A Computer Adaptive Measure of Reading Motivation

Marcia H. Davis
Johns Hopkins University
marcy@jhu.edu

Wenhao Wang
University of Kansas

Neal M. Kingston
University of Kansas

Michael Hock
University of Kansas

Stephen M. Tonks
Northern Illinois University

Gail Tiemann
University of Kansas

To appear in *Journal of Research in Reading*:
Davis, M. H., Wang, W., Kingston, N. M., Hock, M., Tonks, S. M., & Tiemann, G. (2020). A computer adaptive measure of reading motivation. *Journal of Research in Reading*. https://doi.org/10.1111/1467-9817.12318

Journal of Research in Reading peer review process: Manuscripts submitted to this journal undergo editorial screening and peer review by anonymous reviewers.

Corresponding Author:
Marcia H. Davis (marcy@jhu.edu)
2800 North Charles Street, Suite 420
Center for Social Organization of Schools, The Education Building
Johns Hopkins University, Baltimore, MD 21218,

Authors' Note
This article is based on data published in Kingston et al. (2018) and Davis et al. (2017). We received funding for this work from the Institute of Education Sciences, under Grant Number R305A110148. The present research was also supported by the Center for the Interdisciplinary Study of Language and Literacy at Northern Illinois University and a grant awarded to the fifth author by the Institute of Education Sciences (R305A150193). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Institute of Education Sciences. Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Abstract

Background: The importance of reading motivation has led to the development of a large number of self-report reading motivation measures; however, there is still a need for a usable measure of adolescent reading motivation that captures a large number of theoretically and empirically distinct constructs.

Methods: The current paper details the development and validation of a computer adapted measure of reading motivation, the Adaptive Reading Motivation Measure (ARMM), which assesses constructs of curiosity, involvement, interest, value, challenge, grades, recognition, competition, avoidance, self-efficacy, perceived difficulty, preference for autonomy, social motivation, prosocial goals, and antisocial goals for reading.

Results: Model fit indicated that hierarchical multidimensional models fit better than models without a hierarchical structure. The validation results indicate that females scored higher than males and younger students scored higher than older students on most ARMM scores when scores were derived using a higher-order model. In addition, these scores correlated significantly to reading behavior, engagement, and achievement and indicated high reliability.

Conclusions: The findings suggest that the ARMM would be a valid measure to assess a large number of reading motivation constructs in a short period of time within a classroom setting.

Keywords: Reading Motivation, Measurement, Adolescents, Validity, Computer Adaptive

What is already known about this topic

- Motivation to read is considered a critical contributor to reading achievement.
- Although there are a number of adolescent reading motivation scales, most of these scales only measure a few reading motivation constructs.

What this paper adds

- The paper describes the development of the ARMM, which measures 15 separate constructs as well as a general reading motivation construct and, due to the computer adaptive nature, is only 45 items long.
- Findings show that the ARMM scores were sensitive to gender and grade differences consistent with prior reading motivation research and correlated significantly to reading behavior, engagement, and achievement.

Implications for theory, policy or practice

- Being able to assess a large number of constructs could provide useful information to teachers implementing reading interventions and improving instruction.
- The ARMM was developed for fifth through twelfth grade, which would facilitate grade comparisons in research studies.

A Computer Adaptive Measure of Adolescent Reading Motivation

Motivation to read is considered a critical contributor to reading achievement (Retelsdorf et al., 2011; Schiefele et al., 2012). If students lack the motivation to engage in reading, reading improvement will be limited (Guthrie & Wigfield, 1999) and may actually decline (Baker & Wigfield, 1999; Unrau & Schlackman, 2006). In the end, it is motivation that activates the behavior to engage in reading, making motivation an important factor in efforts to improve literacy (Guthrie & Wigfield, 2000). The importance of reading motivation has led to the development of a large number of self-report reading motivation measures, as described in a recent review (Davis et al., 2018). Although there are many reading motivation measures for elementary school students, measures for adolescents have only been developed recently, and many measure only a few reading motivation constructs (Davis et al., 2018). There is a need for a usable measure of adolescent reading motivation that captures a large number of theoretically and empirically distinct constructs.

Theoretical Perspectives

In the review of reading motivation scales Davis et al. (2018) found that while quite a few scales of reading motivation were directed by one (De Naeghel et al., 2012) or even multiple theories of motivation (Wigfield & Guthrie, 1997), there were still a large number of scales that had no theory identified. To add to the confusion, although quite a few different theoretical perspectives have driven item development for these scales, items appear similar despite having different construct labels (Davis et al., 2018; Neugebauer & Fujimoto, 2018). Like Guthrie and Coddington (2009) we believe that focusing on only one theory of motivation may limit the scope and multidimensionality of a measure. Our understanding of reading motivation derives from several theories including self-determination theory (Ryan & Deci, 2000), achievement goal theory (Meece et al., 2006), expectancy-value theory (Wigfield & Eccles, 2000), social cognitive theory (Schunk, 2003), and interest development theory (Hidi & Renninger, 2006). We define reading motivation as "students' goals, values, beliefs, and dispositions towards reading" (Guthrie et al., 2013, p. 10), which implies that reading motivation is multidimensional and is based within many different theories related to goals, values, and beliefs.

Adolescent Reading Motivation

Reading motivation research has indicated that reading motivation declines over time (Schaffner et al., 2016). The decline in intrinsic reading motivation may be related to the typical reading practices of secondary schools such as less reading instruction in content area classes, lack of choice in reading, poorer personal connections with teachers, less connection with reading and real-world interactions, and more complex texts compared to elementary school (Guthrie & Davis, 2003). Due to this decline, it could be argued that it is important for secondary teachers to monitor reading motivation in their classrooms. However, assessing reading motivation through observation is difficult, especially in secondary schools where teachers see students for only a short time (Guthrie & Davis, 2003). Measuring engagement and motivation in reading can be difficult and time-consuming even for trained observers (Lutz et al., 2006; Neugebauer, 2016).

