Studying the Factorial Structure of Ghanaian Twelfth-Grade Students' Views on Mathematics

Chapter · August 2014

DOI: 10.1007/978-3-319-06808-4_18

2 authors:

Emmanuel Adu-tutu Bofah
University of Helsinki

Markku S. Hannula
University of Helsinki

Studying the Factorial Structure of Ghanaian Twelfth-Grade Students' Views on Mathematics

Emmanuel Adu-tutu Bofah and Markku S. Hannula

Abstract Researchers often import and adopt surveys from one cultural setting to another in order to collect comparative data or to simplify the laborious process of instrument development. Even when the instrument has been proven to have high reliability in the original setting, the reliability may prove to be much weaker in the new setting, especially when Western instruments are imported into non-Western countries. In this chapter, we discuss the problems of importing an instrument from one culture to another and associated methodological challenges. More importantly, we present a detailed account of using structural equation modeling (SEM) and MPlus software to validate a survey instrument imported to Ghana. The students' Views of Mathematics (VOM) instrument is based on earlier Western research and was further developed in Finland, where it had been validated to have high reliability. First, we used confirmatory factor analysis to test whether the seven factors identified in Finland were identifiable in Ghana. As the original factor structure was found not to fit the Ghanaian data, we continued with an explorative approach to identify the Ghanaian factor structure, resulting in a four-factor structure. For cross-validation purposes, the sample was randomly split into two, one-half of the sample assigned as the calibration sample and the other half as the validation sample. Measurement invariance was established at the configural, metric and structural levels between the calibration and validation sample. We further discuss the measurement artifacts and cultural differences as possible causes for the observed differences in the factor structures between the Ghanaian and the Finnish sample.

Keywords Cross-cultural affects • Views of mathematics • Affect and students' beliefs structures • Factor analysis • Measurement invariance • Multigroup analysis • Construct reliability and validity • Structural equation modeling • Survey instrument

---

E.A. Bofah (✉) • M.S. Hannula
Faculty of Behavioral Science, Department of Teacher Education, University of Helsinki, Helsinki, Finland
e-mail: emmanuel.bofah@hotmail.com; markku.hannula@helsinki.fi

B. Pepin, B. Roesken-Winter (eds.), *From beliefs to dynamic affect systems in mathematics education*, Advances in Mathematics Education,
DOI 10.1007/978-3-319-06808-4_18

Introduction

The important role of students' beliefs on learning mathematics is widely acknowledged. We know that students' affective dispositions influence their learning of mathematics, for the better or for the worse (for an overview, see Hannula 2012). The declining numbers of students that are studying science, technology, engineering and mathematics at university level in the Organization for Economic Cooperation and Development (OECD) countries (Ainley et al. 2008) can at least be partially attributed to students' negative views towards mathematics.

Surveys are an important method for studying students' mathematics-related beliefs. Examples of surveys include The Trends in International Mathematics and Science Study (TIMSS) and The Programme for International Student Assessment (PISA). The first widely used instrument for studying mathematics-related affect was the Mathematics Anxiety Rating Scale (Richardson and Suinn 1972). In the long run, the Mathematics Attitude Scales (Fennema and Sherman 1976) has been much more influential in the field with a total of nine scales, namely: scales for students' anxiety, confidence, success, and effectance motivation, students' perceptions of mathematics as a male domain, the perceived usefulness of mathematics, their ratings of their respective teacher's perceptions of themselves and their parents' (mothers' and fathers') interest in mathematics. In addition, the "Indiana Mathematics Belief Scales," which was designed for secondary school and college level students (Kloosterman and Stage 1992), has been influential in the field. The multiscale instruments presented in the aforementioned studies and others inspired by them are useful in exploring the structural properties of mathematics related beliefs (McLeod 1992; Op't Eynde et al. 2006; Roesken et al. 2011).

Although international comparative studies such as TIMSS and PISA have measured student's mathematics related beliefs worldwide, their instruments have been developed in Western countries, usually in North America. Their approach relies on an unwarranted assumption that the structure of affect is cross-culturally invariant (Van de Vijver and Leung 2000). Empirical studies have revealed that the reliabilities of TIMSS and PISA scales vary across countries, being highest in the Western countries and lower in non-Western countries (Metsämuuronen 2012a; Rutkowski and Rutkowski 2010).

The present study reports the implementation and utility in Ghana of one such instrument "Views On Mathematics (VOM)" scale, which was developed in Finland by Pehkonen's research team (Hannula et al. 2005; Roesken et al. 2007, 2011). Finland, a Nordic welfare state and a member of the OECD, has repeatedly scored very high in human development indexes (e.g., Malik 2013). Specifically, Finland is known to have an excellent educational system, which has achieved eminence in the recent TIMSS and PISA results. Ghana is a sub-Sahara African country that has medium level human development index, slightly above the Africa average (Malik 2013). Ghana has not participated in PISA, and it has been performing poorly in TIMSS (Mullis et al. 2012). First, we will discuss the methodological issues relating to validation of an instrument in a new cultural context. Second, we will also provide a detailed account of applying exploratory and confirmatory factor analysis to analyze the structure of beliefs. Third, we will report our empirical findings regarding the belief structure in Ghana for mathematics in twelfth-grade Ghanaian students.

Theoretical Background

According to McLeod (1992), four structural qualities distinguish students' mathematics belief systems: (a) beliefs about mathematics; (b) beliefs about the self; (c) beliefs about mathematics teaching and (d) beliefs about the social context. The first classification, beliefs about mathematics, includes students' beliefs such as thinking that mathematics is difficult, and the belief about the usefulness of mathematics. The second categorization, beliefs about the self, includes the self-concept, confidence, and causal attributions. These, in turn, include success and failures related to mathematics. The third category, beliefs about teaching, includes beliefs about what is expected of a teacher to help students learn mathematics. In other words, this measures the importance that students attach to mathematics instruction (Op't Eynde et al. 2002). McLeod's fourth category, "beliefs about the social context", includes the cultural issues associated with mathematics education, influence of parents and others outside the school on one's mathematics learning in addition to one's home environment.

Op't Eynde and colleagues (2002, 2006) further developed a framework of students' mathematics-related belief systems. Based on relevant literature reviews, they clustered students' mathematics-related beliefs systems into implicitly or explicitly held subjective conceptions students hold to be true for:

1. "Beliefs about mathematics education: (a) beliefs about mathematics, (b) beliefs about mathematical learning and problem solving, (c) beliefs about mathematics teaching;
2. Beliefs about the self as a mathematician: (a) intrinsic goal orientation beliefs, (b) extrinsic goal orientation, (c) task-value beliefs, (d) control beliefs, (e) self-efficacy beliefs;
3. Beliefs about the mathematics class context: (a) beliefs about the role and the functioning of the teacher, (b) beliefs about the role and the functioning of the students in their class (c) beliefs about socio-mathematical norms in their own class." (Op't Eynde et al. 2006, p. 63)

Studies on Finnish teacher training students (Hannula et al. 2005) and upper secondary school students (Roesken et al. 2007, 2011) have provided data on beliefs and motivation. Roesken and her colleagues argued that it is possible to empirically distinguish between students' cognitive beliefs, motivations, and their emotional relationship with mathematics. They reported five dimensions for students' cognitive beliefs (*ability, success, teacher quality, family encouragement,* and *difficulty*), and separate dimensions for student motivation (*effort*) and emotions (*enjoyment of mathematics*).

Due to the high reported reliability of the scales for a similar age group (Cronbach's alpha (α): 0.800–0.910: Roesken et al. 2011), and inclusion of emotional and motivational dimensions, we decided to use the VOM instrument for our study in Ghana for measuring upper secondary students' mathematical beliefs systems. Moreover, we used existing measurement scales, which allowed us to compare and synthesize what is already known. This study is also based on the fact that indigenous research and theorizing, as well as research that integrates different cultural perspectives, are crucial to the establishment of more useful and universal theories (Leung and Zhang 1995; Van de Vijver and Leung 2000). Many researchers have lamented about the Western bias in cross-cultural research (e.g., Van de Vijver and Leung 2000). The bias is reflected in the methods used, and the theoretical orientations adopted. For example, there has been severe criticism of validity and reliability problems associated with the importation of Western instruments into non-Western countries (e.g., Cheung 1996; Van de Vijver and Leung 2000).

