

Test-Taking Motivation on Low-Stakes Tests: A Swedish TIMSS 2003 Example

Hanna Eklöf

Department of Educational Measurement, Umeå University, Sweden

The study's objective was to investigate student test-taking motivation in the Swedish TIMSS 2003 context. Swedish eighth-grade students formed the study sample and the focus was on mathematics. Test-taking motivation was measured using questionnaire items and interviews, and reported level of test-taking motivation was regressed on test score. The questionnaire study showed that the Swedish students in general reported that they were well motivated to do their best in TIMSS. According to regression analysis, test-taking motivation was positively and significantly though rather weakly related to mathematics achievement. The interview study mainly corroborated the main results from the questionnaire study but added some complexity to the results. Although most interviewed students reported that they were well motivated to do their best in TIMSS and that they valued a good performance, they nevertheless seemed aware of the fact that the test was low-stakes for them personally. Many students further reported competitive, comparative, or social responsibility reasons as motivating, while other students seemed more intrinsically motivated to do their best. Findings from quantitative as well as qualitative analyses suggest that the Swedish TIMSS result is unlikely to be negatively affected by a lack of student motivation. However, nothing is known about student test-taking motivation in other countries participating in TIMSS, and further research exploring this issue in an international context is warranted.

Key words: TIMSS 2003; test-taking motivation; test performance, mathematics, Swedish eighth-grade students

Introduction

The present paper presents parts of a research project in the Swedish TIMSS 2003 context where different aspects of student achievement motivation are studied from a measurement perspective and a validity perspective (see Eklöf, 2006a, 2006b; Eklöf, in press). Student motivation is a core issue in educational settings as achievement motivation is assumed to interact with achievement behavior in important ways (Pintrich & Schunk, 2002; Wigfield & Eccles, 2002). Achievement motivation can be conceptualized and measured on different levels of generality. The most common type of motivational measure is domain-specific and measures achievement motivation for a particular domain (e.g., mathematics, science). However, achievement motivation can also be conceptualized and measured on a situation-specific level, i.e., motivation to perform well in a given situation, or on a given test. Situation-specific motivation or test-taking motivation is the focus of the present paper.

A positive motivational disposition towards the test is often assumed to be a necessary though not sufficient condition for a good test performance (Cronbach, 1988; Zeidner, 1993; Wainer, 1993; Robitaille & Gardner, 1996) and Messick (1988) noted that a poor test performance could be interpreted not only in terms of test content and student ability, but also in terms of lack of motivation. If different groups of students would differ systematically in level of motivation, and if less motivated students are disadvantaged in that they score below their actual proficiency level, test-taking motivation would be a possible source of bias (Zeider, 1993; Wainer, 1993; Mislevy, 1995; O'Leary, 2002; Baumert & Demmrich, 2001; O'Neil, Sugrue, Abedi, Baker, & Golan,

1997; Robitaille & Gardner, 1996) and hence a threat to the validity of score interpretation and use (Messick, 1995). The issue of student test-taking motivation thus is an issue of validity, and an issue of the trustworthiness of test results. However, knowledge of how individuals perceive the tests they are designated to complete, and their motivation to do their best on these tests, is scarce (Baumert & Demmrich, 2001; Nevo & Jäger, 1993), not least in the context of large-scale, comparative studies.

Test-Taking Motivation and Low-Stakes Tests

Tests that have no personal consequences, i.e., low-stakes tests, are often assumed to cause a decrease in motivation and performance (Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995; Wise & DeMars, 2003). TIMSS is, in several aspects, a low-stakes test and the issue of test-taking motivation is therefore highly relevant in the TIMSS context. Indeed, a rather common concern in the TIMSS context is that not all students are motivated to do their best on the test and that the results therefore can be an underestimation of student knowledge (Baumert & Demmrich, 2001).

First, the result on the TIMSS test has no impact on student grades in mathematics or science. Second, the results in TIMSS are mainly summarized at a national level and no individual results are given to the students or the schools. Thus, neither the students, their teachers, parents, nor peers will ever know the result of an individual student.

On the other hand, one may argue that the fact that the students represent their country in a world-wide comparative study is motivating for the students. One may also argue that the low stakes of the test make the students less anxious, and that they therefore achieve as well as they would on an ordinary test, although they are not maximally motivated.

