education Quarterly*, 18(2), 77–99. doi:[10.1525/aeq.1987.18.2.04x0667q](https://doi.org/10.1525/aeq.1987.18.2.04x0667q).
- Papanastasiou, E. (2002). Factors that differentiate mathematics students in Cyprus, Hong Kong, and the USA. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 8(1), 129–146.
- Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2013). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, 34(1), 1–20.
- Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *The Journal of Applied Psychology*, 98(1), 194–198. doi:[10.1037/a0030767](https://doi.org/10.1037/a0030767).
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, 88(5), 879–903. doi:[10.1037/0021-9010.88.5.879](https://doi.org/10.1037/0021-9010.88.5.879).
- Quilty, L. C., Oakman, J. M., & Risko, E. (2009). Correlates of the Rosenberg Self-Esteem scale method effects. *Structural Equation Modeling*, 13(1), 99–117. doi:[10.1207/s15328007sem1301_5](https://doi.org/10.1207/s15328007sem1301_5).
- Raykov, T. (2001). Bias of coefficient a fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, 25(1), 69–76. doi:[10.1177/01466216010251005](https://doi.org/10.1177/01466216010251005).
- Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. H. Hoyle, *Handbook of structural equation modeling* (pp. 472–492). New York, NY: The Guilford.
- Reddy, V., Prinsloo, C., Visser, M., Arends, F., Winnaar, L., Rogers, S., Janse van Rensburg, D., Juan, A., Feza, N., Mthethwa, M. (2014). Highlights from TIMSS 2011: The South African perspective. Retrieved April 19, <http://www.hsrc.ac.za>.
- Reynolds, A. J. (1992). Comparing measures of parental involvement and their effects on academic achievement. *Early Childhood Research Quarterly*, 7(3), 441–462. doi:[10.1016/0885-2006\(92\)90031-S](https://doi.org/10.1016/0885-2006(92)90031-S).
- Roszkowski, M. J., & Soven, M. (2010). Shifting gears: consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113–130. doi:[10.1080/02602930802618344](https://doi.org/10.1080/02602930802618344).

- Rutkowski, L., & Rutkowski, D. (2010). Getting it 'better': the importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies*, 42(3), 411–430.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Saha, L. J. (1994). Aspirations and expectations of students. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (pp. 354–358). Oxford, UK: Pergamon Press.
- Sanders, C. E., Field, T. M., & Diego, M. A. (2001). Adolescents' academic expectations and achievement. *Adolescence*, 36(144), 795–802.
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, 29, 299.
- Schafer, J. L., & Olson, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, 545–571. doi:10.1207/s15327906mbr3304_5.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. doi:10.1037/1082-989X.7.2.147.
- Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem scale in 53 nations: exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, 89, 623–642. doi:10.1037/0022-3514.89.4.623.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychoeducational Assessment*, 8, 350–353.
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9, 367–373. doi:10.1177/014662168500900405.
- Schoon, I., & Parson, S. (2002). Teenage aspirations for future careers and occupational outcomes. *Journal of Vocational Behavior*, 60, 262–288. doi:10.1006/jvbe.2001.1867.
- Schoon, I., Parsons, S., & Sacker, A. (2004). Socioeconomic adversity, educational resilience, and subsequent levels of adult adaptation. *Journal of Adolescent Research*, 19, 383–404. doi:10.1177/0743558403258856.
- Seaton, M., Parker, P., Marsh, H. W., Craven, R. G., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology*, 34(1), 49–72. doi:10.1080/01443410.2013.825232.
- Sharp, R. M. (1995). *Scribble scrabble: Ready-in-a-minute math games*. Blue Ridge Summit, PA: TBA Books.
- Sheldon, S. B. (2002). Parents' social networks and beliefs as predictors of parent involvement. *Elementary School Journal*, 102(4), 301–316.
- Shen, C., & Pedulla, J. J. (2000). The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education: Principles, Policy & Practice*, 7(2), 237–253.