The cultural backgrounds of students' in Ghana differ from Finland in many respects (e.g., school types, educational resources, disparity between and within schools, socialization norms, daily experiences). Ghana has had relatively stable economic development, which is reflected in its comparatively high human development in relation to its gross national income per capita (Malik 2013). In their educational structures, these two countries have similarities and differences. Compulsory education in Finland starts from age seven, whereas in Ghana compulsory education starts from age 4. Both countries have 6 years of secondary education. In Finland, all teachers are required to have a master's degree including at least 1 year of pedagogical studies, whereas in Ghana, teaching requires a diploma or a Bachelor's degree. The gross enrolment ratio in senior high schools is 34 % in Ghana whereas that of Finland is above 100 % (UNESCO 2011). The share of girls' enrolment in senior high schools in Ghana is 44 % (Ghana Education Service 2013) and in Finland 57 % (Statistics Finland 2013). These vast differences makes it interesting to investigate how the students' in these two countries view themselves as learners of mathematics.

Studies on mathematics related affects in Ghana have been using various survey instruments. For instance, Eshun (2004) and Nyala (2008) used The Mathematics Attitude Scales (Fennema and Sherman 1976) to measure students Mathematics self-belief. Asante (2012) used the Attitude Towards Mathematics Inventory scale (ATMI), (Tapia and Marsh 2000). Asante (2012) reported the Cronbach's alpha reliability for ATMI to be 0.940, and Nyala (2008) reported 0.630 for Fennema & Sherman Mathematics Attitude Scale. In those studies, on the scale of mathematics self-confidence, Asante (2012) and Eshun (2004) reported significantly higher scores for males at the senior secondary school whereas Nyala (2008) reported no significant different between both sexes at the junior secondary school level. Also on the usefulness of mathematics scale Eshun (2004) reported higher scores for males at the senior secondary school whereas Nyala (2008) reported higher scores for females at the junior secondary school. Similar findings were reported for the mathematics as a male domain and anxiety at both school levels with girls reporting lower scores (Eshun 2004; Nyala 2008).

Multiscale construction and development is usually a multistage multifaceted process. Over the past several decades, scales for measuring students' affective structure have become the norm. Their possible widespread usefulness is because they provide multiple converging pieces of information about the studied constructs and can involve unlimited sample size in addition to robust methods for analyzing the sample to facilitate generalizing the findings. Most of these instruments or constructs are imports from Western research. Translations of such constructs are an inevitable tool to conduct such studies. However, translation does not guarantee that the translated instruments will measure the same as in the original. Differences in linguistics, cultural or both can make translations of the instruments difficult and meaningless. As such, the adaptation of these instruments should be based on theory, construct reliability analysis, exploratory and confirmatory factor analysis (Marsh et al. 2012). As Marsh and his colleagues argued "from a construct validation perspective, theory, measurement, statistical analysis, empirical research, and practice are inexorably intertwined, so that the neglect of one will undermine the others." (ibid. p. 111)

Researchers, policy makers and educators interest in cross-national comparative studies such as the TIMSS and PISA have gained considerable attention recently. However, challenges to TIMSS and PISA studies are that the target populations have unique social conventions-cultures, school systems and cognitive structures and styles (Metsämuuronen 2012a, b). Implementing the instruments developed in one cultural setting into a new cultural setting is problematic regardless of their high reliabilities in the original settings. For example, Metsämuuronen (2012a), and Rutkowski and Rutkowski (2010) reported that some scales (e.g., math self-concept) that had been used in PISA and TIMMS studies showed less reliable scores in East Asia, the Middle East, and Europe, when compared to data from North America, where the scales were originally constructed. Metsämuuronen (2012b) found that in the TIMSS2007, the math attitudes scale were not invariant and manifested "*fragmentation*" in most of the participating countries (in most low achieving countries) due to different cultural values. With empirical examples, Rutkowski and Rutkowski (2010) found that the possible cause to this was too much missing data: a possible sign of respondent misinterpretation.

Other researchers have argued that TIMSS and PISA uses robust psychometric, sampling methods, and translations methods, yet the math motivational construct is still affected by construct bias, method bias and item bias (Van de Vijver and Leung 2011). Rutkowski and Rutkowski (2010) argued that, the country composition of PISA makes it impossible to have motivational constructs that will measure the desired goals for non-OECD countries. On the other hand, for each successive cycle, TIMSS have been dropping or adding new mathematics motivational constructs in response to reported validity, reliability, psychometric properties of the data and feedback from various countries coordinators (Marsh et al. 2012; Rutkowski and Rutkowski 2010).

Aims of the study

The present study objectives are:

- to test for the factorial validity of the VOM for Ghanaian upper secondary students,
- in the event of a model misfit (i.e. the seven-factor structure), propose and statistically test an alternative factorial structure,
- to cross-validate the new factorial structure across a second independent sample from the Ghanaian data,
- to test for factorial and structural invariance across a subsample (gender) from the Ghanaian sample, and
- to affirm the theoretical structure of the VOM construct.

The Present Investigation: *A Priori* Predictions and Research Questions

The present study examined three research questions that give support for construct validity and reliability. First, we will compare the reliabilities of the scales in the Ghanaian sample to the reliabilities observed in Finland. Second, a more robust approach, confirmatory factor analysis will be used to validate the constructs. If the theoretical model is not supported, exploratory factor analysis will be used to determine the factor structure of the Ghanaian sample. The sample will be split into calibration and validation sample. Third, confirmatory factor analysis will be used to test the equivalence (measurement invariance) of the derived constructs with the validated sample.

Hypothesis 1 Research have showed that imported constructs regardless of their high reliabilities in the original settings, often shows a very low reliabilities when imported to a different cultural setting (e.g., Cheung 1996; Metsämuuronen 2012a; Rutkowski and Rutkowski 2010; Van de Vijver and Leung 2000). Given that the constructs come from Western research, we hypothesized that reliability estimates (e.g., Cronbach's alpha (α) will be lower than in the reported studies in Finland.

Hypothesis 2 All things being equal, we could have hypothesized that the students' views of themselves as mathematics learners using VOM could be explained by seven factors *(ability, effort, success, teacher quality, family encouragement, difficulty, and enjoyment*). However, since these items have not been fully used in other countries apart from Finland, we will leave open the research question as to whether there is support for the seven-factor structure identified in previous studies (Hannula et al. 2005; Roesken et al. 2011). We hypothesized *a priori* that: (a) each item has a non-zero loading on the VOM factor it was designed to measure, and zero loadings on all other factors, (b) the factors are correlated and, (c) the error/uniqueness term

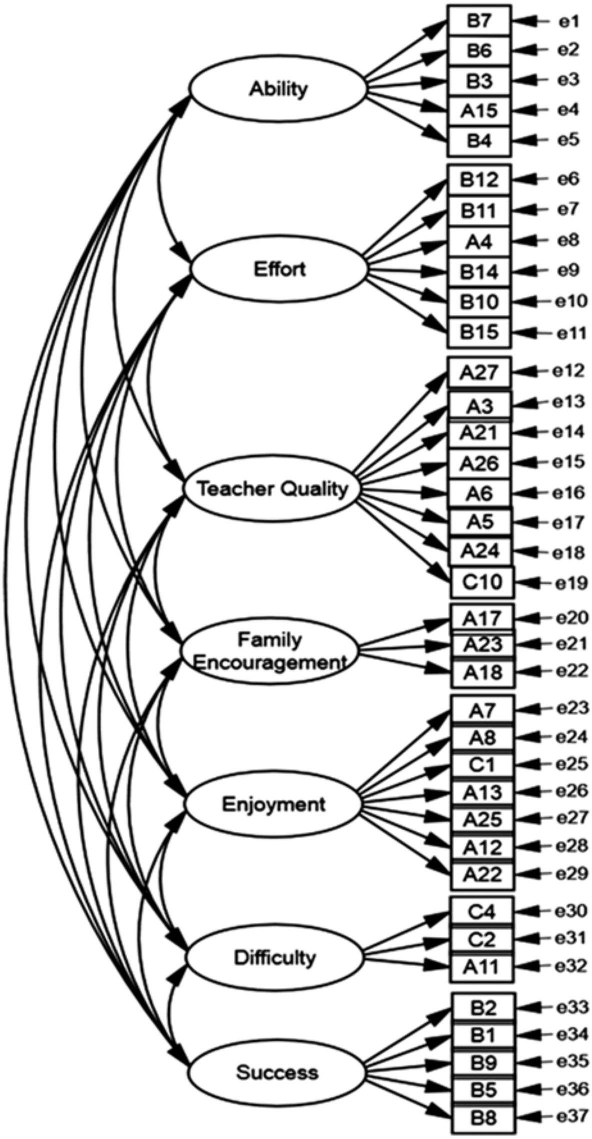

Fig. 1 The seven-factor model of students' view of themselves as learners of mathematics as identified in Finland. Short one-way arrows denote measurement error terms associated with the observed measures. This diagram was drawn based on the findings in Roesken et al. (2011)

for the item variables are uncorrelated. A schematic representation of this postulated model is presented in Fig. 1.