A dynamic adolescent reading motivation measure could help teachers examine the nuances of reading motivation and determine interventions that could target specific constructs of reading motivation. However, out of the 16 measures reviewed by Davis et al. (2018) and additional two published after the review, only seven, which the Adapted Reading Motivation Measure (ARMM) is included, were written specifically for adolescent students. Of these seven adolescent reading scales, only three measure a wide range of motivational concepts; however, one of these three can only be used to measure reading of non-fiction texts of middle school students, which can limit its use. Also, only three of the seven scales were developed for both middle and high school students. Limiting to only a few grades may make comparisons between grades or longitudinal studies over a series of grades more difficult. Further, only two of the scales measured extrinsic motivation. Although elementary studies indicate that extrinsic motivation may not correlate as highly to engagement and achievement compared to intrinsic motivation (Wang & Guthrie, 2004), as intrinsic motivation decreases with age (Lepper et al., 2005; Schaffner et al., 2016), extrinsic motivation may play a larger role in motivating reluctant adolescent readers. Finally, social motivation can be highly important for adolescents (Moje et al., 2008); however, it is only measured by two of the measures.

Computer Adaptive Testing

One way to include more constructs on a measure without increasing the number of items is by using computer adapted technology. Computer adaptive measures use Item Response Theory (IRT) to select items for each respondent based on their previous answers, so that each respondent only has to answer a small subset of available items. In traditional measures, all respondents answer the same items, which makes them longer. Although the use of adaptive measures for questionnaire development is well established (e.g., Edelyn & Reeve, 2007) there are no computer adaptive measures of reading motivation (Davis et al., 2018).

The Current Study

The goal of the current paper was to describe the development and large-sample validation of the Adaptive Reading Motivation Measure (ARMM), an adaptive adolescent reading motivation survey that assesses fifteen separate reading motivational constructs. The development process was a multi-stage process, which included an item-writing stage, a pilot test, a large field test, and a validation study. In this paper the development process is explained and validation findings are be presented and discussed. The following questions in the validation study were addressed:

1. Which of four models, varying on degree of multidimensionality and number of hierarchical levels, fit the ARMM data the best?
2. Were the ARMM scores as measured by the computer adapted version reliable?
3. Were there differences between male and female students on the ARMM, and if so, did those differences align to previous research?
4. Were there differences between younger and older students on the ARMM, and if so, did those differences align to previous research?
5. Did the ARMM scores correlate with measures of reading behavior, engagement, and achievement? Were these correlations higher than correlations with math achievement?

Method

Participants

Development and Pilot Test

In the pilot test we administered items to 2,258 fifth through twelfth students from 32 schools in the Midwest and West Coast United States. At the school level there was an average of 76.2% white students, 3.0% black students, 12.0% Hispanic students, and 3.7% Asian students across the participating schools. In addition, there was an average of 41.8% of students receiving free or reduced meal prices across the schools.

Field Test and Cognitive Interviews

Participating in the field test were 7,457 public school students recruited from different research and teaching networks in the United States (813 fifth grade, 1,428 sixth grade, 1,160 seventh grade, 1,090 eighth grade, 1,355 ninth grade, 563 tenth grade, 576 eleventh grade, 413 twelfth grade students, and 59 students who did not identify their grade). Self-identified gender included 3,030 males and 2,711 females; 1,716 students gave no response in regards to gender. At the school level there was an average of 36.5% white students, 34.0% black students, 21.0% Hispanic students, and less than 1% other across the 209 participating schools. In addition, there was an average of 59.5% of students receiving free or reduced meal prices across the schools. At the same time as the field test cognitive interviews with students from two elementary, one middle, and one high school (28 girls, 25 boys) were conducted.

Validation Study

Participating in the validation study were 1,905 students from 43 schools located in the Midwestern United States (720 fifth-grade, 1,046 sixth- to eighth-grade, and 139 high school students). Of these students, 1.8% were Black, 0.6% were Asian, 4.0% were Native American, 93.1% were White, and 0.5% were other. Each student took the reading behaviors, engagement, and ARMM scales. One participating district provided achievement data for 605 students in the fifth grade and 287 in sixth to eighth grade.

Measures

Item Development

A goal of the ARMM developers was to measure a wide range of reading motivation constructs; therefore, the team systematically reviewed reading motivation measures over the last 25 years and consulted with reading motivation experts in order to build a comprehensive list of constructs. In their review of past measures, the team found that some constructs, like self-efficacy and self-concept, were too similar at the individual item-level to warrant separate constructs, and therefore these constructs were collapsed into one scale for the ARMM. See the final construct list in Table 1 and six-point scale in Figure 1.

Seven middle school and six high school teachers with expertise in reading and language arts instruction were recruited to attend a summer item-writing workshop. ARMM project staff, including principal investigators and consultants, and an author of the Motivation for Reading Questionnaire, also attended. The workshop began with an overview of the hypothesized sub-constructs of adolescent reading motivation. The teachers received a document with a definition of each construct, sample items, and an overview of item-writing procedures. Working in pairs, members of the panel then wrote over 700 items. Given the size of the item pool, the ARMM staff created a text mining program to calculate the proportion of identical words in every possible pair of items and deleted the items that were too similar. Additionally, an experienced test-item editor revised the items for clarity and reading level.

Pilot Test. ARMM staff selected 600 items (40 for each of the 15 factors) for inclusion in the pilot study. A total of 10 different basic forms of items were specified, using a blocked design, with each individual form containing 20 unique items from each of three constructs, or 60 total items per form. Across all the forms, each of the 15 constructs was presented twice; thus each construct was represented by 40 unique items, for a total of 600 items. Students who participated in the pilot were randomly assigned to an assessment form on their school computers.

Field Test

Classical test theory item statistics from the pilot test were used to select a final pool of 20 items per construct (300 items in total) for a model comparison study. These items had higher item total correlations and non-extreme item average scores based on pilot data. A total of 10 forms containing the 300 unique items (20 items for each of the 15 constructs) were administered to public school students on their school computers. In order to collect enough student responses to calibrate for IRT models, sparse-matrix design was used to build the test form. Sparse-matrix design is a calibration data collection involving items overlapping across forms. The resulting ten forms were named Form A, Form B, Form C, … , Form J. Each form had 60 items, four items for each of the 15 constructs. All items appeared on two different forms, but each student only had to respond to 60 items on one form. This item overlapping across test forms allows items to be calibrated on a common set of IRT metrics and also collects enough data without students taking all 300 items at one time.