- Shen, C., & Tam, H. P. (2008). The paradoxical relationship between student achievement and self-perception: a cross-national analysis based on three waves of TIMSS data. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14(1), 87–100.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120.
- Simpkins, S. D., & Davis-Kean, P. E. (2005). The intersection between self-concepts and values: Links between beliefs and choices in high school. *New Directions for Child and Adolescent Development*, 110, 31–47. doi:10.1002/cd.148.
- Sy, S. R. (2006). Rethinking parent involvement during the transition to first grade: A focus on Asian American families. *School Community Journal*, 16(1), 107–126.
- Stankov, L., & Lee, J. (2014). Quest for the best non-cognitive predictor of academic achievement. *Educational Psychology*, 34(1), 1–8. doi:10.1080/01443410.2013.858908.
- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–90.
- Topor, D. R., Keane, S. P., Shelton, T. L., & Calkins, S. D. (2010). Parent involvement and student academic performance: A multiple mediational analysis. *Journal of prevention & intervention in the community*, 38(3), 183–197.
- Tomçş, J. M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling*, 20(2), 299–313. doi:10.1080/10705511.2013.769394.
- Tuohilampi, L., Hannula, M., Giacconi, V., Laine, A., Näveri, L. (2013). Comparing the structures of 3rd graders' mathematics-related affect in Chile and Finland. In *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education (CERME8)*. Antalya: ERME.
- UNESCO International Bureau of Education. (2011). World Data on Education: Seventh edition. <http://www.ibe.unesco.org/en/services/online-materials/world-data-on-education/seventh-edition-2010-11.html>. Accessed 28 July 2015
- UNESCO Statistics. (2014). <http://www.uis.unesco.org/Pages/default.aspx>. Accessed on 30 July 2015.
- Urbán, R., Szigei, R., Kökönyei, G., & Demetrovics, Z. (2014). Global self-esteem and method effects: Competing factor structures, longitudinal invariance, and response styles in adolescents. *Behavior Research Methods*, 46(2), 488–498. doi:10.3758/s13428-013-0391-5.
- Van de Vijver, F. J. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31, 33–51.
- Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, 92(2), 308–315. doi:10.1037/0022-0663.92.2.308.
- Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, 75(5), 1556–1574. doi:10.1111/j.1467-8624.2004.00757.x.
- Watt, H. M., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental psychology*, 48(6), 1594.

- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81.
- Wilkins, J. L. M., Zembylas, M., Travers, K. J. (2002). Investigating correlates of mathematics and science literacy in the final year of secondary school. In D. F. Robataille, & A. E. Beaton, *Secondary analysis of the TIMSS data* (pp. 291–316). Boston: Kluwer Academic.
- Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education*, 72(4), 331–346.
- Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology*, 102(2), 453–466. doi:[10.1037/a0017271](https://doi.org/10.1037/a0017271).
- Wilson, P. M., & Wilson, J. R. (1992). Environmental influences on adolescent educational aspirations: A logistic transform model. *Youth & Society*, 24(1), 52–70.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 189–194. doi:[10.1007/s10862-005-9004-7](https://doi.org/10.1007/s10862-005-9004-7).
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57–67. doi:[10.1023/A:1007999204543](https://doi.org/10.1023/A:1007999204543).
- Yang, Yanyun, & Green, S. B. (2011). Coefficient Alpha: A Reliability Coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, 29, 377–392. doi:[10.1177/0734282911406668](https://doi.org/10.1177/0734282911406668).
- Ye, F., & Wallace, T. L. (2013). Psychological sense of school membership scale: Method effects associated with negatively worded items. *Journal of Psychoeducational Assessment*, 32(3), 202–215. doi:[10.1177/0734282913504816](https://doi.org/10.1177/0734282913504816).
- Zhu, Y., & Leung, F. K. (2011). Motivation and achievement: Is there an East Asian model? *International Journal of Science and Mathematics Education*, 9(5), 1189–1212.