Student Motivation and Effort in the Swedish TIMSS Advanced Field Study

Hanna Eklöf
Department of Applied Educational Science, Umeå University, Sweden

The purpose of the present study was twofold. Firstly, the study aimed to test a 9-item test-taking motivation scale that could be used in large-scale, low-stakes assessment contexts. The scale was assumed to measure motivation, invested effort and perceived importance in the assessment situation. Secondly, the study aimed to describe student reported effort, motivation to do their best and perceived importance of the TIMSS Advanced field-test. The scale was administered to a Swedish sample (n = 163) participating in the TIMSS Advanced 2008 field-test. Confirmatory factor analysis supported a two-factor solution as hypothesized, but the importance factor was unreliable and a one-factor solution with an effort factor only provided a better fit to data. Based on these results, a number of countries participating in the TIMSS Advanced 2008 Main Study chose to include the effort scale in their student questionnaires. Results further showed that the pupils in the present sample perceived the TIMSS field-test as less important and that they reported a low level of motivation to do their best on the test. Furthermore, many students omitted several items in the test booklet. It is concluded that the results from the Swedish TIMSS Advanced field-test might be biased by a lack of motivation and effort among students. Preliminary findings from the subsequent Swedish TIMSS Advanced 2008 Main Study also reveal a modest level of motivation and effort among the students, and a significant relation between reported level of effort and achievement on the test.

Keywords*: test-taking motivation, effort, low-stakes assessment, validity*

Introduction

Test-takers' reactions and behaviors in the assessment situation are important to consider as they can have an impact on test outcomes. One important aspect of the psychology of test-taking is motivation. The issue of test-taker motivation is particularly relevant in the context of assessments that have few consequences for the test-takers, but that may be highly consequential for policy makers and other stake holders. TIMSS is an example of such an assessment. In TIMSS, the stakes of the test are (seemingly) low for the participating students as performance has no consequences for the individual: students do not get any feedback on their performance, individual test results are not returned and test results do not count toward student grades. On the other hand, on a systemic level, good performance in a study like TIMSS may be important. Results are used for cross-national comparisons and evaluation of educational quality, and sometimes even as a ground for educational reform. It follows that it is important that the results obtained in a study like TIMSS are valid measures of student proficiency.

However, when assessments results have no consequences for the test-takers, it cannot be taken for granted that all students take these tests seriously, that they perceive the test as important and that they are motivated to invest effort in completing test items. Ignoring the test-taking motivation component could be a threat to the validity of the interpretation and use of test results, in the sense that results could be a measure of motivation and

effort as much as a measure of student knowledge (Thelk, Sundre, Horst, & Finney, 2009). This is a potential problem not only in TIMSS, but also in other international comparative studies; assessments that are used for accountability purposes; assessments for program evaluation or pilot tests for high-stakes tests. On the other hand, it cannot be taken for granted that all students *lack* motivation to do their best on low-stakes tests either. It is not uncommon that low performance on low-stakes tests is attributed to a lack of motivation, but unless the motivational components of test-taking have been assessed and evaluated, this might be an invalid attribution (Thelk et al., 2009). Indeed, if student motivation and effort is not taken into consideration in low-stakes assessment contexts, the true meaning of a test score might be difficult to evaluate. What complicates matters even more in a study like TIMSS is that different nations with different assessment cultures are to be compared. Should it be the case that students in some nations perceive the test as important and invest effort when completing the test while students in other nations do not, and that level of invested effort affects actual performance, this could be a serious threat to the validity of obtained results and comparisons made. For an assessment like the field-test to TIMSS, the validity of the interpretations and uses of assessment results can also be threatened if item characteristics are affected by student motivation and effort rather than their proficiency. There are of course many other variables that also pose threats to the validity of the interpretations and uses of test results, but the motivational component of test-taking is surely one of them.

Empirical studies of student motivation in low-stakes assessment contexts have yielded mixed results. Several studies have shown a clear relationship between test stakes and test-taking motivation on the one hand, and between test-taking motivation and test performance on the other (Schiel, 1996; Thelk, Sundre, Horst, & Finney, 2009; Wise & DeMars, 2005, Wolf & Smith, 1995). Other studies have failed to demonstrate clear relationships between test stakes, test-taking motivation and test performance (Baumert & Demmrich, 2001; O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005).