Model Comparisons. A confirmatory IRT model comparison method was used to compare models generated based on different construct relationship assumptions. Four either unidimensional or multidimensional graded response models (Samejima, 1969) were calibrated. In the unidimensional model only one general reading motivation dimension is extracted from the data. In the second model, the multidimensional model, fifteen construct factors were allowed to correlate with each other and items load on one of these fifteen. In the third model, the higher order model, the fifteen construct factors correlated with the general factor directly and correlated with each other indirectly through the relationships with the general factor. Finally, the last model was a bi-factor model (Gibbons & Hedeker, 1992), which allows for a general factor as well as multiple secondary factors. However, unlike the higher-order model, the fifteen construct factors do not correlate with the general factor. In the current paper, the fifteen construct factors using the bi-factor model did not correlate with each other and the covariance of all sixteen latent factors were set to zero.

Models were estimated using IRTPRO 2.1 (Cai et al., 2011), which uses the expectation–maximization (EM) algorithm (Bock & Aitkin, 1981) and the Metropolis–Hastings Robbins–Monro (MH-RM) algorithm (Cai, 2010). Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used to compare models. AIC and BIC are often used to choose among non-nested models, which one cannot do with regular fit indices like CFA.

Cognitive Interviews. Cognitive interviews with students from two elementary, one middle, and one high school (28 girls, 25 boys) were conducted using an approach developed by Karabenick et al. (2007) to examine if students' interpretation of the ARMM items matched with the research team's intended meanings (Tonks et al., 2014). Information from the cognitive interviews and IRT item discrimination parameters from the field test was used to further limit the number of items.

Adaptive Reading Motivation Measure

Findings from a simulation study (Wang & Kingston, 2018) indicated that a fixed length hierarchical IRT adaptive test with medium test length, administering 3 items per construct, could provide an accurate estimation of student scores. Since the ARMM has 15 constructs, the final adaptive measure would administer 45 items, 3 items per construct. In order to improve the quality of items administered in the ARMM, only 12 out of 20 items field tested per construct were selected to form the adaptive test item pool (see sample items in Table 1). These 12 selected items per construct all had high IRT item discrimination parameters. However, for the grades construct the number of items with high IRT item discrimination parameters was very small, so only 6 items were kept for the adaptive item pool. The adaptive algorithm for ARMM only used the bi-factor secondary level item parameters calibrated in the field test to select items in order to speed the item selection and scoring process (Wang & Kingston, 2018).

During administration of the ARMM students are first shown one set of 15 items, one for each construct. Like in computer adaptive achievement tests, the items selected for the first round are selected to measure moderate levels of motivation. After a student takes all 15 items, the computer displays a new set of 15 items, one for each construct, that are selected based on the student's responses from the first set. Thus, if students seem to be answering on the higher end of the reading motivation continuum, the student would receive items that differentiate more on the higher end of the continuum ("Getting good grades in reading is important to me" and "I enjoy reading about topics that interest me"). Likewise, if students seem to be answering on the lower end of the reading motivation continuum, the student would receive items that differentiate more on the lower end of the continuum ("I find ways to avoid reading in class" and "I make fun of students who like to read"). A third set is selected based on the student's responses to the first two sets.

The theta scores derived from the ARMM are decimal values that can include negative numbers. In order to make the scores more interpretable for teachers and researchers, we used a linear transformation (multiplying the scores by 16 and adding 50 points) that kept the rank and relevant ratings among students consistent, but changed the scores to positive integer numbers. Item discrimination parameters of three negative constructs (avoidance, perceived difficulty, and antisocial) were set to be negative (e.g. lack of avoidance) for the higher order model since that model assumes only positive correlation among constructs. Finally, the team used differential item functioning (DIF) analysis (Hidalgo & Lopez-Pina, 2004) to examine potential biases for grade or gender. No ARMM items showed gender or grade level DIF.

Reading Behavior

Ten items measuring reading behavior were administered to the students after they were administered the ARMM assessment. Four of the items, three of which were adapted from the PIRLS 2011 student survey (International Association for the Evaluation of Educational Achievement, 2011), asked students to rate how often they completed certain reading tasks such as reading for fun or reading for homework. Another four items, adapted from four items from the PIRLS 2011 student survey, asked students to rate how often they read certain material (in print or on an electronic device) outside of school such as magazines, nonfiction texts, novels, or comic books. Finally, two items asked how often students read outside of school on a computer or electronic device. Each item had response categories of “0 - Never or almost never, 1 - Once or twice a month, 2 - Once or twice a week, 3 - Every day or almost every day.” These 10 items had a Cronbach reliability of .71. The IRT graded response model (Samejima, 1969) was used to score the responses of these 10 items to get IRT scores for reading behavior.

Self-Reported Reading Engagement

Ten reading engagement items were administered to the students after they completed the reading behavior items. Skinner et al. (2009) define engagement as “the quality of a student’s connection or involvement with the endeavor of schooling and hence with the people, activities, goals, values, and place that compose it” (p. 494). They further distinguish between emotional engagement which reflects emotional states such as “enthusiasm, interest, and enjoyment” and behavioral engagement which reflects “effort, exertion and persistence” (p. 495). We applied this to a reading context and had students rate two emotional reading engagement items (“I am very excited when the teacher gives us reading to do”) and eight behavioral engagement items (“I read a lot during free reading time in class”). Answer choices were on a scale of “0 – Not at all like me” to “5 – Very much like me.” These 10 items had a Cronbach reliability of .89. The IRT graded response model (Samejima, 1969) was used to score the responses of these 10 items to get IRT scores for engaged reading.

Measures of Academic Progress (MAP)

MAP scores were obtained from one participating school district in the ARMM validation study and were used as measures of achievement in reading and mathematics. The *Measures of Academic Progress* is a computer adaptive achievement measure and was developed by the Northwest Evaluation Association (NWEA, 2003) in order to assess achievement level and growth in the areas of reading, language, math, and science. The reading MAP items are multiple choice and the assessment uses Rach Units that were developed by NWEA. According to the NWEA (2013, p. 5) the “numerical (RIT) value assigned to a student represents the level of test item difficulty at which he or she is capable of answering correctly approximately 50% of the time. The RIT scale is continuous across grades, making it ideal to track student achievement growth both within a school year and across adjacent school years.” Marginal reliability estimates (Green et al., 1984), used in IRT analysis, were high (.90 to .95) as was test-retest reliability (.76 to .91).