Previous Research on Test-Taking Motivation

A vast amount of research has investigated various aspects of general and domain-specific achievement motivation. The research on situation-specific motivation or test-taking motivation is anything but vast. Studies are scattered in time and place, theoretically and methodologically. However, the expectancy-value theory of achievement motivation (Pintrich & De Groot, 1990; Eccles & Wigfield, 2002; Wigfield & Eccles, 2002) has been applied to a number of studies investigating test-taking motivation (Wolf, Smith, & Birnbaum, 1995; Wolf & Smith, 1995; Baumert & Demmrich, 2001; Sundre & Kitsantas, 2004), and was the theoretical framework used in the investigation of test-taking motivation in the Swedish TIMSS 2003 context as well (see Eklöf, 2006b). The expectancy-value theory is comprehensive in order to mirror as many as possible of the processes underlying motivated behavior and includes many contextual and psychological aspects that have been shown to interact and influence achievement choices and achievement behavior. Although comprehensive, the model has two core components; one expectancy component that corresponds to the question “Can I do this task?” and one value component that corresponds to the question “Do I want to do this task and why?” The expectancy component in the model thus refers to the individual’s beliefs and judgments about his or her capabilities to do a task and succeed at it. The value component in the model refers to the various reasons individuals have for engaging in a task or not (see Eccles & Wigfield, 2002; Eklöf, 2006b). In the present paper, the focus is mainly on the value component in the model (see Eklöf, 2006b, for a more elaborated presentation of the expectancy-value theory).

The results from earlier studies actually focusing on test-taking motivation have been somewhat inconclusive and in many cases, the link between reported level of motivation and actual achievement has been weak. Studies have found that the students are quite motivated even when the test is low-stakes for the students (The Center for Educational Testing and Evaluation, 2001), that raising the stakes does not always contribute to a corresponding rise in motivation and achievement (Baumert & Demmrich, 2001; O’Neil, Abedi, Miyoshi, & Mastergeorge, 2005), and that reported level of test-taking motivation is weakly associated with

subsequent performance (O'Neil et. al., 2005; Zeidner, 1993). On the other hand, other studies have found that the stakes of the test indeed have an impact on motivation and performance (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995).

In summary, it is not clear from previous empirical studies whether the validity of low-stakes tests like TIMSS is threatened by a lack of motivation among the participants because a) it is not clear if the participating students are lacking motivation at all and b) it is not clear whether rated level of test-taking motivation interacts with test performance. The present study explores these issues in a Swedish TIMSS 2003 context.

The study presented in the present paper is also concerned with issues of measurement validity. According to validity theorist Samuel Messick, an important aspect to consider from a validity viewpoint is the "social psychology of the assessment setting" which, according to Messick, "requires careful attention" (Messick, 1989, p. 14). How do respondents react to the tasks that are presented to them in TIMSS 2003? Do they perceive the tests they are about to complete as valid? As important? These are issues that are vital for the validity of interpretation and use of test scores. This is particularly true for studies like TIMSS, which involve so many students from so many countries and cultures. Unfortunately these are also issues that mainly have been forgotten in the research on the validity of large-scale, comparative studies.

Study Objective

The main objectives of the present study was to investigate the reported level of test-taking motivation and the relation between test-taking motivation and mathematics test performance, as well as to explore student perceptions of test stakes and task value in a sample of Swedish eighth-grade students participating in TIMSS 2003.

Method

Participants

A sample (n= 350) of the Swedish eighth-graders participating in TIMSS 2003 participated in the study and completed a test-taking motivation questionnaire before they took the TIMSS test. Of these students, 343 students were valid cases in the TIMSS data base and this sample was the sample used in the present study. The sample consisted of 174 boys (50.7%) and 169 girls (49.3%). Students came from 17 randomly sampled classes that participated in TIMSS. Approximately half the sample was 14 years old at the time of testing; the other half was 15 years old. A previous study based on the same sample of students showed that the present sample was representative of the Swedish TIMSS 2003-participants (Eklöf, 2006a). Of these 343 students, 329 students completed the open-ended questionnaire item also analyzed in the present study. Further, 30 students (15 boys and 15 girls) from this sample agreed to be interviewed about their experience of participating in TIMSS 2003.