Studies have further found near-zero correlations between reported motivation/effort and academic ability (Wise & DeMars, 2005; Wise & Kong, 2005;). These findings are important as one could otherwise suspect that students choose to attribute poor performance to lack of effort rather that lack of ability.

A previous study of test-taking motivation in the Swedish TIMSS 2003 context (see Eklöf, 2006, 2008) could show that Swedish Grade 8 students perceived the test as rather important and that they reported a high level of motivation, despite the seemingly low stakes of the test. Results from this study however also revealed a positive and significant relationship between motivation and performance. The amount of variance in test score that was explained by motivation was not very high when other variables were held constant, but considering the fact that a few points systematical difference on the individual level can mean a significant difference on an aggregated level, findings may still be important.

In sum, it is not entirely clear what role motivation plays in performance on low-stakes assessments, whether assessment that have no consequences are actually perceived as low-stakes and unimportant by the test-takers, which groups of test-takers that are more likely to suffer from a lack of motivation, or which strategies that are best to use in motivating students taking low-stakes tests. It is also not clear how motivation and effort is best assessed, and psychometrically described instruments designed for measuring student test-taking motivation are rare.

What is clear, however, is that a valid instrument measuring motivational aspects in low-stakes assessment contexts could be of great value, and that increased knowledge about the dynamics of test-taking motivation would be beneficial to our interpretations and uses of assessment results, not least in the context of international comparative studies.

Drawing on items from the Student Opinion Scale, originally developed by Wolf & Smith (1995) and later refined by Donna Sundre and her collegues (Sundre & Finney, 2002; Sundre & Moore, 2002) as well as items previously used to study test-taking motivation in the TIMSS context (Eklöf, 2006, 2008), the aim of this primarily descriptive study was to

a) try out and evaluate a short instrument that could possibly later be augmented to existing instruments and used across languages for assessing student motivation and the relationship between motivation and performance in low-stakes assessment contexts

b) to explore reported and actual test-taking behavior of a sample of students in one of the countries participating in the TIMSS Advanced field-test.

The present study also continues an empirical study of test-taking motivation that was initiated in connection with TIMSS 2003 in Sweden. Where appropriate, results from the present study are compared to results from TIMSS 2003.

Materials and Method

Participants

The sample consisted of 163 Swedish students, 67 girls (41 %) and 94 boys (58 %), all in the last year of secondary schooling (Grade 12, age 18-19 years) and all participating in the TIMSS Advanced 2008 field-test. Two individuals have not reported either gender or their identification code on the questionnaire why these are not possible to identify. All students were taking advanced courses in mathematics and/or physics. In two schools the students completed the TIMSS mathematics test and in three school three the students completed the TIMSS physics test. The sampling of schools was made by the Swedish National Agency for Education which is responsible for implementation of the TIMSS study in Sweden.

Measures

The Motivation Scale

The measure evaluated and used in the present study is based on the propositions of the expectancy-value theory of achievement motivation (Atkinson, 1964; Eccles & Wigfield, 2002; Pintrich & DeGroot, 1990) and on scales developed by Donna Sundre (1999) and Hanna Eklöf (2006, 2008).

In connection with TIMSS 2003 (Grade 8 pupils) in Sweden, Eklöf developed and used a short 4-item scale with items asking for motivation, effort and importance. Evaluation of this short scale indicated sound psychometric properties (Eklöf, 2006) and a positive relation to achievement on the TIMSS test (Eklöf, 2007, 2008). Indices of discriminant validity further suggested that test-taking motivation, as measured with this four-item scale, was distinct from general attitudes towards, self-concept or valuing of mathematics (Eklöf, 2006). Still, this scale needed revisions for a number of reasons. For instance, the scale used in TIMSS 2003 included pretest as well as posttest items, while the purpose of the present scale was to be a posttest scale.