Hypothesis 3 In the confirmatory factor analysis (CFA) that follows the EFA, we expect support for the invariance of factor loading, and factor variance-covariance (structural invariance) of the new proposed factor structure for the calibrated and validated independent sample including students' gender. We hypothesized a low to moderate correlation between the constructs.

Methods

Given the cultural background differences between students of the Finnish and Ghanaian samples, it is possible that some items of the instrument function differently, which may lead to different factorial structures. The factorial validity of VOM factors has been examined using only Principal Component Analyses (PCA) approach. Based on deficiencies associated with exploratory factor analysis (EFA) in general (Marsh et al. 2009, 2011; Sass and Schmitt 2011; Schmitt 2011) and PCA in particular (Marsh et al. 2009; Schmitt 2011), Confirmatory Factor Analysis (CFA) procedures were used to provide a more robust test of factorial validity. Moreover, the factorial structure of the VOM had not been validated using two independent samples of any data.

The Need for Factor Analysis

Factor analysis (FA) has become a highly popular statistical method in the behavioural sciences. In fact, it is especially relevant for test construction and development, as we will see throughout the rest of this chapter. FA is a method generally used to help uncover the relationships between assumed latent variables and manifest variables.

There are two main types of factor analysis: EFA and CFA. EFA is a data-driven approach such that no *a priori* specifications are made concerning the number of common factors and the indicators (i.e. factor loadings). In contrast, CFA is used to test the extent to which *a priori*, theoretical model of factor loadings provides an adequate fit for the actual data. Thus, in EFA the statistical method determines the factors and loadings, whereas CFA detects how well our theoretical model matches reality (the actual data) (Hair et al. 2010). Thus, CFA is a tool that enables one to confirm or reject *a priori* theory. FA bears resemblances to a statistical approach often used in the behavioral and social sciences for data reduction, and has been used in all analyses involving VOM which is called principal component analysis (PCA) (Raykov and Marcoulides 2010). PCA has often been assumed to be a factor analytic method. However, from a technical perspective, PCA is not a member of the FA family (Schmitt 2011; Raykov and Marcoulides 2010). One main difference between PCA and FA is that PCA assumes no measurement of error, whereas FA methods account for the measurement error (Schmitt 2011). Moreover, in FA the common factors are interpretable in addition to reduction of complexity whereas PCA is only for data reduction. Schmitt argued that though evidence suggests that PCA can produce similar results as FA when measurement reliability is high and when factor items are many, estimations of PCA will be less close to CFA than any factor analysis method (Schmitt 2011).

There are some limitations associated with EFA such as (a) not being able to yield a unique factor structure, (b) not defining a testable model, (c) not assessing the extent to which a hypothesized model will fit given data, (d) not being able to suggest model improvements, and, (e) not offering a strong analytic framework for evaluating the equivalence of measurement models across distinct groups (e.g., gender) (Byrne 1991; Brown and Moore 2012). Thus, CFA is a more powerful tool for the testing of factorial validity and construct validation, which necessitated its use in the present study.

It is also important to know that structural equation models (SEMs) based on CFAs may produce very different structural coefficients and model fit statistics than EFAs, as the CFA approach can depict the factor structure differently (Marsh et al. 2011; Sass and Schmitt 2011; Schmitt and Sass 2011). Therefore, specifying the appropriate measurement model (EFA or CFA) has direct implications for replicating factor structures and interpreting structural coefficients (Marsh et al. 2009; Marsh et al. 2011; Sass and Schmitt 2011).

In determining the number of factors, different statistical methods were used, the Minimum Average Partial (MAP) method using the IBM SPSS Statistic 21 (O'Connor 2000) and Parallel Analysis (PA) (Henson and Roberts 2006) procedure in Mplus. The use of the above factor retention methods were used as recommended by Schmitt (2011). The determining of the factors was also guided by the quality of the variables measuring the factors, size of the loadings (>0.300) on the standardized scale, size of indicator communalities, number of variables that load on the factor (min 3), factor homogeneity, and factor determinacy—correlation between the estimated factor score and the factor.

Participants

The sample consisted of 2034 twelfth-grade Ghanaian students (*M*age = 18.49, *Mdn*age = 18, *SD*age = 1.25; 58.2 % girls). Nine Senior High Schools were selected from urban and rural schools in Ghana based on their rankings by the Ghana Education Service. The first author gave the questionnaire to the students during their normal class hours in the summer of 2011. Participants' permissions were collated and received by the heads of institutions. The participation in the survey was voluntary and students had the right to withdraw or skip any question that they did not wish to answer. The schools were selected to represent the most representative variety of school types in Ghana, and they included single-sex, coed, private, religious and public schools. Some schools fell under more than one of these categories. The students were enrolled in different mathematics classes; core mathematics (49.3 %) and elective mathematics (50.7 %). They were enrolled in either General Arts (33 %), Business (19.2 %), Science (29.1 %), or Vocational Science (18.7 %) streams.

We cannot claim that the sample is representative of the entire student population of secondary schools in Ghana, but the schools were chosen to represent the most commonly occurring types of high schools in terms of the social intake, disciplines and rates of academic success and failure. Therefore, the results cannot be applied to the students of all schools, though they are representative of students in a range "typical" of the secondary school system in Ghana. There were 63 different student classrooms with an average class size of 32 students.

Measures

We used the VOM instrument (Hannula et al. 2005; Roesken et al. 2011). The instrument consists of 55 items, most of which had originated from a qualitative study on student-teachers' views of mathematics (Pietilä 2002). An additional four items originated from a previous study on Finnish comprehensive schools (Nurmi et al. 2003), and 10 items originated from the self-confidence scale of Fennema-Sherman mathematics attitude scales (Fennema and Sherman 1976), and some novel items developed by the team to measure student perceived success in mathematics. Apart from the 10 Fennema-Sherman items, all the other items were originally in Finnish and had been translated into English.

Items were assessed using a 5-point Likert scale. The statements in the questionnaire were grouped around the following topics: (1) Experiences as a mathematics learner (A1–A29), (2) Image of oneself as a mathematics learner (B1-B15), and (3) View of mathematics and its teaching and learning (C1–C11). The instrument has been successfully implemented in teacher and student settings, whereby reliability and validity of the instrument have been demonstrated. Cronbach's alpha (α) reliability in a study of Finnish upper secondary students (Roesken et al. 2011) was between (0.800–0.910) and in a study of student teachers (0.780–0.910) (Hannula et al. 2005). Abbreviated four-item versions of the scales for *success, ability, effort, difficulty* and *enjoyment* were also used in a study of Finnish comprehensive school students (Hannula and Laakso 2011), and again, the reliabilities were found to be good: for eighth-grade students (0.780–0.880) and with exception of *effort* (0.660) reliabilities were also good for fourth-grade students (0.750–0.810).