Results

Field test

The first research question asked, “Which of four models, varying on degree of multidimensionality and number of hierarchical levels, fit the ARMM data the best?” Findings related to fit can be seen in Table 2. Since lower AIC and BIC estimates indicate greater fit, the bi-factor model was shown to be the best fitting of all four models for both AIC and BIC as well as the log likelihood estimate, with the higher-order model coming in a close second. The worst fitting was the unidimensional model. These model fit results also indicate the construct relationships reflected by bi-factor or higher-order model, i.e. fifteen separate constructs and one general level, is closer to the true construct relationships than the one combined construct reflected by the unidimensional model or fifteen separate constructs without general level reflected by the multidimensional model. Due to these fit statistics, we decided to use both the bi-factor and higher-order models for the validation study with scores derived from the bi-factor model labeled ARMM-B and scores derived from the higher-order model labeled ARMM-H.

Reliability

The second research question asked, “Were the ARMM scores as measured by the computer adapted version reliable?” To answer this question we determined the reliabilities of the scores from the higher-order and bi-factor models. The marginal reliability (Green, et al., 1984), calculated as (true score variance – the marginal posterior error variance of measurement) / true score variance, is typically used in CAT IRT analysis (Wang, 2014) to provide a single value estimation of reliability. Thus, marginal reliabilities were used to estimate reliability for the ARMM-H and ARMM-B scores. Table 3 shows the reliabilities of the general and fifteen constructs measured with either a higher order or a bi-factor model. The reliabilities of the ARMM-H scores (.75 to .95) were generally higher than those of the bi-factor ARMM-B scores (.50 to .95) with the general scores for both were higher than all the other constructs. The lowest reliabilities were the bi-factor scores curiosity, value, interest, and involvement. Additionally, we examined the correlation between the ARMM-H and ARMM-B general scores and found it to be very high ($r = .994$, $p < .001$). Due to assumptions inherent in the hierarchical and bi-factor models, the ARMM-H construct scores correlated highly to the general ARMM-H score (average .79) and other ARMM-H construct scores (average .61), and the ARMM-B construct scores did not relate highly to the general ARMM-B score (average .045) and with the other

ARMM-B construct scores (average .02). On account of the low reliabilities of the fifteen ARMM-B scores, only the general ARMM-B score was used in further analysis.

Gender Differences

The third research question asked, “Were there differences between male and female students on the ARMM, and if so, did those differences align to previous research?” To answer this question, we examined differences in means among the female and male students for the ARMM-H scores and ARMM-B general score. Table 4 presents the means of the scores by gender. Females were significantly higher than males on all ARMM scores. Higher effect sizes were found for differences between genders on the two general scores, intrinsic reading constructs (involvement, value, and interest), grades, and social motivation. Those with the highest mean difference at the lower end of the confidence interval include social and antisocial motivation.

Grade Differences

The fourth research question asked, “Were there differences between elementary and secondary students on the ARMM, and if so, did those differences align to previous research?” To answer this question, we examined differences in means among elementary (Grade 5) and secondary (Grades 6-12) students for all of the ARMM-H scores and ARMM-B general score. Table 5 presents the means of the scores by elementary and secondary grades. Elementary students were significantly higher than secondary students on all ARMM scores except competition. Higher effect sizes were found for differences between grades on the two general scores, value, and anti-social motivation. Constructs with the highest mean difference at the lower end of the confidence interval include prosocial motivation and anti-social motivation.

Reading Behavior, Engagement, and Achievement

The fifth research question asked, “Did the ARMM scores correlate with measures of reading behavior, engagement, and achievement?” In addition, “were these correlations higher than correlations with math achievement?” The correlations of the ARMM-H scores and ARMM-B general score with reading behavior and engagement are on Table 6. All ARMM-H scores were significantly correlated to both reading behavior and engagement with correlations somewhat higher for secondary students than elementary students. The constructs with the highest correlation to behavior and engagement included both general reading motivation scores, intrinsic reading constructs (curiosity, challenge, involvement, value, and interest), and grades.

Table 7 presents correlations of the ARMM-H scores and ARMM-B general score with reading achievement scores. All ARMM-H scores were significantly correlated to reading achievement for eighth grade students with correlations ranging between .29 and .60. Constructs with high correlations to reading achievement (.3 or higher) across grades included both general reading motivation scores, challenge, involvement, grades, and difficulty. Looking across the grades there is variability on which constructs correlate highest in each grade. For example, difficulty reading texts was correlated the highest with reading achievement for the fifth and sixth grade students, but not in seventh and eighth grades. In seventh grade self-efficacy was just as high as difficulty reading texts. In eighth grade the autonomy construct was the most related to reading achievement.

Using regression analysis we examined the relationship of ARMM scores with reading achievement while controlling for math achievement. As can be seen in Table 8, reading achievement significantly predicted most of the reading motivation scores even when controlling for math achievement for students in grades five, seven, and eight. Math achievement did not significantly predict reading motivation scores when reading achievement was controlled. This finding indicates that reading achievement, and not math achievement, was uniquely associated with reading motivation.

Discussion

While some measures of adolescent reading motivation exist, there was a need for a flexible measure of adolescent reading motivation that could be used to assess a wide range of reading motivation constructs for both middle and high school students, but be feasible in length to use in a classroom setting. The ARMM measures 15 separate constructs as well as a general reading motivation construct and, due to the computer adaptive nature, is only 45 items long. The goal of the current paper was to describe the development and validation of the ARMM. We presented data regarding the structure and reliability of the measure, as well as examined gender and grade differences. In addition, we examined construct validity through relating scores from the ARMM with measures of reading engagement, behavior, and achievement.

Model Comparisons

In the development of the ARMM, four models (univariate, multivariate, higher order, and a bi-factor models) were examined for fit. Out of all the four models, the univariate (one general factor) had the worst fit, indicating that reading motivation measured by the ARMM is multidimensional. This aligns to past research on motivation measurement showing that multivariate models fit better than one-dimensional models (McKenna et al., 2012; Tunmer & Chapman, 1991).

The current study also examined two models that measure a hierarchical structure of reading motivation. Using a hierarchical structure allows the researcher to study general reading motivation while retaining the multidimensional nature of reading motivation. Not accounting for the multidimensional nature of reading motivation is the reason why univariate models have a poor fit compared to other models. In the current paper we examined two models that included a general reading motivation score in addition to fifteen sub-scores. These two models (higher order and bi-factor) fit better than the multidimensional model with no general reading motivation factor. This indicates that reading motivation constructs on the ARMM were related in a hierarchical structure. Due to the goodness of fit for both the bi-factor and higher-order models, and the similarity of the higher-order model to traditional measures, we decided to use scores from both in the validation study; however, due to low reliabilities of the scores for the ARMM-B model, only the general score was used in the analysis.