One of the few instruments that are developed to measure examinee motivation and effort and that has been used and evaluated in a number of studies is the Student Opinion Scale (SOS) (Sundre, 1999). The SOS is a 10-item scale that consists of two 5-item

subscales (*Reported Effort* and *Perceived Importance of the Test*). Favorable reliability and validity evidence has been found for the SOS and its two subscales (Sundre,1999; Sundre & Moore, 2002; Sundre & Wise, 2003), and the effort scale in particular. The SOS has been used for a number of years and on a large number of students, though only in English and mainly on students in one U.S. university. One option for the present study could have been to translate all items in the SOS and try them on a Swedish sample. However, all items in the SOS are not suitable for use in assessments where results are not returned to the students (such as in TIMSS). Also, other items, not in the SOS scale but rather in the scale developed by Eklöf (2006) was judged relevant to include as they have been proven valuable and demonstrated a positive relationship with achievement.

The work with a scale that could be used in the TIMSS context therefore resulted in a scale with nine items (see Table 1) adapted from the SOS as well as from the test-taking motivation scale used in TIMSS 21003. Six items were assumed to measure reported effort and three items were assumed to measure perceived importance of the test. All items were measured on a four-point scale with ratings ranging from a highly favorable attitude to a highly unfavorable attitude (1 = *agree a lot*, 4 = *disagree a lot*).

In addition, the instrument administered to the students in the present sample also contained an open-ended item where the students were asked to describe their motivation to do their best in the TIMSS field-test in their own words (also an item used previously in the TIMSS 2003 context). Further, the instrument contained three items asking for test anxiety, another important aspect of the psychology of test-taking, as well as a number of items asking for students general attitudes towards tests in school.

The TIMSS Advanced test.

The achievement test in the present study were the mathematics and physics tests used in the TIMSS Advanced 2008 field-test. The test could be regarded a true low-stakes test for the test-taker as a) results are not returned to the students and do not count towards student grades, and b) the assessment was a field-test, aimed to try out and evaluate mathematics and physics items rather than to evaluate student proficiency and make cross-national comparisons.

Procedure

The questionnaires were administered to five schools (~200 students) out of the 20 Swedish schools participating in the TIMSS Advanced 2008 field-study. Teachers as well as students were informed that participation in the motivation study was voluntary. Still, all sampled schools administered the questionnaire and the questionnaire was completed by 163 students, which corresponds to the total number of students in these schools participating in the TIMSS Advanced field-study. A larger number of students were sampled to participate in TIMSS Advanced but did not do so, and hence did not complete the motivation questionnaire either. The questionnaire was administered directly after administration of the TIMSS mathematics/physics booklets.

Statistical Analysis

All data analyses were made in SPSS Version 17.0 and AMOS 17.0. Initially, questionnaire data was screened for missing values. To test the distributional properties of data, univariate skewness and kurtosis of the variables were analyzed. As part of the psychometric evaluation of the scale, Mardia's measure of multivariate kurtosis was consulted, and data was screened for outliers. Reliability (internal consistency) of scores for the present sample was assessed through coefficient alpha and the hypothesized factor structure was tested in confirmatory factor analysis. After this intitial screening and evaluation of the scale used, descriptive statistics were calculated in order to describe

students' perceptions and behavior in the assessment situation: means and standard deviations on the item level, correlations between items and the percentage of students agreeing or disagreeing with each statement. Where appropriate, results are compared with corresponding results from TIMSS 2003 (Swedish Grade 8 students). Students answers to the open-ended item in the questionnaire were then analyzed and categorized, and their response patterns on the the TIMSS Advanced test were analyzed with regard to omitted items. Finally, some preliminary results from using the effort scale in the TIMSS Advanced 2008 main study (in Sweden) were calculated in terms of mean values and relation with test score.

Results

Psychometric evaluation of the Motivation scale

Screening of items

All variables were in the range of -1 - +1 as concerns univariate skewness as well as kurtosis. However, Mardia's coefficient suggested violation of the multivariate normality assumption (Mardia's normalized coefficient = 9.40). This value fall outside what can reasonably can be viewed as multivariate normality. A screening for multivariate outliers across variables revealed that ten cases were to be counted as outliers. These ten cases were consistently very negative towards their participation in the field study (not at all motivated, spent no effort, test felt unimportant). When these ten cases were removed from data, Mardias coefficient dropped to .791 which is well below the stipulated limits. CFA's using Maximum Likelihood estimation (ML) were performed for the sample with 163 individuals as well as for the sample with outliers removed (n= 153). Both analyses yielded comparable results and as the sample was so small to begin with, below is reported the CFA's using maximum likelihood estimation with 163 cases.