Measure, Procedure, and Data Analysis

No established measures of test-taking motivation were available and there were no items asking for test-taking motivation in the TIMSS student background questionnaire. Therefore, a test-taking motivation questionnaire was developed and applied in the Swedish TIMSS context (see Eklöf, 2006a, for a description of the development and validation of this questionnaire). The students completed the test-taking motivation questionnaire before they took the TIMSS test. Also, two items asking for test-taking motivation were added to the TIMSS Student background questionnaire as national options. The students completed the Student Background Questionnaire after the TIMSS mathematics and science test. Further, for validation purposes,

posttest interviews were performed with a smaller sample of the students ($n = 30$) (see Eklöf, 2006b, for a detailed study design).

The Test-Taking Motivation Questionnaire. The Test-Taking Motivation Questionnaire is a self-report instrument developed to measure aspects related to student test-taking motivation (Eklöf, 2006a, see also Eklöf, in press). Eccles and Wigfield's *Expectancy-value model of achievement motivation* was used as the general theoretical basis in the development and interpretation of this questionnaire. The questionnaire was administered before the students completed the TIMSS test booklet. In addition, two items, asking for perceived level of motivation to do well on the TIMSS mathematics and science test, were administered after the TIMSS test as well. According to exploratory factor analysis on data from the present sample (see Eklöf, 2006a), four items (see Table 1, items translated from Swedish) formed a scale that was used as a measure of mathematics test-taking motivation in the present study. This four-item scale was named Test-Taking Motivation (TTM). All items in the scale were measured on a four-point scale with ratings ranging from a highly unfavorable attitude to a highly favorable attitude (e.g., 1 = *not at all motivated*, 4 = *very motivated*) (see Eklöf, in press, for a more detailed presentation of the items in the scale).

Table 1. Items belonging to the mathematics test-taking motivation scale

Variable	Item
TTM	
1	How motivated are you to do your best on TIMSS' mathematics items? (pretest measure)
2	How important is it for you to do your best in TIMSS?
3	How much effort will you spend on answering the mathematics items in TIMSS?
4	How motivated were you to do your best on TIMSS' mathematics items? (post-test measure)

The test-taking motivation questionnaire also contained an open-ended item that was read "*Describe in your own words how motivated you feel to do your best in TIMSS and why*" (translated from Swedish). The item was assumed to generate answers revealing something about the students' perceptions of task value and of the stakes of the TIMSS-test. All students responded to this item before they completed the TIMSS mathematics and science test.

The interviews. Short interviews were performed with 30 students using a semi-structured interview guide. The interview guide contained a list of topics that were to be explored in each interview. The topics included in the interview guide focused on the students' perceptions of TIMSS 2003 in terms of test stakes and task value; their experienced level of test-taking motivation before and during the test, perceived importance of a good performance, and invested effort when completing the TIMSS mathematics and science items. In the interviews, the students were also asked to compare the TIMSS test to regular achievement tests in school.

TIMSS 2003 mathematics test. In TIMSS, each student completes a booklet containing only a sample of the total number of mathematics/science items that are used in the study. It is therefore impossible to calculate a total score that can be compared over populations and subpopulations using raw data. To obtain comparable achievement scores, each student obtains a scaled score, which represents an estimation of his or her score, if the student had answered all

items (see Martin, Mullis, & Chrostowski, 2004). Two different kinds of scores are estimated for each student. One is the national Rasch score (see Eklöf, in press, for analyses of the relation between test-taking motivation and test performance using the Rasch score as dependent variable). The other score consists of five “plausible values” for each student. These plausible values are imputed values obtained through complex item response modelling. The five values an individual obtains are random excerpts from the distribution of possible values for that individual. The mean plausible value is set to 500, with a standard deviation of 100. All achievement results reported in TIMSS internationally are based on these plausible values and in the present study, the first plausible value was used as the dependent variable.

Quantitative Data Analysis

First, descriptive statistics and correlations between variables were computed. Then, student ratings of test-taking motivation were regressed on mathematics score. In this regression, two motivational scales used in TIMSS internationally: Mathematics Self-Concept (MSC) and Students Valuing of Mathematics (VoM) were included in the analysis and held constant in order to investigate whether the test-taking motivation scale explained any variance in the mathematics score not explained by these two motivational variables. All analyses were performed in SPSS. All tests of significance were two-tailed and the alpha level was set to .05.