Validity Results of the Higher-Order Model

The examination into the validity of the higher-order model scores produced results similar to previous research. The ARMM-H scores had high reliabilities for all constructs. Females were significantly higher than males on all ARMM-H constructs, which aligns to past research regarding gender differences in reading motivation (Clark, 2011; Schaffner et al., 2013). Further, younger students scored significantly higher than older students on all but one of the ARMM-H scores, which also aligns to past research which found motivation decreases with age (Lepper et al., 2005; Unrau & Schlackman, 2006; Schaffner et al., 2016). Finally, these scores correlated significantly and positively to reading behavior, engagement, and achievement, which again relates to past research (Guthrie et al., 2007; Schaffner et al., 2016; Stutz et al., 2016). Most of the ARMM-H scores correlated significantly to behavior, engagement, and achievement for both elementary and secondary students. Most of the ARMM-H scores related to reading achievement even when controlling for math achievement, indicating discriminate validity. In addition, the ARMM-B general reading motivation score, correlated with each of these at the same magnitude and direction as the higher order model.

In examining extrinsic motivation in particular, the construct of motivation to receive good grades correlated to behavior, engagement, and reading achievement just as highly, if even higher than some of the intrinsic motivation constructs of involvement, value, interest, and challenge. For the eighth-grade participants, the motivation of receiving good grades correlated even higher than self-efficacy. Past research has indicated that secondary school teachers place a higher emphasis on performance goals, like getting good grades, than do elementary school teachers (Wigfield et al., 1998). This may translate to students also placing an emphasis on getting good grades. The finding of a positive and significant relationship of extrinsic motivation and reading behavior, engagement, and achievement is interesting since most of adolescent reading motivation scales do not measure extrinsic motivation constructs (Davis et al., 2018). The only other construct that correlated to achievement higher than grades for eighth grade students was autonomy. Research has shown that secondary school teachers often provide less choice than elementary school teachers (Guthrie & Davis, 2003), but it may be that adolescents may want more choice than they are currently receiving.

Usefulness of the ARMM to Teachers and Researchers

Due to the computer adaptive nature of the ARMM it can measure 15 different reading motivation constructs and general reading motivation with only three items per construct and still maintain high levels of reliability and predictive validity. The ARMM is quite different from other measures of reading motivation. First, although it has very similar items to the Motivation for Reading Questionnaire (MRQ; Wigfield & Guthrie, 1997), the ARMM was written by secondary teachers for secondary students, so, unlike the MRQ, the ARMM items would be inappropriate for ages below fifth grade. Second, although there are other adolescent reading motivation scales, most of these scales only measure a few reading motivation constructs (Davis et al., 2018). We argue that being able to assess a large number of constructs could help capture the multidimensional nature of adolescent reading motivation as well as provide useful information to teachers on implementing reading interventions. For example, teachers could use the ARMM to inform the development of interventions that align to what motivates their students to read (selecting their own texts, being social around reading) based on their ARMM scores. As teachers implement particular reading interventions, such as adding more collaboration around reading or helping students feel more competent reading the texts, they can use the ARMM to examine motivational changes over the course of a few months or the school year. Being able to see change in their students over time might encourage teachers to implement more reading interventions into their secondary classrooms. Finally, only a handful of adolescent reading motivation measures were developed for both middle and high school students and no other measure has been developed for fifth through twelfth grade. However, this range may be necessary for reading motivation researchers interested in examining grade comparisons and longitudinal studies of reading motivation, especially if the researcher wanted to examine a number of reading motivation constructs at the same time.

Limitations

There are several limitations of the ARMM. First, the ARMM was developed for teacher use, so it primarily measures general or academic reading motivation and not specifically motivation to read at home. In addition, the ARMM, like the MRQ, is a general measure of reading motivation and does not differentiate among school subjects, non-fiction and fiction reading, nor digital and print reading. Although it can be argued that general reading motivation measures may not pick up on individual contexts, both general reading motivation and specific reading motivation scores are highly related (Neugebauer, 2014). In the current study we found that the ARMM-H scores were highly related to the reading engagement measure, which asked students to report on engagement of reading both at home and school as well as the reading behavior measure that asked students to rate how often they read fiction, non-fiction, and on electronic devices.

One major limitation of this study was the lack of diversity of participants in the pilot and validation studies due to the availability of participants. It could be argued that the field test, which included large percentages of white, black, and Hispanic students, was the most influential step for the selection of items and calculation of item characteristics for the final computer adaptive measure. Therefore, the ARMM would be appropriate for use in schools with diverse students. However, findings regarding the relationships between the ARMM scores with engagement, behavior, and achievement might not generalize to all students.

Future Directions

In summary, we found the ARMM-H scores and ARMM-B general score from the ARMM to be reliable. These scores were sensitive to gender and grade differences consistent with prior reading motivation research. Further, the ARMM-H scores and ARMM-B general score correlated significantly to reading behavior, engagement, and achievement. Although these scores were sensitive to gender and grade differences, future research will need to be conducted to determine if these scores are sensitive to interventions related to reading motivation and in more diverse settings.

References

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly, 34*(4), 452–477. https://doi.org/10.1598/RRQ.34.4.4

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459. https://doi.org/10.1007/BF02293801

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika, 75,* 33–57. https://doi.org/10.1007/s11336-009-9136-x

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software]. Chicago, IL: Scientific Software International.

Clark, C. (2011). Setting the baseline: The National Literacy Trust's first annual survey into young people's reading - 2010. *National Literacy Trust*. https://eric.ed.gov/?id=ED541400

Davis, M. H., Tonks, S. M., Hock, M., Wang, W., & Rodriguez, A. C., (2018). A review of reading motivation scales. *Reading Psychology, 39*(2), 121-187. https://www.tandfonline.com/doi/full/10.1080/02702711.2017.1400482

Davis, M. H., Tonks, S. M., Hock, M., & Wang, W. (2017, April). *Validation of an adaptive measure of reading motivation* [Poster presentation]. American Educational Research Association, San Antonio, TX. https://www.aera.net/Events-Meetings/Annual-Meeting/Previous-Annual-Meetings/2017-Annual-Meeting

De Naeghel, J., Van Keer, H., Vansteenkiste, M., & Rosseel, Y. (2012). The relation between elementary students' recreational and academic reading motivation, reading frequency, engagement, and comprehension: A self-determination theory perspective. *Journal of Educational Psychology, 104*(4), 1006-1021. https://doi.org/10.1037/a0027800

Edelyn, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*(Suppl1), 5-18. https://link.springer.com/article/10.1007/s11136-007-9198-0

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*(3), 423-436. https://doi.org/10.1007/BF02295430

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347-360. https://doi.org/10.1111/j.1745-3984.1984.tb01039.x

Guthrie, J. T., & Coddington, C. S. (2009). Reading motivation. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 503–526). New York, NY: Routledge.

Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading and Writing Quarterly*, 19, 59-85. https://www.tandfonline.com/doi/abs/10.1080/10573560308203

Guthrie, J. T., & Wigfield, A. (1999). How motivation fits into a science of reading. *Scientific Studies of Reading, 3*(3), 199–205. https://doi.org/10.1207/s1532799xssr0303_1

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 403–422). Mahwah, NJ: Erlbaum.

Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, *32*(3), 282-313. https://www.sciencedirect.com/science/article/pii/S0361476X06000269?via%3Dihub

Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, *48*(1), 9-26. https://doi.org/10.1002/rrq.035

Hidalgo, M. D., & Lopez-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measuremen*t, 64, 903-915. https://doi.org/10.1177/0013164403261769

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111-127. https://doi.org/10.1207/s15326985ep4102_4

International Association for the Evaluation of Educational Achievement (2011). *PIRLS 2011 Student Questionnaire, Grade 4.* https://timssandpirls.bc.edu/pirls2011/downloads/P11_StuQ.pdf

Karabenick, S.A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., De Groot, E., Gilbert, M. C., Musu, L., Kempler, T. M. & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist 42,* 139-151. https://doi.org/10.1080/00461520701416231

Kingston, N., Wang, W., Davis, M. H., Tonks, S. M., Tiemann, G. & Hock, M. (2017). *Development of the adaptive reading motivation measures: Technical report.* Retrieved from the Center for Educational Testing and Evaluation website:

https://aai.ku.edu/sites/aai.ku.edu/files/docs/Technical_Reports/ARMM_Summary_Technical_Report.pdf

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology*, *97*(2), 184. https://doi.org/10.1037/0022-0663.97.2.184

Lutz, S. L., Guthrie, J. T., & Davis, M. H. (2006). Scaffolding for engagement in elementary school reading instruction. *Journal of Educational Research, 100*(1), 3-20. https://www.tandfonline.com/doi/abs/10.3200/JOER.100.1.3-20

McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: Results of a U.S. survey. *Reading Research Quarterly, 47,* 283–306. https://doi.org/10.1002/rrq.021

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487-503. https://www.annualreviews.org/doi/full/10.1146/annurev.psych.56.091103.070258

Moje, E., Overby, M., Tysvaer, N., & Morris, K. (2008). The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard educational review*, *78*(1), 107-154. https://doi.org/10.17763/haer.78.1.54468j6204x24157

Neugebauer, S. R. (2014). Context-specific motivations to read for adolescent struggling readers: Does the Motivation for Reading Questionnaire tell the full story? *Reading Psychology*, *35*(2), 160-194. https://doi.org/10.1080/02702711.2012.679171

Neugebauer, S. (2016). Stable or situated understandings of adolescent reading engagement across readers and raters. *The Journal of Educational Research*, *109*(4), 391-404. https://doi.org/10.1080/00220671.2014.968914

Neugebauer, S. R., & Fujimoto, K. A. (2018). Distinct and overlapping dimensions of reading motivation in commonly used measures in schools. *Assessment for Effective Intervention*, https://doi.org/10.1177/1534508418819793

Northwest Evaluation Association (2003). *Technical manual for the Northwest Evaluation Association (NWEA) Measures of Academic Progress and Achievement Level Tests.* Portland, OR: Author.

Northwest Evaluation Association (2013). Measures of Academic Progress: A comprehensive guide to the MAP K – 12 computer adaptive interim assessment. https://www.nwea.org/content/uploads/2014/07/Comprehensive-Guide-to-MAP-K-12-Computer-Adaptive-Interim-Assessment

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, *21,* 550–559. https://doi.org/10.1016/j.learninstruc.2010.11.001

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55,* 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Samejima, F. (1969). *Estimation of a latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. http://www.psychometrika.org/journal/online/MN17.pdf

Schaffner, E., Philipp, M., & Schiefele, U. (2016). Reciprocal effects between intrinsic reading motivation and reading competence? A cross-lagged panel model for academic track and nonacademic track students. *Journal of Research in Reading*, *39*(1), 19-36. https://doi.org/10.1111/1467-9817.12027

Schaffner, E., Schiefele, U., & Ulferts, H. (2013). Reading amount as a mediator of the effects of intrinsic and extrinsic reading motivation on reading comprehension. *Reading Research Quarterly*, *48*(4), 369-385. https://doi.org/10.1037/t56496-000

Schiefele, U., Schaffner, E., Moller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly, 47*(4), 427-463. https://doi.org/10.1002/Rrq.030

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 19,* 159–172. https://doi.org/10.1080/10573560308219

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*, 493-525. https://doi.org/10.1177/0013164408323233

Stutz, F., Schaffner, E., & Schiefele, U. (2016). Measurement invariance and validity of a brief questionnaire on reading motivation in elementary students. *Journal of Research in Reading*, *40*(4), 439-461. https://doi.org/10.1111/1467-9817.12085

Tonks, S., Davis, M. H., Rodriquez, A., & Schwartz, J. (2014, October). *Cognitive validity of a new reading motivation measure for adolescents.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Evanston, IL.

Tunmer, W. E., & Chapman, J. W. (1991). *An investigation of language-related and cognitive-motivational factors in beginning reading achievement.* Research proposal submitted to the Ministry of Education, New Zealand. Palmerston North, New Zealand: Educational Research and Development Centre, Massey University.

Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *The Journal of Educational Research*, *100*(2), 81-101. https://doi.org/10.3200/JOER.100.2.81-101

Wang, C. (2014). Improving measurement precision of hierarchical latent traits using adaptive testing. *Journal of Educational and Behavioral Statistics, 39,* 452–477. https://doi.org/10.3102/1076998614559419

Wang, J. H. Y., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between US and Chinese students. *Reading Research Quarterly, 39*(2), 162-186. https://doi.org/10.1598/RRQ.39.2.2

Wang. W. & Kingston, N., (2019). Adaptive testing with a hierarchical item response theory model. *Applied Psychological Measurement, 43*(1) 51-67. https://doi.org/10.1177/0146621618765714

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, *25*(1), 68-81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*, 420-432. https://doi.org/10.1037/0022-0663.89.3.420

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school context. *Review of research in education*, *23*, 73-118. https://doi.org/10.2307/1167288

Table 1
Definitions and Sample Items for the 15 Measured Reading Motivation Constructs

Construct	Sample Item
Self-efficacy	I am one of the best readers in my class
Perceived Difficulty	The books that teachers assign are often hard for me to read
Challenge in Reading	I enjoy reading difficult material
Curiosity for Reading	I get excited when reading about new things
Involvement in Reading	I get so involved in my reading that I often lose track of time
Value for Reading	It’s very important to read a lot
Interest in Reading	I have favorite topics I like to read about
Preference for Autonomy	Choosing what I want to read is important to me
Reading for Grades	Getting good grades in reading is important to me
Reading for recognition	I feel proud when I am recognized as a good reader
Reading to Compete	It’s important to me that I read better than my classmates
Social Motivation	I like to talk with my friends about what we read in class
Pro-Social Goals for Reading	I like to help my classmates understand what they’ve read
Antisocial Goals for Reading	My friends and I laugh at classmates who don’t read well
Reading Avoidance	I find ways to avoid reading in class

Note. A list of all items can be viewed at Kingston et al. (2017)

Table 2

Models Used in Comparisons

Model	-2*loglikelihood	AIC	BIC
Unidimensional	1,323,204.73	1,326,804.73	1,339,263.83
Multidimensional	1,262,970.64	1,266,780.64	1,279,966.52
Higher-Order	1,261,866.41	1,265,492.41	1,278,041.50
Bi-factor	1,252,993.77	1,257,193.77	1,271,729.39

Note. AIC = Akaike information criterion. BIC = Bayesian information criterion.

Table 3

Reliabilities of Scores within Higher-Order and Bi-factor Models

	Reliabilities	
Factor	ARMM-H	ARMM-B
General	0.95	0.95
Self-efficacy	0.85	0.70
Curiosity	0.85	0.50
Challenge	0.86	0.69
Involvement	0.85	0.54
Value	0.87	0.51
Interest	0.88	0.52
Autonomy	0.86	0.79
Recognition	0.88	0.85
Grades	0.88	0.57
Competition	0.85	0.85
Avoidance	0.83	0.73
Difficulty	0.84	0.80
Social	0.85	0.75
Prosocial	0.85	0.76
Antisocial	0.75	0.66

Table 4
Differences in ARMM-H Scores and ARMM-B General Score by Gender

	Females	Males	Significance	Effect Size	Mean Difference Lower CI	Mean Difference Upper CI
ARMM-B						
General	45.46	40.98	0.00	0.40	3.47	5.48
ARMM-H						
General	45.01	40.70	0.00	0.40	3.34	5.28
Self-efficacy	41.51	38.70	0.00	0.20	1.58	4.04
Curiosity	44.20	39.66	0.00	0.34	3.35	5.74
Challenge	45.58	41.58	0.00	0.29	2.78	5.23
Involvement	44.31	39.30	0.00	0.43	3.98	6.05
Value	43.95	39.09	0.00	0.40	3.76	5.97
Interest	44.42	40.10	0.00	0.39	3.32	5.32
Autonomy	37.53	33.18	0.00	0.33	3.16	5.54
Recognition	41.44	37.05	0.00	0.29	3.05	5.74
Grades	43.99	39.77	0.00	0.39	3.27	5.18
Competition	41.43	38.70	0.01	0.12	0.70	4.75
Avoidance (R)	57.94	53.19	0.00	0.34	3.51	6.00
Difficulty (R)	62.94	60.99	0.03	0.10	0.18	3.72
Social	48.70	42.57	0.00	0.42	4.83	7.44
Prosocial	46.14	40.83	0.00	0.36	3.98	6.65
Antisocial (R)	56.52	50.23	0.00	0.36	4.72	7.87

Note. Females N=914, Males N=991, Scores with (R) indicate lack of a construct for the Higher Order Model, CI=Confidence Interval.

Table 5

Differences in ARMM-H Scores and ARMM-B General Score by Grade Level

	Elementary	Secondary	Significance	Effect Size	Mean Difference Lower CI	Mean Difference Upper CI
ARMM-B						
General	46.03	41.37	0.00	0.42	3.67	5.65
ARMM-H						
General	45.51	41.10	0.00	0.41	3.45	5.37
Self-efficacy	42.06	38.83	0.00	0.24	2.00	4.46
Curiosity	45.16	39.82	0.00	0.40	4.15	6.53
Challenge	46.89	41.44	0.00	0.40	4.25	6.65
Involvement	43.94	40.34	0.00	0.31	2.56	4.64
Value	44.72	39.42	0.00	0.43	4.21	6.38
Interest	44.91	40.50	0.00	0.40	3.42	5.39
Autonomy	36.87	34.29	0.00	0.19	1.40	3.77
Recognition	42.62	37.05	0.00	0.37	4.22	6.92
Grades	44.35	40.24	0.00	0.38	3.16	5.06
Competition	40.23	39.88	0.74	0.02	-1.72	2.42
Avoidance (R)	58.89	53.39	0.00	0.40	4.28	6.72
Difficulty (R)	63.50	60.97	0.01	0.13	0.74	4.31
Social	48.32	43.81	0.00	0.31	3.21	5.82
Prosocial	46.97	41.20	0.00	0.39	4.42	7.12
Antisocial (R)	57.75	50.51	0.00	0.41	5.64	8.85

Note. Females N=914, Males N=991, Scores with (R) indicate lack of a construct for the Higher Order Model, CI=Confidence Interval

Table 6

Correlations of ARMM-H Scores and ARMM-B General Score with Reading Behavior and Engagement