Reliability

Internal consistency reliability for the hypothesized six-item effort/motivation scale was acceptable (α = .81) for the present sample. When the reliability of the hypothesized three-item Importance scale was assessed, the coefficient was α = .57 which must be regarded as rather poor. Correlations between the three items in the Importance scale also showed that one item was weakly related to the other items in the scale, and item analysis showed that some students gave awkward answers to this item. The value of this already very brief Importance scale could thus be questioned.

Confirmatory Factor Analysis

The hypothesized scale structure as well as a number of competing models was subjected to confirmatory factor analysis. In all models, the covariance matrix was used as the measure of association. Multiple model fit indices; the normed χ^2 statistic, the Comparative Fit Index (CFI), the Root Mean Square Error of Association (RMSEA) and the Standardized Root Mean Residual (SRMR) were consulted.

Three different models were tested. The first model (a) had an effort factor with six indicators and an importance factor with three indicators. This was the hypothesized structure of the data. The second model (b) had only one motivation factor with nine indicators, in line with the original model of test-taking motivation suggested by Wolf & Smith (1995). The third model tested (c) contained one six-item effort/motivation factor only (the less reliable importance factor excluded). Fit indices for the respective models are shown in Table 1.

Insert Table 1 about here

All three specified models fitted the data rather well. In particular, the two-factor model with one 6-item Effort factor and one 3-item Importance factor, as well as the one-factor model with 6 effort items only showed a very good fit to data and had non-significant χ^2/df-values. The correlation between the Effort factor and the Importance factor was r = .68. The standardized pattern coefficients (factor loadings) were acceptable for all items in the Effort scale (ranging from .45 (item E4, see Table 2), to .84 (E3)). Loadings for the three items in the Importance scale were also acceptable, although one item had a loading of only .35 (I3). The other two items had loadings of .58 and .83, respectively (for more detailed results of the psychometric evaluation of the motivation scale, see Eklöf, 2010). Based on the results from these analyses, it was decided to use the six-item effort scale in a number of countries (Sweden, Norway, Slovenia and Russia) participating in the TIMSS Advanced 2008 Main Study.

Student responses to items in the Motivation scale

In the data set, there were only 5 cells (out of 1,467) with missing values. These five missing values were distributed on different variables and different cases and in the analyses, they were replaced with the mean value for the respective variables. Of the 163 students that completed the motivation scale, 157 students were valid cases in the achievement data base. One group of students (n = 64) took a mathematics test, the other group (n = 93) took a physics test.

Mean values on the item level (see Table 2) reveal a modest interest among the students as concerns their participation in the TIMSS Advanced field-test. Actually, a majority of students claimed that they did not perceive the test as important, that they did not feel motivated to do their best and that they did not invest as much effort as they could have.

Insert Table 2 about here

Descriptive results also show a noticeable difference between how the Grade 8 students participating in TIMSS 2003 and the Grade 12 students participating in the field trial to TIMSS Advanced 2008 responded on comparable items (see Table 2). When looking at percentages of students agreeing or disagreeing with the various statements, it is shown that a majority of student in the present Grade 12 sample (58 %) disagreed or disagreed a lot with the statement that they had felt motivated to do their best on the test, as compared to 24 % of the Grade 8 students in 2003. In the present sample, 63 % of the students did not think that it was important to do well on the test, as compared to 26 % of the Grade 8 students in 2003. A relatively large proportion (57 %) of the students in the present sample (given their answers to other items in the scale) reported that they gave their best effort on the test. Still, corresponding percentage in the TIMSS 2003 study was 90 %.

Further, in the TIMSS Advanced sample, 82 % of the students agreed or agreed a lot with the statement that they could have tried harder on the test (even if they on a previous item reported that they gave their best effort); 78 % agreed or agreed a lot with the statement that they try harder on other tests they have in school, and 93 % of the students disagreed or disagreed a lot with the statement that the test felt important to them. 66 % of the students in the sample did not agree with the statement that they worked on every item even when the item seemed difficult, and more than half of the students claimed that they were not concentrated while working with test items.