The studies reported high correlations between core dimensions of the beliefs (up to 0.790; Roesken et al. 2011). We are aware that the high correlation (>0.750) between constructs is a possible sign of multicollinearity (Byrne 2012; Hair et al. 2010). Multicollinearity violations may lead to the wrong interpretation of the findings because it makes it difficult to predict the individual importance of a predictor. Moreover, instances where even a proper solution can be obtained, multicollinearity can lead to inaccurate parameter estimates and a high incidence of Type II errors, particularly when reliability is weak, sample size is small, and explained variance is low (Grewal et al. 2004, p. 526). Although multicollinearity was high, it was not a concern for the authors because reliabilities were high, a high R^2, and the large sample size offset the problems caused by the multicollinearity (M. Hannula, personal communication, October 8, 2013). Other literature also supports the argument (see: Grewal et al. 2004; Mason and Perreault 1991). Grewal and colleagues, further argue that the problem of multicollinearity should not be viewed in isolation unless the multicollinearity is severe.

Analyses

For cross-validation purposes, the whole sample was randomly split into two, with one-half of the sample, ($N = 1{,}017$) assigned as the calibration sample and the other half ($N = 1{,}017$) as the validation sample. The reason for this split was to ascertain whether a model that has been specified in one sample could be replicated over a second independent sample from the same population. The objective was to find a robust model that was replicable among the sample and avoid the problem of capitalization on a chance outcome that can appear when only one sample is analyzed.

Data were analyzed in three-stages. First, CFA procedures were conducted on the whole sample to investigate whether the established dimensionality (seven-factor structure) and factor-loading pattern fitted the Ghanaian twelfth-graders' sample. This was the confirmatory aspect of the analysis.

Second, the data did not fit the hypothesized model, therefore analyses proceeded in an exploratory mode using both EFA and CFA approaches to identify the course of the misfit, and specify an alternative model for the factor structure. The EFA method was used to examine the number of underlying factors and the CFA-post hoc procedures were used to identify item parameters that contributed to the model misfit. Information from the exploratory analyses (both EFA and CFA-post hoc) was used to propose a final factorial structure based on the calibration sample. CFA was then used again to investigate whether the established dimensionality and factor-loading pattern fitted the independent validation sample. Third, VOM equivalency across the calibration and validation sample was tested in respect of (a) factor form invariance or configural invariance- that freely estimated the item loadings on both samples, (b) factor loading invariance or metric invariance for the calibrated and validated samples, and (c) the common characteristics of individuals by examining factor variances and covariances (FVCV) relationship in both samples (structural invariance). FVCV invariance will also help to ascertain the homogeneity (unidimensionality) of the constructs.

Goodness of Fit and Reliability Estimates

Evaluation of a model fit was based on multiple criteria that reflected statistical, theoretical, and practical perspectives. A goodness of fit was evaluated by using Chi-Square Difference Testing using the Satorra-Bentler Scaled Chi-Square test statistic ($SBS\Delta\chi^2 - MLR\chi^2$), the Root Mean Square Error of Approximation (RMSEA), the comparative fit index (CFI) and the Tucker-Lewis Index (TLI), which are relatively independent of sample size (Chen 2007).

The hypothesized model and the final model were compared for the best fit using, the information Criteria indices such as Akaike (AIC), Bayesian (BIC), and Sample-Size Adjusted (SSBIC) because the models were not nested. The CFI and TLI vary along 0–1 and values ≥0.90 and 0.95 are deemed acceptable and excellent threshold respectively, and RMSEA ≤0.08 and 0.05 for close and reasonable fitting model (Brown 2006). For AIC, BIC, and SSBIC, the model with the smallest value information criterion is preferred. When evaluating the worth of individual parameters, statistical significance values as indicated by the Mplus *z*–values, goodness-of-fit based on the normalized residual values, modifications indices (MIs), and model meaningfulness were also taken into account.

The SEM analyses in the present study were done using Mplus 7.11 (Muthén and Muthén 1998–2012). All analyses were based on the Mplus robust maximum likelihood estimator (MLR), with standard errors and test-of-fit that were robust to non-normality of the observations to control for the non-independence of observation (Muthén and Muthén 1998–2012). In addition, the choice of MLR, rather than categorical variable estimator procedure was based on research studies (Rhemtulla et al. 2012) that indicated how categorical methods make little or no differences when Likert scales of five or more categories are treated as categorical variables or continuous variables. In order to include all of the observed data, missing data patterns were handled with Mplus feature of full information maximum likelihood (FIML).

We analyzed the normality assumptions, by investigating the normality of each variable in terms of its kurtosis and skewness. With guidelines of normality (i.e., skewness: <3; kurtosis: <7) proposed by Curran et al. (1995), there were few non-normality items that supported the use of robust maximum likelihood estimator (MLR).

Cronbach's alpha has been used as the standard measure of reliability for a long time, although it is known to either underestimate or overestimate reliability (Geldhof et al. 2014; Novick and Lewis 1967). Composite reliability (ω) (Geldhof et al. 2014; Raykov 2012) used in conjunction with structural equation modeling (SEM) will be estimated to complement the α estimates of the new VOM scales. Composite reliability (ω) takes into account the computed factor loadings, and produces more precise estimates of reliability than those provided by α (Geldhof et al. 2014; Raykov 2012). It is interpreted in the same way as Cronbach's alpha. Generally, ω values of 0.600–0.700 are acceptable in exploratory research (Hair et al. 2010).

Invariance Model Testing

Measurement invariance is the equivalence of a measured construct in two or more groups, such as people from different cultures (Chen 2008). It assures that the same constructs are being assessed in each group (Sass 2011). Invariance model testing usually begins with a baseline model often called the *configural model* in which all parameters in the model are freely estimated across groups. When the baseline model fits adequately in each group, this indicates that the same number of factors best represents the data for all, and the same variables define each factor across groups. Then one can test if the factor structures are equal by restricting the factor loadings to be equal across groups. The model in which the factor loadings are held equal is usually called the *metric invariance or weak invariance.* When metric invariance holds, we can conclude that the constructs are similarly manifested in each of the groups. Finally, we imposed constraints to factor variance and covariance to test for structural invariance. A non-invariance structural model would suggest a differential structure for the construct being measured across the groups (i.e. the associations among the underlying factors varying across groups). Thus, structural invariance indicates the homogeneity (unidimensionality) of the constructs, which is a necessary condition for both reliability and validity.

Results

Stage 0: Computing Cronbach's Alpha (α) for the Hypothesized Scales

Our *a priori* model (Fig. 1) posited that the VOM constructs could be explained by seven-factors. The seven factors were the *ability, effort, teacher quality, family encouragement, enjoyment of mathematics, difficulty of mathematics, and success* (Roesken et al. 2011). The first confirmatory approach was to compute the Cronbach's alpha for each factor. The Cronbach's alpha coefficients were calculated as indicators of factor reliability. These alpha coefficients for the Ghanaian sample were within the acceptable standard for *ability* (0.863) and *enjoyment* (0.764), with the rest below the acceptable threshold: *effort* (0.538), *teacher quality* (0.190), *family encouragement* (0.623), *difficulty* (0.565) *and success* (0.661). The Cronbach's alpha values indicated that most scale reliabilities were considerably lower than those of the Finnish sample. The scale for *teacher quality* was unacceptably low, but two of the scales were above the usual considered 0.700 reliability threshold and two of the remaining three were sufficiently reliable for some researchers to consider them acceptable (Hair et al. 2010), or even all three when their content coverage and unidimensionality were sufficient (Schmitt 1996). Since Cronbach's alpha does not index unidimensionality of the constructs together with what have been discussed earlier, there is good reason to apply a more robust approach (CFA-stage 1; EFA-stage 2) to test the whole model, which was stage 1.

Stage 1: Test for Factorial Validity; Confirmatory Factor Analyses

CFA indices for the hypothesized seven-factor model were poor from both statistical ($\text{MLR}\chi^2_{(608)} = 1922.993$) and a practical (CFI = 0.843, TLI = 0.828, RMSEA = 0.046) perspective. This model was therefore rejected. We also tested the model fit after removing the scale with the lowest reliability (*teacher quality*), but the model fit was only marginally improved (MLR$\chi^2_{(362)} = 1361.447$, CFI = 0.849, TLI = 0.831, RMSEA = 0.052).

A further look at the correlations indicated very high factor correlations between some of the constructs, which indicated multicollinearity: the correlations between the *ability* and *difficulty* factors ($r = 0.853$), and *ability* and *enjoyment* ($r = 0.847$), *difficulty* with *enjoyment* ($r = 0.871$), suggested that the factor structures were not statistically distinguishable, thus they measured the same dimension.