	Behavior			Engagement		
	All	Elementary	Secondary	All	Elementary	Secondary
ARMM-B						
General	0.61^{**}	0.52^{**}	0.64^{**}	0.84^{**}	0.80^{**}	0.85^{**}
ARMM-H						
General	0.61^{**}	0.52^{**}	0.64^{**}	0.83^{**}	0.79^{**}	0.85^{**}
Self-efficacy	0.42^{**}	0.34^{**}	0.44^{**}	0.66^{**}	0.60^{**}	0.68^{**}
Curiosity	0.60^{**}	0.53^{**}	0.62^{**}	0.78^{**}	0.73^{**}	0.80^{**}
Challenge	0.56^{**}	0.43^{**}	0.60^{**}	0.77^{**}	0.71^{**}	0.80^{**}
Involvement	0.56^{**}	0.44^{**}	0.60^{**}	0.78^{**}	0.72^{**}	0.80^{**}
Value	0.59^{**}	0.51^{**}	0.62^{**}	0.81^{**}	0.77^{**}	0.82^{**}
Interest	0.61^{**}	0.51^{**}	0.64^{**}	0.82^{**}	0.77^{**}	0.83^{**}
Autonomy	0.44^{**}	0.33^{**}	0.48^{**}	0.58^{**}	0.50^{**}	0.60^{**}
Recognition	0.42^{**}	0.34^{**}	0.43^{**}	0.55^{**}	0.48^{**}	0.57^{**}
Grades	0.59^{**}	0.51^{**}	0.62^{**}	0.81^{**}	0.77^{**}	0.83^{**}
Competition	0.35^{**}	0.29^{**}	0.39^{**}	0.42^{**}	0.34^{**}	0.46^{**}
Avoidance (R)	0.42^{**}	0.29^{**}	0.46^{**}	0.66^{**}	0.55^{**}	0.71^{**}
Difficulty (R)	0.20^{**}	0.11^{**}	0.24^{**}	0.42^{**}	0.29^{**}	0.47^{**}
Social	0.52^{**}	0.42^{**}	0.55^{**}	0.70^{**}	0.61^{**}	0.73^{**}
Prosocial	0.50^{**}	0.42^{**}	0.53^{**}	0.66^{**}	0.57^{**}	0.69^{**}
Antisocial (R)	0.30^{**}	0.21^{**}	0.33^{**}	0.45^{**}	0.36^{**}	0.48^{**}

Note. * $p < .05$, ** $p < .01$, scores with (R) indicate lack of a construct for the Higher Order Model.

Table 7

Correlations of ARMM-H Scores and ARMM-B General Score with Reading Achievement

	Reading			
	Grade 5	Grade 6	Grade 7	Grade 8
ARMM-B				
General	0.32**	0.31**	0.38**	0.52**
ARMM-H				
General	0.32**	0.30**	0.37**	0.53**
Self-efficacy	0.36**	0.29**	0.47**	0.49**
Curiosity	0.23**	0.23*	0.30**	0.45**
Challenge	0.33**	0.32**	0.45**	0.45**
Involvement	0.33**	0.34**	0.47**	0.52**
Value	0.28**	0.33**	0.33**	0.53**
Interest	0.28**	0.26*	0.31**	0.49**
Autonomy	0.28**	0.33**	0.32**	0.60**
Recognition	0.14**	-0.01	0.25**	0.37**
Grades	0.31**	0.30**	0.34**	0.54**
Competition	0.09*	-0.13	0.28**	0.30**
Avoidance (R)	0.33**	0.25*	0.34**	0.51**
Difficulty (R)	0.41**	0.51**	0.47**	0.41**
Social	0.16**	0.12	0.09	0.43**
Prosocial	0.16**	0.11	0.05	0.37**
Antisocial (R)	0.23**	0.35**	0.13	0.29**

Note. * $p < .05$, ** $p < .01$, scores with (R) indicate lack of a construct for the Higher Order Model.

Table 8

Standardized Regression Coefficients from Regression Analysis of Reading Achievement Predicting ARMM Scores, While Controlling for Math Achievement and Analysis of Math Achievement Predicting ARMM Scores, While Controlling for Reading Achievement

	Reading				Math			
	5	6	7	8	5	6	7	8
ARMM-B General	0.33**	0.16	0.39**	0.51**	-0.01	0.24	-0.02	0.02
ARMM-H								
General	0.33**	0.16	0.38**	0.53**	-0.01	0.23	-0.01	0.00
Self-efficacy	0.36**	0.18	0.42**	0.49**	-0.00	0.18	0.08	0.01
Curiosity	0.26**	0.16	0.29*	0.49**	-0.04	0.12	0.02	-0.04
Challenge	0.32**	0.17	0.44**	0.48**	0.03	0.25	0.02	-0.01
Involvement	0.34**	0.21	0.50**	0.51**	-0.01	0.20	-0.04	0.03
Value	0.28**	0.17	0.41**	0.47**	0.00	0.24	-0.13	0.08
Interest	0.30**	0.10	0.32**	0.54**	-0.03	0.26	-0.02	-0.04
Autonomy	0.31**	0.25	0.27*	0.58**	-0.04	0.14	0.08	0.04
Recognition	0.15**	-0.01	0.22	0.29	-0.01	-0.01	0.06	0.10
Grades	0.31**	0.17	0.32**	0.54**	0.00	0.22	0.04	0.01
Competition	0.12**	-0.23	0.23	0.38*	-0.04	0.15	0.10	-0.10
Avoidance (R)	0.32**	0.23	0.42**	0.41**	0.02	0.03	-0.14	0.14
Difficulty (R)	0.37**	0.53**	0.38**	0.40*	0.06	-0.02	0.15	0.02
Social	0.18**	0.06	0.13	0.36*	-0.02	0.10	-0.07	0.08
Prosocial	0.20**	-0.08	0.08	0.51**	-0.06	0.32	-0.04	-0.16
Antisocial (R)	0.20**	0.29*	0.20	0.35*	0.04	0.10	-0.13	-0.06

Note. * p < .05, ** p < .01, scores with (R) indicate lack of a construct for the Higher Order Model.

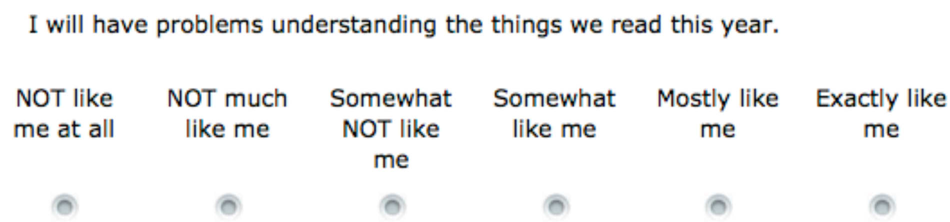

Figure 1. ARMM presentation model 1; 6-point scale, with labels on all points.