Preliminary results from the TIMSS Advanced 2008 *Main study* in Sweden further show that the Grade 12 students participating in the Main Study also reported a rather low

level of effort on the test (see Table 2). The effort scale was used also in Norway, Slovenia and Russia but unfortunately, these results have not yet been analyzed.

The effort scale has been used also in connection with high-stakes national tests in science in Sweden and there is a significant difference between reported effort among the students taking the high-stakes national test and the students taking the low-stakes TIMSS test (see Eklöf, 2010). Students taking the national test to a high degree perceive the test as important and report that they invested effort when completing test items.

The questionnaire used in the present study also contained items asking for test anxiety and general attitudes towards tests in school. Findings suggest that few students (between 2 and 4 % on the three test anxiety items) were adversely affected by test anxiety when taking the test. This is in agreement with theoretical assumptions and previous empirical findings; when a test is perceived as less important, the level of test anxiety is also low, except for a small group of students who always feel anxious in assessment situations. Students general attitudes towards tests in school further seem rather positive. For example, only 1 % of the students reported that they found tests in school unnecessary.

The open-ended item

Ninety per cent of the students completed the open-ended item in the questionnaire where they were asked to describe in their own words how motivated they felt to do their best on the test and why. A preliminary coding of this item show that 50 % of the students expressed themselves in negative terms as regards motivation to do their best on the TIMSS test. 30 per cent of the students were categorized as positive according to their answers on the open-ended item, while 20 % were categorized as indifferent. The same categorization was done in TIMSS 2003 for the Grade 8 students and corresponding numbers there were 72, 13 and 15 %, respectively. Among the students who reported a lack of motivation to do their best on the test in the present sample, many referred to the low stakes of the test (the fact that it did not count towards their grade; that it was only a field trial) as detrimental to their motivation. Others claimed that they lacked motivation as they had already finished all their mathematics/physics courses and gotten their grades, while others claimed that they lacked motivation for opposite reasons: that they had other more important tests in mathematics or physics to concentrate on. Among those students who were categorized as positive, some reported that they always try their best on every test, others felt a social responsibility and wanted “to help” with the study.

Test-taking behavior

The relationship between reported level of motivation and total test score has not been analysed in the field-test as no achievement scores have been calculated, but when looking at achievement data on an item level, it is obvious that many students chose to omit items in the test booklets, particularly constructed-response items. In the group taking the mathematics test (n = 64), only 14 % of the students had attempted to answer all constructed-response items, and 61 % of the students had omitted more than 25 % of the constructed-response items. In the group taking the physics test (n = 93), 16 % of the students made an attempt at every constructed-response item, and 39 % had omitted more than 25 % of these items. It should also be noted that an “attempt” does not equal a serious effort to solve the task at hand. The scoring guide does not allow separation of those apparently trying to solve an item from those only making some kind of mark before moving on to the next item. To be able to say something more about how students actually worked with the items in the test booklets, more detailed analysis of response patterns and item characteristics are needed.

Some preliminary findings from the TIMSS Advanced 2008 Main study. As noted, the effort scale was used also in the TIMSS Advanced 2008 Main Study in Sweden. Here,

correlations between ratings on the effort scale and achievement have been calculated using the *IEA IDB Analyzer*. Preliminary results show statistically significant correlations between reported effort and achievement (r = .30 for the mathematics test and r = .24 for the science test). In the official TIMSS Advanced report published by the Swedish National Agency for Education, the effort scale was also included in the analysis as an index of motivation. This index was one of the (relatively few) indices that had a significant effect on test score, corresponding to the effect of socio-economic status or language spoken at home (Swedish National Agency for Education, 2009). Again, no corresponding analyses have yet been performed for the other countries using the effort scale in TIMSS Advanced. They are underway during 2010, together with more thorough analyses of student motivation and achievement in the TIMSS context.