Stage 2: Exploratory Factor Analysis

After rejecting the *a priori* model, the next logical step was to take an exploratory approach to analyze these data in order to identify a better fitting model. A particularly important question was (a) whether the Ghanaian data could be described more reasonably by a model that specified less than, or more than the seven factors, and (b) whether an independent sample from the Ghanaian data exhibits the same pattern of loadings for all factors. The data were reanalyzed using exploratory factor analysis (EFA) to answer these questions. Previous research indicates different dimensions of mathematics-related beliefs correlate (e.g., Roesken et al. 2011), thus Geomin (oblique) rotation was used as the rotation procedure to get a cleaner simple factor structure that is similar to CFA (Schmitt 2011).

All 55 items from the original questionnaire were used in the EFA analysis. The results from the Minimum Average Partial (MAP) method indicated a six-factor solution whereas Parallel Analysis indicated a seven-factor solution**.** EFA analyses for 4–7 factors were run on the data simultaneously to determine if there were plausible models that could explain the relationships among the items. A four-factor solution was included and tested because of the high correlation that was identified early by pre-supposing three factors to be measuring the same dimension. The residual variance of all items, i.e. the proportion of variance in the indicators that has not been explained by the latent variables, were checked in respect of all the proposed factor structures (four, five, six, and seven factors). Items with very high residual variance (>0.800), loadings of less than 0.300 and high cross-loadings were deleted. When the EFA was re-run, no item loaded for the seventh factor and analysis was continued with four, five and six factor models. Again, items with a high residual variance and which loaded less than 0.300 or high cross-loadings were deleted and the EFA was re-run. Whereupon only two items loaded on the sixth factor and thus we removed the items and continued comparing four and five factor solutions. The high factor correlation (unstandardized: $r = 0.914$, standardized $r = 0.849$) indicated that two of the constructs in the five-factor solution were not statistically distinguishable. The EFA for a four-factor structure was acceptable as the final model.

The Four-Factor Structure

Given both substantive and statistical considerations discussed above, the EFA suggested the four-factor solution as the most optimal to represent the Ghanaian data. The *a priori* hypotheses model included 37 items, whereas only 29[1] items out of 55 items exceeded the threshold for inclusion in the analysis and were included in the present EFA model. The four factors were labeled as *self-confidence, self-concept*, *family encouragement,* and *teacher quality.*

The *self-concept* factor includes all five *ability* items, five out of seven *enjoyment* items, two out of three *difficulty* items, one *success* item, one *effort* item and two new items making a total of 16 items. The two new items were item A10: *My eagerness to study mathematics is seasonal* and item A19: *Mathematics has been a clear and precise subject to study*. Therefore, this study has demonstrated that the *ability*, *success,* and *enjoyment* factors loads on the same factor and therefore can be treated empirically as the same construct. Absolute target loadings were high between (0.395 and 0.789) with non-target loadings between (0.003 and 0.159). The *self-confidence* factor include three items from the *success* factor (B9, B2, B1), and one *effort* factor item (B15). Absolute target loadings on the *self-confidence* factor were between 0.435–0.688, and very low non-target loadings of between 0.003 and 0.088. The items of the *teacher quality* factor were all from the Finnish based factor except that, two items from the original solution were not included (A5, C10), because they failed to surpass the threshold value in addition to being cross loading items. Target loadings were between 0.358 and 0.740, non-target loadings between 0.011 and 0.118. All items on the *family encouragement* factor exceeded the threshold for inclusion in the analysis. Target loadings were between 0.496 and 0.596, whereas non-target loadings were between 0.002 and 0.218. All factor loadings were statistically significant (p <0.001). The patterns of the correlations were consistent with a low-moderate (0.141–0.430) correlation in line with the *a priori* hypothesis. These results support the assertion that the VOM structure of the Ghanaian data is different.

The four factors can be identified within McLeod's (1992) structural qualities associated with VOM. The *self-confidence* and *self-concept* factors corresponded to 'beliefs about self', whereas the *teacher quality* factor corresponded to that of 'beliefs about mathematics teaching' and the *family encouragement* factor to 'belief about the social context'. Factor determinacies were 0.958 for self-*concept*, 0.878 for *teacher quality*, 0.862 for *self-confidence* and 0.798 for *family encouragement*.

[1] Four items were deleted due to content overlap detected from the post hoc confirmatory factor analysis in the next section.

Post hoc Confirmatory Factor Analysis

The EFA suggested a four-factor structure for the VOM. We ascertained the extent to which the newly specified model fitted the data over the hypothesized model, by using the CFA approach. For a better model over the hypothesized model,[2] we compared the AIC, BIC, and SSBIC information criteria between two models because they were not nested. This is because the new four-factor model had been structurally revised (i.e., factors had been partially collapsed into a single latent variable—for example, *ability*, *enjoyment* and *difficulty* factors). To improve the VOM constructs, modification indices (MI) were consulted. Six consecutive CFA analysis guided by the MIs, correlations between items and item residual variance, led to items B7, A7 and C1 (i.e. all on the self-*concept* factors) being deleted and the error covariance between items B4 and B3 included in the final model (for detail analysis, see Bofah and Hannula 2014).

There is considerable discussion in the CFA literature regarding the interpretation and need for the inclusion of the error covariance in addition to what the appropriate solution to the problem is (Byrne 2012; Byrne 1993; Marsh et al. 2012). Studies have indicated that including the error covariances in a model improves the model fit whereas excluding them is likely to bias and inflate parameter estimates (Byrne 2012; Marsh et al. 2004, 2012). The inclusion of the error covariance has been justified when these parameters represented non-random measurement errors (Byrne 1991) due to *method effects; and* as such their presence was expected. Brown and Moore (2012, p. 362), and Edelen and Reeve (2007) argued that the possible causes for such covariation are results from the following: common assessment methods (e.g., questionnaires); reverse items, or similarly worded items, items that are presented sequentially, and items with high content overlap. They also listed items prone to differential susceptibility to other influences such as self-report items, demand characteristics, reading difficulties, item translation, acquiescence, and the format of the instrument or social desirability. In our study data set, translation, and content overlap was the case in all but one of the error covariance (i.e., between B3 and B4).

Moreover, there seems to be a good explanation for the error covariance between items B3 ("Mathematics is my weakest subject") and B4 ("Mathematics is difficult for me"). Crosstabulation of responses revealed that most respondents tended to have similar agreement with both items, though there was a significant subset of respondents who disagreed strongly with B3 and yet agreed with B4. We assume that these respondents are the students who find mathematics difficult, and that who also struggle more with other subjects. Another explanation was that items B3 and B4 were presented sequentially in the questionnaire and both measure students' weakness in mathematics that display local dependence of both items. In addition, both items were negatively worded. Including error covariance between these items improves the model fit because it adjusts for this response pattern.

Reviewing the error covariance, we see that not only is it very highly significant ($p < 0.0001$) in the model, it was also very high ($r = 0.444$). We evaluated the

[2] Indices not reported.

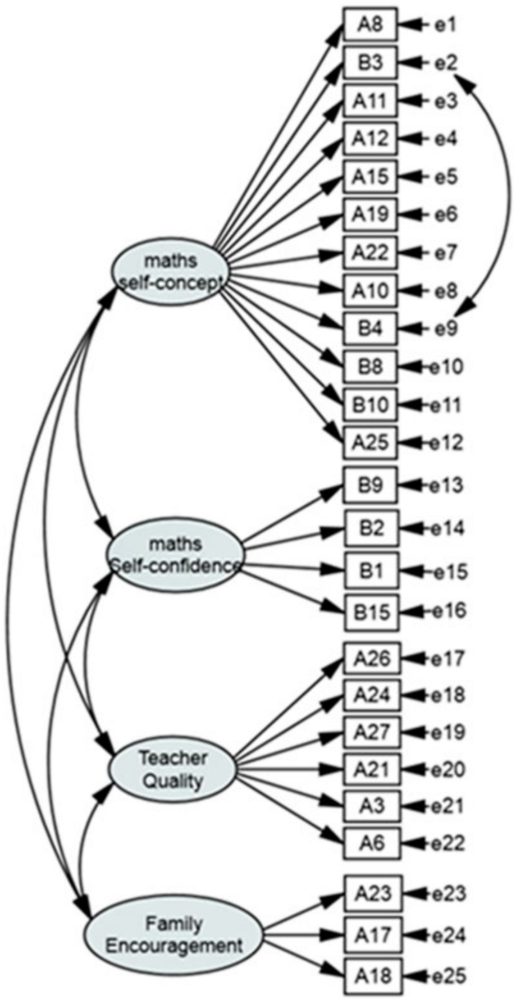

Fig. 2 Final hypothesized model and Baseline model of the VOM for Ghanaian twelfth-grade students. Short one-way arrows stand for measurement error terms associated with the observed measures