Qualitative Data Analysis

The wording of the open-ended item in the Test-Taking Motivation Questionnaire (*Describe in your own words how motivated you feel to do your best in TIMSS and why*) was not intended to lead the students into any in advance specified response categories. Accordingly, the analysis of this item initially followed a rather unbiased bottom-up procedure, where response categories were not specified in advance, but the students’ answers were content-analyzed and interpreted mainly in terms of the students’ perceptions of the stakes of the TIMSS test and their perceptions of task value. Then, common themes in the students’ responses were identified and students giving similar responses were joined together in one category. Only the main common themes are reported below.

The analysis of the interviews was mainly descriptive in nature. The students’ responses to, and elaborations on, the topics included in the interview guide were by and large taken at face value and interpreted in terms of task value perceptions and perceptions of test stakes. Student responses to the open-ended item and the interview topics were assumed to reveal more about their perceptions of the TIMSS test than questionnaire items with a closed item format can reveal. The open-ended item and the interviews were also used as tools for validation of the questionnaire results (for more details, see Eklöf, 2006b).

Findings

First, the students’ ratings on the test-taking motivation (TTM) scale as well as their ratings of the individual items in the TTM scale are described. Second, multiple linear regressions exploring the relation between the TTM scale and mathematics score, with students’ ratings of mathematics self-concept and of their valuing of the mathematics subject held constant, are reported. Third, the results from the open-ended item in the test-taking motivation questionnaire are briefly summarized. Fourth, the main findings from the interviews are presented.

Reported Level of Test-Taking Motivation

The TTM scale, which included the four items in Table 1, had a score reliability coefficient of $\alpha = .79$ which is acceptable given that the scale consisted of only four items. The maximum value of the TTM scale was 4.0, and the mean value for the present sample was 3.09 ($SD =$

.55), which indicates that the students in the sample on average reported a fairly high level of test-taking motivation.

As concerns the individual items in the TTM scale, a majority of the students in the sample ($n = 343$) reported that they were either very motivated or somewhat motivated to do their best on TIMSS mathematics items before (89 %) as well as after (76 %) they took the test. A majority of the students further said that it was either very important or rather important for them to do their best in TIMSS (74 %), and that they would spend a lot of effort or a fair amount of effort (90 %) when answering the TIMSS mathematics tasks (see Eklöf, in press, for more detailed results).

Relations Between Ratings of Test-Taking Motivation and Mathematics Score

For the total sample, the TTM scale was significantly but rather weakly correlated with the mathematics score (the 1st plausible value) ($r = .25, p < .01$).

To investigate whether the TTM scale accounted for any variation in the TIMSS mathematics score when other relevant variables were held constant, a regression model was built with the TTM scale and two motivational scales used in TIMSS internationally, Mathematics Self-Concept (MSC) and Students Valuing of Mathematics (VoM) as independent variables. The 1st plausible value in mathematics was used as the dependent variable

According to this model, the three independent variables together explained about 39 % (R^2) of the variation in the mathematics score for the present sample. Most of this variation was explained by the MSC variable ($\beta = .60, t = 12.87, p < .01$). The TTM variable had a positive and statistically significant, though rather weak, relation to the mathematics score when the other independent variables were partialled out ($\beta = .11, t = 2.36, p < .05$). The VoM variable was weakly negatively related to the mathematics score when the effect of the other independent variables was partialled out ($\beta = -.02, t = -.44, p = n.s.$).

The Open-Ended Item

There were 329 valid responses to the open-ended item in the Test-Taking Motivation Questionnaire that was administered to the students ($n = 343$) before they completed the TIMSS mathematics and science test. In their answers, 238 of these 329 students (72 %) expressed themselves in positive terms as concerned their participation in TIMSS and their motivation to do their best. Forty eight students (15 %) gave rather indifferent answers to this open-ended item, and 43 students (13 %) reported a negative motivational disposition towards the TIMSS test.

What motivates students to do their best on a test like TIMSS? Students reporting a positive motivational disposition towards the TIMSS test in their answers to the open-ended item were grouped into three major categories: one category in which the students mainly gave comparative/competitive reasons (CR) for their motivation to do well (they wanted to do their best as they were to be compared to other countries; they wanted to show that Sweden is a prominent country), one category in which they expressed a social responsibility (SR) as the main reason for why they wanted to do their best (they wanted to do their best as they had been chosen for this study; they wanted to do their best to help with the research), and one category in which the students gave more personal reasons (PR) for their motivation to do well (they always do their best, they wanted to do their best to test themselves, to see how much they knew). Below are some examples from students belonging to these categories.