Conclusions and Implications

The first aim of the present study was to test a scale that could possibly be used in future studies of test-taking motivation in low-stakes assessment contexts. Psychometric evaluation of the 9-item motivation scale, used on a sample of Swedish Grade 12 students participating in the TIMSS Advanced 2008 field-test, suggest that a 6-item effort scale seems to be a psychometrically sound scale that can be used for assessing motivation and effort in low-stakes assessment contexts. Based on these findings, it was also decided to use the Effort scale in a number of countries participating in the TIMSS Advanced 2008 Main Study. The importance scale, however, consisted of only three items and was not very reliable in the present study. The value of this scale can be questioned even if a confirmatory model with a three-item Importance scale showed an acceptable fit to data. It is also possible that an anxiety scale should be added to a future Motivation Questionnaire as test anxiety also is an important feature of test-taking psychology. In general, we feel that there is a need for continued work on the motivation scale in order to arrive at a psychometrically sound scale that is sufficiently brief for potentially unmotivated subjects to respond to, but that still generates reliable and valid data that it can be used as an effective measure of motivation in the assessment situation.

The second aim of the present study was to explore student perceptions of the TIMSS Advanced field-trial, mainly in terms of motivation, perceived importance of the test and reported invested effort. It can be concluded that the students in the present sample did not perceive the TIMSS field-test as very important, that a majority of students reported that they did not feel very motivated to do their best, that many students claimed that they could have invested more effort, and that they claim to invest more effort on other tests they have in school. In line with their answers to the items in the Motivation scale, there was also a large number of non-responses (and sometimes nonsense-responses) in the test booklets.

Omissions are a typical result of non-effortful behaviour and as data from the TIMSS field study is used to calibrate test items and evaluate item quality, a large number of non-responses or nonsense-responses pose a threat to the quality of assessments of this kind.

To summarise the findings from this study, it is possible that the Swedish results from the TIMSS Advanced 2008 Field Study are biased by a lack of motivation among the participating students. The design of the present study does not allow any far-reaching generalizations, but one could speculate that what is being evaluated in field-tests and try-outs is perhaps not always item quality, but sometimes also (lack of) student effort. If this would be the case, the value of field-testing is diminished. There is little meaning in spending time and resources on trying out items if they do not work as they will/should have done in the real test.

Results imply that it is important to monitor student motivation and effort also in pilot tests and try-out studies in order to be able to control for this variable and thereby get better evaluations of obtained findings. Results also imply that it might be important to make efforts to raise the situational interest in (low-stakes) tests among students as well as teachers. It has been reported from the PISA study that raising situational interest by, for example, simply emphasizing the importance of trying hard or emphasizing the societal value of a good performance, can be enough to motivate students to try a bit harder (Baumert & Demmrich, 2001).

Results from the present study also show a considerable difference between how Grade 8 students participating in TIMSS 2003 and the Grade 12 students in the present sample responded to comparable items. While the Grade 8 students in general seemed positive towards their participation in TIMSS, the Grade 12 students in general seemed rather negative. There might be several explanations for these differences. Firstly, there is research showing that with increasing age, years of schooling and experience with standardized testing and grading, students get more aware of where effort need to be invested and not. Further, their workload increases and students are sometimes tired of being assessed and evaluated. Secondly, there is a difference between a TIMSS field-test and a TIMSS Main Study. If TIMSS Main Study is a low-stakes test, the TIMSS field-test basically is a no-stakes test. There are some potential external motivators present in the TIMSS Main Study as students are "representing their country" and as results are going to be evaluated and compared nationally and internationally. It was therefore expected that they Grade 12 students participating in the TIMSS Advanced 2008 Main study would report higher levels of motivation and effort than the field-test students, particularly as the students participating in TIMSS Advanced is a selected group of students that have chosen to study advanced courses in mathematics and physics. However, preliminary results suggest that the Swedish students participating in the Main Study were not more motivated and did not spend more effort than the students participating in the field-test.

Still, in a study like TIMSS, the comparability of results is essential. Should it be the case that all students in all countries are equally unmotivated, or that level of effort can be shown to have a negligible effect on performance, test outcomes can still be validly compared between countries - at least as concerns the effect of motivation on test score. As noted above, the results from the other nations using the Effort scale in TIMSS Advanced has not yet been analyzed, but these will be interesting and important comparisons to make.