strength associated with the error covariance term linking item B3 and B4 together with the indicated rationale, and considered it to be more realistic to include this parameter in the model, rather than ignoring its presence. We tested models with and without the error covariance B3 and B4 to show that including the error covariance in the final model improved the model. Fit indices for model with error covariance (MLR$\chi^2{}_{(268)}=612.874$, CFI = 0.932, TLI = 0.923, RMSEA = 0.036), and model without error covariance (MLR$\chi 2{}_{(269)}=690.419$, CFI = 0.916, TLI = 0.907, RMSEA = 0.039). All the fit indices together with the SB$\Delta\chi 2-$MLRχ^2 value of 61.314 and $\Delta df=1$ indicate that the model with the error covariance provides a significantly better fit to the data than the model without the error covariance (the critical value for SBS$\Delta\chi^2-$MLRχ^2 is 3.84; $\alpha=0.05$, $df=1$).

Therefore, our final model illustrated in Fig. 2 became the final model that represented the VOM structure for the Ghanaian data. It provided the baseline model for subsequent analyses related to cross-validation and multi-group invariance testing of the Ghanaian data.

Stage 3: Cross-Validation Analyses

We continued to cross-validate the factor structure using the results obtained from both the EFA and CFA analyses. Our hypothesized new model illustrated in Fig. 2 provided the baseline model. Cross-validation of our hypothesized model was achieved by testing for invariance across separate calibration and validation samples of the data. Measurement invariance modeling starts with testing for configural invariance. There were support[3] for the *configural, and metric models*, which indicated support for the construct validity across the calibrated and validated samples. Of substantial interest were the two specified residual covariances and the extent of their invariance across the calibration and validation samples. We considered it worthwhile for psychometric reasons and to remove any doubt of capitalizing on chance for their inclusion in the model (MacCallum et al. 1992). We postulated a model in which, the factor loadings, factor variances, factor covariances and residual covariances were constrained to be equal call the structural model.

The sample supported the structural invariance model. Consistent with our study hypotheses, all correlations were in the low to modest range ($r = 0.191$–0.535) between the dimensions (see Table 1). Support for the structural invariance, indicated the unidimensionality of the constructs..

In addition, gender invariance (not reported) was tested and there was support for *configural*, *metric*, and *structural invariance*, which gives a further support for the validity and reliability of the constructs.

Reliability of the New VOM Scales

In response to the reliability hypothesizes, we began by evaluating the Cronbach's coefficient alpha (α) reliability and the Composite reliability (ω) used in conjunction with SEM models of the new four factor VOM scale.

Given that the original items and the constructs come from Western research, it would be expected to find lower reliability estimates. Moreover, the brevity of some of the constructs, (e.g., *family encouragement*) coupled with the positive relationship of Cronbach's alpha reliability values to the number of items on a construct, led some of the coefficient alpha (α) estimates to be below the acceptable threshold. Reliability values of some scales reached the desirable standard of 0.700 (self-concept, $\alpha = 0.872$; $\omega = 0.868$), (*teacher quality*: $\alpha = 0.706$; $\omega = 0.716$). However, they also fell below an acceptable value of 0.700. Although the thresholds may decrease to 0.600 (*self-*

[3] Fit indices not reported.

Table 1 Factor structure relating the VOM items

Item	Factor loadings	*SE*	Item wording
Mathematics self-concept factor(1)			
A8	0.589	0.028	Doing calculations has been enjoyable.
RB3	1.124	0.027	Mathematics is my weakest subject.
RA11	0.862	0.027	Mathematics has been difficult in upper secondary school.
A12	1.118	0.024	Mathematics is my favorite subject.
A15	0.851	0.025	I have made it well in mathematics.
A19	0.882	0.029	Mathematics has been a clear and precise subject to study.
RA22	1.044	0.028	Mathematics has been the most boring part of my study.
RA10	0.685	0.032	My eagerness to study mathematics is seasonal.
RB4	1.061	0.027	Mathematics is difficult for me.
B8	0.641	0.029	I can handle advanced Mathematics tasks.
RB10	0.948	0.030	I have a wrong attitude about mathematics
A25	0.518	0.033	I enjoy pondering over mathematics tasks.
Mathematics self-confidence factor (2)			
B9	0.586	0.038	I know that I can do well in mathematics
B2	0.530	0.036	I can get a good grade in mathematics.
B1	0.497	0.037	I am certain that I can learn mathematics.
B15	0.260	0.032	It is important for me to get a good grade in mathematics.
Teacher quality factor (3)			
A26	0.961	0.031	The teacher has so far been a positive example.
RA24	0.929	0.032	The teacher rushes through the teaching of mathematics.
RA27	1.029	0.034	I would need a better teacher.
RA21	0.787	0.038	The teacher does not inspire me to study mathematics.
A3	0.544	0.033	The teacher explains the studied topics.
A6	0.470	0.037	Teacher explains what the studied topics are needed for.
Family encouragement factor (4)			
A23	0.799	0.044	My family encourages me to study mathematics.
A17	0.863	0.045	The importance of competence in mathematics has been emphasized at my home.
A18	0.825	0.046	The example of my parent (s) has had a positive influence on my motivation.
			Items removed during CFA.
B6: I am not the kind of person that knows mathematics well.			
B7: I am not good in Mathematics.			
A7: Studying mathematics is boring.			
C1: Mathematics is a mechanical and boring subject.			

(continued)

Table 1 (continued)

Item	Factor loadings	*SE*	Item wording	
Factor correlations				
	1	2	3	4
1	1.000			
2	0.423 (0.026)	1.000		
3	0.535(0.024)	0.232(0.030)	1.000	
4	0.260(0.033)	0.248(0.037)	0.191(0.034)	1.000

NB: All significant at p <0.001. These results are based on the metric invariance model. Factor loadings are unstandardized estimates. Correlations are constrained to be equal across calibration and validation sample. For model identification, all factor variances were fixed at 1. VOM = students' views on mathematics. For the correlations, parenthesis are standard errors (*SE*), items with R in front are reversed coded

confidence: $\alpha=0.690$; $\omega=0.697$), (*family encouragement*: $\alpha=0.619$; $\omega=.621$) as reported in exploratory research and adapted constructs such as those being used in our present study (Hair et al. 2010).

Results for ω and α were roughly the same, with α values slightly underestimating the *teacher quality, self-confidence, family encouragement* constructs and overestimating the *self-concept* construct. The lower reliabilities we obtained may imply substantial error of measurement and/or limited true individual differences, hence may attenuate the validity of interpretations based on manifest scale scores, weaken statistical power, and effect sizes (Raykov 2012; Schmitt 1996). It is thus advisable to base any comparisons on latent-variable models that account for the unreliability and measurement errors as suggested by Marsh and colleagues (2012).

Discussion

The factorial validity of the Finnish View of Mathematics (VOM) instrument was tested on a sample of Ghanaian twelfth-grade students. The original seven-factor model fitted the data poorly. In addition, further exploratory factor analysis (EFA) revealed that the Ghanaian data can best be explained by a four-factor structure. The alternative factorial structure was validated, further refined and then cross-validated with an independent sample from the Ghanaian data using a confirmatory factor analysis (CFA) approach. Moreover, measurement of invariance was established at the configural, metric and structural parameter levels. In respect of the new four-factor model, two scales from the Finnish model were partly confirmed and one other scale fully confirmed. The reliability values of the new scales were not very much higher than the values of the original scales. These may be due in part to the brevity of scale consisting of only three or four items. However, the overall fit of the four-factor model was significantly better than the fit of the original seven-factor model. We do know and attest to the use of Cronbach's alpha reliability (α) as a non-dependable general index of reliability for multidimensional scales, irrespective of whether their component errors are correlated

(Raykov 2012). Cronbach's alpha underestimated the reliabilities when there was no error term and overestimated the reliability when there was an error term within the construct. Moreover, because α is sensitive to the number of items in a scale, it underestimated the *family encouragement* and the *self-confidence* constructs. The reliabilities for the Ghanaian sample were generally acceptable.