I am fairly motivated to do my best as it is a competition. And you'd rather win.

I am motivated to do my best. I think it is an important test to see how children in different parts of the world work and how they solve problems.

I want to do my best to see how much I have learned over the years

About half the students reporting a positive motivational disposition towards TIMSS were categorized either in the CR category (67 students) or in the SR category (50 students). Nineteen per cent (44 students) of the students reporting a positive motivational disposition towards TIMSS were categorized in the PR category.

Among the 43 students who claimed that they were *not* well motivated to do their best, two main categories were also identified. One of the categories included students who reported the low stakes of the test as the reason for why they were not maximally motivated (the result did not count for their grades; they would never know the results). Sixteen students were coded as belonging to this category. Another group of students reported that they were not motivated as they did not like school, the school subjects tested, or tests in general. Twelve students were coded as belonging to this category. The remaining students gave various reasons as to why they felt motivated/not motivated to do their best in TIMSS.

The Interviews

How motivated were the interviewed students to do their best on TIMSS? During the interviews, all students were asked how motivated they were to do their best on the TIMSS test, before and during the test. Most interviewed students stated that they had been rather motivated to do their best on the TIMSS test. Most students further said their level of motivation was approximately the same during the time they took the tests. A couple of students even maintained that they got even more motivated, once they started to work with the mathematics and science tasks. On the other hand, a number of students reported that their motivation to do their best decreased during the test.

Perceived value of a good performance and amount of effort invested. When asked whether a good performance in TIMSS was important to them and why, a majority of the students reported that they thought TIMSS was a rather important study and that it was rather important for them to do their best, perhaps not always for their own sake, but more for reasons external to themselves. In accordance with the answers to the open-ended item, several of the respondents put forward comparative/competitive reasons for why they felt that a good performance in TIMSS was important: "It felt rather important to kind of show, what we can", "You don't want Sweden to look stupid". On the other hand, a few students did not think that TIMSS or a good performance in TIMSS was that important due to the low stakes of the test: "It didn't feel as important, because it wasn't, it didn't count for the grades".

The students were asked how much effort they invested when answering the mathematics and science items in TIMSS. The typical answer was that they tried the best they could, as there is little point in not trying. Many students also reported that the amount of effort they invested shifted from item to item. Some students exerted more effort on items they found interesting and thought they had a chance on, while a few students exerted more effort on items they found difficult.

Again, a number of students put forward comparative/competitive reasons for why they tried hard to do their best, while others put forward social responsibility reasons: "I wanted to do the best I could because the investigation gets wrong if not everyone try their best".

The TIMSS test compared to other tests in school. In connection with the discussion about the value of a good performance, the students were also asked to compare the TIMSS test to their regular school tests. Here, a number of students said that, although they tried to do their best on TIMSS, they are more motivated to perform well on their regular tests and exert more effort on these tests: "I think that you try harder on a regular test because then you are to be graded and then you want to do a better result".

The students seemed to be well aware of the fact that the TIMSS test did not have any consequences for them personally. Most of them still claimed that they tried their best, however. One student even said he was more motivated to do well and exerted more effort on the TIMSS test than on their regular school tests, because “it’s so big, like, Sweden in the world”. Other students said that they worked as hard on the TIMSS test as they would have on any other test and that it did not matter that the test did not count for their grades. Rather, they felt no pressure and could work in a more relaxed manner: “It wasn’t like a test, it was like an investigation and therefore, you weren’t under stress and so”.

Concluding Remarks

The main purpose of the work presented in this paper was to study aspects related to student test-taking motivation in the TIMSS 2003 context. The issue of test-taking motivation is an issue about validity and about the trustworthiness of test results and therefore, it is an issue worthy of attention in the context of large-scale, comparative studies. Students’ perceptions of TIMSS 2003 and their motivation to perform well in the study were investigated through questionnaire items and interviews, and the association between test-taking motivation and performance was explored through correlation and regression. Swedish eighth-grade students participating in TIMSS 2003 formed the study sample. From the obtained results, some conclusions can be drawn.