Again and to conclude, the main implication from the present study is that may be important to continue monitoring student motivation and effort in studies like TIMSS, and to do that using high-quality measures of motivation.

References

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: D. Van Nostrand Company.

Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*, 441-462.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109-132.

Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement,* 66, 643–656.

Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*, 311-326.

Eklöf, H. (2008). Test-taking motivation on low-stakes tests: A Swedish TIMSS example. In *Issues and Methodologies in Large-Scale Assessments*, IERI Monograph Series, Vol. 1. Hamburg: IEA-ETS Research Institute.

Eklöf, H. (2010). *Psychometric evaluation of the Swedish Student Opinion Scale*. Submitted manuscript.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33-40.

Schiel, J. (1996). Student effort and performance on a measure of postsecondary educational development (ACT Research Report No. 96-9). Iowa City: ACT.

Sundre, D. L. (1999). *Does examinee motivation moderate the relationship between test consequences and test performance?* Report No TM029964. Harrisonburg, Virginia: James Madison University.

Sundre, D. L., & Finney, S. J. (2002). *Enhancing the validity and value of learning assessment: Furthering the development of a motivation scale*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Sundre, D. L., & Moore, D. L. (2002). The student opinion scale: A measure of examinee motivation. *Assessment Update, 14*, 8-9.

Sundre, D. L., & Wise, S. L. (2003). *Motivation filtering: An exploration of the impact of low examinee motivation on the psychometric quality of tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 2003.

Swedish National Agency for Education. (2009). *TIMSS Advanced 2008. Svenska gymnasieelevers kunskaper i avancerad matematik och fysik i ett internationellt perspektiv*. [TIMSS Advanced 2008. Swedish students' proficiency in advanced mathematics and physics in an international perspective]. Report 336, 2009. Stockholm: Skolverket.

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *The Journal of General Education, 58*, 129-151.

Wise, S. L., & DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and possible solutions. *Educational Assessment,* 10(1), 1-17.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*, 162-183.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227-242.

Table 1. Fit indices obtained from confirmative factor analyses

Model	ML χ^2	*df*	χ^2/df	CFI	RMSEA	SRMR
1. One 9-item factor	60.47	27	2.24	.92	.09	.06
2. Two factors (6 Eff, 3 Imp)	40.36	26	1.55	.96	.06	.06
3. One 6-item effort factor	15.30	9	1.70	.98	.07	.04
4. Three factors (6 Eff + 3 Imp + 3 Anx)	90.83	51	1.78	.92	.07	.07

Table 2. Items in the Motivation Scale, together with means and standard deviations for the TIMSS Advanced field-test (Grade 12) and corresponding values for comparable items in the TIMSS 2003 Main Study (Grade 8) and the TIMSS Advanced 2008 Main Study (Grade 12, mathematics and physics separately)

Item	Wording	TIMSS Advanced field-test (n = 163)		TIMSS 2003 Main Study (n = 350)		TIMSS Adv. 2008 math (n = 2,244)		TIMSS Adv. 2008 phys (n = 2,250)	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
E1	I gave my best effort on this test	2.61	.83	3.22	.65	2.65	.82	2.77	.77
E2	I worked on each item in the test and persisted even when the task seemed difficult	2.26	.82	-	-	2.21	.83	2.35	.85
E3	I felt motivated to do my best on this test	2.30	.83	2.92	.71	2.18	.88	2.24	.88
E4	I did not give this test my full attention while completing it	2.44	.85	-	-	2.31	.87	2.37	.86
E5	While taking this test, I could have worked harder on it	1.88	.76	-	-	1.90	.85	1.93	.82
E6	I tried less hard on this test as I do on other tests we have at school	1.88	.87	-	-	1.95	.87	2.01	.88
I1	This was an important test to me	1.59	.64	-	-	-	-	-	-
I2	Doing well on this test was important to me	2.15	.89	3.00	.79	-	-	-	-
I3	I am not concerned about the result I receive on this test	2.20	.88	-	-	-	-	-	-

Note. Items are scored so that a low value is always indicative of a more negative attitude towards the test in terms of perceived importance and reported invested effort (minimum value = 1, maximum value = 4).