The differences between the Finnish and Ghanaian models were interesting. In the Finnish samples, the scales for *ability, enjoyment* and *effort* were discrete, which provided support for separating emotions and motivation from more cognitive beliefs (Roesken et al. 2011). In Ghana, *ability*, *success,* and *enjoyment* factors were loaded on to the same factor and therefore can be treated empirically as one construct. The Ghanaian *self-concept* scale included three *ability* items, one *effort* item, four items from the *enjoyment* factor, one item from the *difficulty* factor, one item from the *success* scale and two items (item A10 and A19) that had not originally been part of the factorial structure of the Finnish samples. The *effort, enjoyment,* and *difficulty* factors were not confirmed in the Ghanaian data. Similarly, a study by Kaldo and Hannula (2012) failed to confirm the reliability for the scale of *effort* in a sample of Estonian university students. Hence, it is possible that the separate scale of *effort* is a characteristic feature of Finnish students.

The observed differences in the factor structures between the Ghanaian and the Finnish sample are not necessarily indicative of cultural variation, since such differences may be due to one or more measurement artifacts unrelated to the constructs. Indeed, careful reading of the content associated with each deleted item-pair reveals a strong content overlap. We do admit that in scale construction, it is important to look for items that are highly inter-correlated in order to establish a high degree of internal consistency and reliability. Nevertheless, in our study there was a translation issue, whereby originally two Finnish terms that emphasized different aspects of boredom (unpleasant versus tiresome) were both translated as 'boredom', which reduced the necessary variation between the different items. This finding highlights that item translation is critical when implementing an instrument into a new linguistic setting.

Also, this study has confirmed early studies (Edelen and Reeve 2007) that when items are measuring the same construct and are negatively worded, placing both items next to each other in a questionnaire can lead to local dependency—the response of the items are based on the response of each other.

The psychometric approach in Finland had been PCA, which technically is not able to determine and evaluate measurement error or to indicate alternative model specifications. Using Structural Equation Modeling (SEM) helped us to detect measurement error and bias, while also understanding the disparities. Although the identified factor structure does not conform to the original seven-factor structure of the VOM, it does reflect those important aspects of students' belief systems defined by McLeod (1992) and by Op't Eynde and colleagues (2006). In the light of the extremely stringent approach used to test the validity, the new 25-item scale proved to be quite sound from a psychometric perspective and it theoretically supports many of the dimensions suggested in the literature (McLeod 1992; Op't Eynde et al. 2006; Roesken et al. 2011). However, our results also indicate that important cultural differences are not restricted to how strongly students hold different beliefs, but also to what the actual belief constructs are.

We also conclude that, differences between the present study and previous studies on VOM reflect genuine differences in how the Finnish and Ghanaian upper secondary students' views of mathematics are structured. Realistically, given the dramatic cultural differences between these two countries it was not surprising that the mathematical self-concept among Finnish students seems to have an underlying structure, whereas for Ghanaian students the construct is a single entity. However, the findings can be partially attributed to the choice of statistical analyses. For example, in the present study, by the combined use of EFA and CFA we assumed that the best and correct factor structure for the Ghanaian sample was identified. The original approach in Finland applied PCA and less robust approaches to identify the factor structure. We do expect our conclusions to be more reliable than the results from previous studies in which these important modeling considerations were neglected.

Being able to independently validate the factor structure in two independent samples and for both genders, allowed us to conclude that (a) there is strong empirical support for a new four factor structure, (b) the same variables define each factor across all subsamples (c) all the latent variables have the same relationship within the sample and any differences in the covariance between the measured variables are due to the common factors. In addition, the finding that the newly formed four-factor model supports metric invariance across students' gender as well as single-sex and coeducational schools (Bofah and Hannula submitted) increases its value as an assessment instrument. This indicates that the instrument educes responses to questions that are being asked in the same way by different groups within a sample.

To summarize, the current study supports the measurement and structural invariance of VOM, as measured in the Ghanaian sample, across student gender, and suggests that further mean comparisons within the belief constructs can be interpreted as representing the underlying mean differences in the Ghanaian data.

Moreover, the present study has shown that, translation of a construct into a different language is more than just producing a text in another language. Knowing the linguistic and cultural differences can help reduce the problems associated with responses to translated adapted constructs. In addition, this paper has raised three important issues educational researchers face when they adopt and validate a construct cross-culturally. The three issues discussed below have been similarly argue for and discussed in Geisinger (1994) and Lin et al. (2005):

First, adaptation issue: an important question that researcher need to ask is, "Does a given construct need to be adapted?" Reliability estimates have been used as a yardstick to circumvent this question. Often or not, researchers adopt survey construct in a new cultural setting because it has a strong reliabilities in the original setting. Although, this issue is not problematic when no marked differences exist between the original population and the target population. Translation is needed when administering a survey instrument to respondents who speak another language either than the language used in original setting. This is where the cultural differences as well as the linguistics of the original and the target populations need to be taken into account. Moreover, a more difficult issue concerning test adaption is subpopulation within a given nation as well as cultural adaption within a single language. For instance, Ghana has five strong ethnic groups with over 100 linguistic and cultural groups within these five ethnic groups (Bodomo 1996). Ghana has adopted English as their national language; will an adopted Western construct translated into the national Language be adapted for use across the whole nation. This "… can be a difficult question, sometimes with more complex answers" (Geisinger 1994, p. 305).

Second, construct validity: this issue mainly deals with the question: "Does the construct measures what it intended purposes were in the new language or culture settings?" To answer this question, the construct validation and the reliability should be demonstrated. This can help establish if the assess construct have the same meaning in new target population. This is an important issue when the new population differs from the original in terms of cultural development. Researchers are encourage to use a more robust methodology such as SEM to validate constructs in cross-cultural research. This can help reduce method effects such as construct bias, method bias, and item bias associated with cross-cultural research (Lin et al. 2005; Sass 2011; Van de Vijver and Leung 2000). This can also help detect problems such as content overlap, item local dependency, and acquiescence.

Third, interpretation issue: importantly, after adapting and validating the instrument, how to interpret the scores of outcome on the new target population, i.e., "what do scores on the adapted measure mean?" Does the outcome support the literature meaningfully? Were the results driven by "acquiescence or substantive cultural differences?" Are the construct different across cultures due to religion and method effect? Can there be any referent group effect?—(see Marsh 2007). As discussed earlier, cultural and linguistic differences, can lead to different interpretations. Thus, the construct and instrument comparability across the cultures should be examined critically before giving interpretations (Lin et al. 2005).

One potential limitation of the study is that 12 items were removed from their designated factors because of model misfit and dimensionality concerns. In addition, two new items were included in the final model. This indicates that the rotation criterion used and how the factor analysis (i.e., EFA or CFA) is parameterized can significantly alter construct correlations and loadings/cross-loadings (Sass 2011; Sass and Schmitt 2011). For these reasons, Sass and Schmitt (2011, p. 301) urge, "model specification, modification, and verification decisions should be made judiciously and researchers must be cognizant of how the modeling approach influences the statistical and theoretical conclusions" (see also, Jöreskog 1993). Although our purpose was not to refine previous measures, implementing this modification should benefit future research using these scales. Another limitation is that the multilevel nature of the data was not taken into account. Because the data were derived from students in intact classes (students' in schools), they are inherently hierarchical. A hierarchical model could have helped us answer the question of whether a particular construct has the same meaning at the individual and classroom levels. Ignoring this nested structure can give rise to problems of bias within-group homogeneity (Fraser 1998). We do think the clustering effect in this study is negligible due to the number of schools involved in the study. A final limitation of this study was that all data were from self-reports and thus subject to social desirability biases.

The outcome of this chapter is one of the indications of the problems associated with the importation of Western instruments into non-Western countries. We conclude that, cross-cultural educational researchers should be conscious of the problems of construct importation and adaptation, such as, item translation—content overlap, acquiescence, reading difficulty, reverse items, similarly worded items, items that are presented sequentially, construct, method, and item bias that could affect the results of studies. We believe it is important that cross-cultural educational researchers acquire both a theoretical understanding of these issues and a practical ability to address them using *MPLUS* or some other SEM software. Furthermore, cross-cultural educational researchers should pay attention to the construct validity and interpretation of the study outcome. Failing, importation of Western constructs into non-Western countries may lead to inferences that are not valid.

This research has laid a solid foundation for future mathematics belief research in Ghana by making readily available a selection of valid, reliable and applicable questionnaires for researchers, teachers and policy makers.

References

Ainley, J., Kos, J., & Nicholas, M. (2008). *Participation in science, mathematics and technology in Australian education* (ACER research monographs). Retrieved from http://research.acer.edu.au/acer_monographs/4

Asante, K. O. (2012). Secondary students' attitudes towards mathematics. *IFE PsychologIA, 20*(1), 121–133.

Bodomo, A. (1996). On language and development in Africa: The case of Ghana. *Nordic Journal of African Studies, 5*(2), 31–51.

Bofah, E. M., & Hannula, M. S. (2014). Structural equation modelling: Testing for the factorial validity, replication and measurement invariance of students' views on mathematics. In *SAGE research methods cases*. London: SAGE Publications, Ltd. doi:http://dx.doi.org/10.4135/978144627305014529518

Bofah, E. A., & Hannula, M. S. (submitted). Students' views on mathematics in single-sex and coed classrooms in Ghana. *International Journal of Mathematical Education in Science and Technology*.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). London: Guilford.

Byrne, B. M. (1991). The Maslach Burnout Inventory: Validating factorial structure and invariance across intermediate, secondary, and university educators. *Multivariate Behavioural Research, 26*(4), 583–605.

Byrne, B. M. (1993). The Maslach Burnout Inventory: Testing for factorial validity and invariance across elementary, intermediate and secondary teachers. *Journal of Occupation and Organizational Psychology, 66*, 197–212.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge.

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*, 1005–1018.

Cheung, F. M. (1996). The assessment of psychopathology in Chinese societies. In M. H. Bond (Ed.), *Handbook of Chinese psychology* (pp. 393–411). Hong Kong: Oxford University Press.

Curran, P. J., West, S. G., & Finch, J. F. (1995). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*, 5–18.

Eshun, B. A. (2004). Sex-differences in attitude of students towards mathematics in secondary schools. *Mathematics Connection, 4*, 1–13.

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales. *JSAS Catalogue of Selected Documents in Psychology, 6*, 31.

Fraser, B. J. (1998). Science learning environments: Assessment, effects and determinants. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 527–564). Dordrecht: Kluwer.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*(4), 304–312.

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.

Ghana Education Service. (2013). *Educational statistics: Secondary education statistics.* Accra: Ghana Education Service. Available from: http://www.ges.gov.gh. Accessed on 11 Oct 2013.

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science, 23*(4), 519–529.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River: Prentice Hall.

Hannula, M. S. (2012, July 8–15). *Emotions in problem solving*. A paper presented in 12th international congress on mathematical education, COEX, Seoul, Korea. Available at: http://www.icme12.org/upload/submission/1983_F.pdf

Hannula, M. S., & Laakso, J. (2011). The structure of mathematics related beliefs, attitudes and motivation among Finnish grade 4 and grade 8 students. In B. Ubuz (Ed.), *Proceedings of the 35th conference of the international group for the psychology of mathematics education* (Vol. 1). Ankara: PME.

Hannula, M. S., Kaasila, R., Laine, A., & Pehkonen, E. (2005). The structure of student teacher's view of mathematics at the beginning of their studies. In M. Bosch (Ed.), *Proceedings of the fourth congress of European research in mathematics education (CERME 4)* (pp. 205–214). Sant Feliu de Guíxols: Fundemi IQS – Universitat Ramon Llull.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393–416.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & L. J. Scott (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park: Sage.

Kaldo, I., & Hannula, M. S. (2012). Structure of university students' view of mathematics in Estonia. *Nordic Studies in Mathematics Education, 17*(2), 5–26.

Kloosterman, P., & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics, 92*, 109–115.

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.

Lin, Y. H., Chen, C. Y., & Chiu, P. K. (2005). Cross-cultural research and back-translation. *The Sport Journal, 8*(4), 1–5.

MacCallum, R. C., Rosnowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490–504.

Malik, K. (2013). *Human development report 2013*. New York: United Nations Development Programme.

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester: British Psychological Society.

Marsh, H. W., Parada, R. H., & Ayotte, V. (2004). A multidimensional perspective of relations between self-concept (self description questionnaire II) and adolescent mental health (youth self-report). *Psychological Assessment, 16*(1), 27–41.

Marsh, H. W., Muthén, B., Asparouhov, A., Lüdtke, O., Robitzsch, A., Morin, A. J., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439–476.

Marsh, H. W., Liem, G. A., Martin, A. J., Morin, A. J., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment, 29*, 322.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, C. K., et al. (2012). 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.

Mason, C. H., & Perreault, W. D., Jr. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research, 28*(3), 268–280.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York: Macmillan.

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.

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1–13. doi:10.1007/BF02289400.

Nurmi, A., Hannula, M. S., Maijala, H., & Pehkonen, E. (2003). On pupils' self-confidence in mathematics: Gender comparisons. In N. A. Pateman, B. J. Dougherty, & J. Zilliox (Eds.), *Proceedings of the 27th conference of the international group for the psychology of mathematics education 3* (pp. 453–460), Hawaii, USA.

Nyala, J. I. (2008). Sex-differences in attitude towards mathematics of junior high school students in Ghana. *Edo Journal of Counselling, 1*(1), 137–161.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396–402.

Official Statistics of Finland. (2013). *Upper secondary general school education*. Helsinki: Statistics Finland. Available from: http://www.stat.fi/til/lop/2012/lop_2012_2013-06-12_tie_001_en.html. Accessed on 11 Oct 2013.

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2002). Framing students' mathematics-related beliefs: A quest for conceptual clarity and a comprehensive categorization. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 13–37). Dordrecht: Kluwer Academic Publishers.

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2006). Epistemic dimensions of students' mathematics-related belief systems. *International Journal of Educational Research, 45*, 57–70.
Pietilä, A. (2002). The ile of mathematics experiences in forming pre-service elementary teachers' views of mathematics. In A. Cockburn & E. Nardi (Eds.), *Proceeding of the 26th conference of the international group for the psychology of mathematics education* (pp. 57–64). Nowich: University of East Anglia.
Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472–492). New York: The Guilford.
Raykov, T., & Marcoulides, G. A. (2010). *Introduction to psychometric theory* (pp. 37–58). New York: Taylor & Francis.
Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373.
Richardson, F., & Suinn, R. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology, 19*, 551–554.
Roesken, B., Hannula, M. S., Pehkonen, E., Kaasila, R., & Laine, A. (2007). Identifying dimensions of students' view of mathematics. In *Proceedings of the fifth congress of the European Society for Research in Mathematics Education* (*CERME-5*) (pp. 349–358), Larnaca, Cyprus.
Roesken, B., Hannula, M. S., & Pehkonen, E. (2011). Dimensions of students' view of themselves as learners of mathematics. *ZDM – International Journal of Mathematics Education, 43*, 497–506.
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.
Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*, 299.
Sass, D. A., & Schmitt, T. A. (2011). Introduction to the special issue: Moving beyond traditional psychometric approaches. *Journal of Psychoeducational Assessment, 29*, 299.
Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350–353.
Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29*, 304–321.
Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*, 95–113.
Tapia, M., & Marsh, G. E. (2000). Attitudes toward mathematics instrument: An investigation with middle school students. *Academic Exchange Quarterly, 8*, 28–44.
UNESCO Institute for Statistics. (2011). *Global education digest 2011: Special focus on secondary education*. Montreal: UNESCO Institute for Statistics.
Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*(1), 33–51.
Van De Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto, & F. J. R. Van De Vijver, *Cross-cultural research methods in psychology* (pp. 17–45). Cambridge: Cambridge University Press.