First, contrary to commonly held preconceptions, the questionnaire study as well as the interview study indicated that the Swedish students participating in TIMSS 2003 in general were well motivated to do their best on the TIMSS test and that they valued a good performance on the test.

Second, the Swedish mathematics result in TIMSS 2003 does not seem to be affected by a lack of motivation among the participating students, as the students reported that they were well motivated to do their best and as ratings of test-taking motivation were positively but rather weakly related to performance.

Third, although the relationship between test-taking motivation and test performance was rather weak, the test-taking motivation scale still explained some of the variance in the result that could not be explained by the domain-specific motivational variables actually measured in TIMSS.

Findings from the present study thus indicate that a majority of the students valued a good performance and did not perceive TIMSS as a low-stakes test in the sense that it was unimportant and not worth spending effort on. It seems that the fact that the test does not have any personal consequences for the individual does not preclude the individual from attaching value to a good test performance. The student might be intrinsically motivated to do his or her best even when there is no feedback on the individual’s performance. The student might also be extrinsically motivated by the fact that he or she is participating in a large international study where student populations are being compared. In their answers to the open-ended questionnaire item and the interviews, some students gave mainly intrinsic reasons for their motivation or lack of motivation, while others gave mainly extrinsic utility reasons. Many students put forward comparative/competitive reasons or social responsibility reasons as the reason for their motivation. Some students reported the low stakes of the test as detrimental to their motivation but these students constituted a minority of the total sample.

It also seems that most students do seem to care about how they perform in studies like TIMSS, but that they nevertheless are aware of the fact that the test result does not count for them personally. Even if most students claimed that they had been well motivated to do their best on the TIMSS test and that they tried the best they could to do their best, some of these students still claimed that they would have tried harder had it been a regular test.

The present paper explores an issue that has largely been ignored in the literature: student motivation to do their best in low-stakes contexts. Test-taking motivation can be relevant to the validity of interpretation and use of test results and ignoring the test-takers and their views is not

compatible with modern conceptions of validity (Messick, 1989). If test-taking motivation would contaminate responses to tests, this is an example of construct-irrelevant variance that affects the validity of test score interpretation (Benson, 1999; Messick, 1989). It follows that students' reactions to tests and their task-specific motivation to do well should be acknowledged in the interpretation and use of test scores, also in TIMSS.

An obvious limitation of the present paper is that the study only includes Swedish TIMSS participants and a possible bias in the international comparisons due to varying levels of test-taking motivation is not possible to study. Nothing is known about student test-taking motivation in other countries participating in TIMSS and other large-scale, international studies. It is possible that level of test-taking motivation differs between countries and cultures and therefore, cross-country comparisons of test-taking motivation and of the effect of test-taking motivation on test achievement are an urgent area of future research in order to strengthen the validity of interpretation of results from studies like TIMSS. Other systematic group differences in test-taking motivation such as gender differences and differences between ethnical and social groups in national and international contexts are also worthy of systematic investigation.

It should be noted that the instrument used in the present study is somewhat tentative and needs continued development and continued validation. Nevertheless, the obtained results imply that even a short measure of test-taking motivation can provide important information, and it would be possible to include a measure of student effort, performance expectancies, perceived importance of a good performance, or level of motivation in the TIMSS test battery. Including such a measure could contribute to the understanding of score meaning, and to the validity of score based inferences.

The present paper also illustrates the possibility of adapting or adding national options to the TIMSS test battery. Each participating country has the possibility of adding national options to the questionnaires and more nations, or possibly collaborating clusters of nations, should take advantage of this possibility. All nations participating in TIMSS have unique characteristics and adding national options to the anchor instruments administered by TIMSS could mirror these unique characteristics and enable large-scale investigation of questions of particular interest.

References

- 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.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*, 10-17.
- Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology, 82*, 300-310.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109-132.
- Eklöf, H. (2006a). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*.
- Eklöf, H. (2006b). *Motivational beliefs in the TIMSS 2003 context: Theory, measurement and relation to test performance*. Doctoral Dissertation, Department of Educational Measurement, Umeå University, Umeå.
- Eklöf, H. (in press). Test-Taking Motivation and Mathematics Performance in TIMSS 2003. *International Journal of Testing*.
- Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-46). Hillsdale, NJ: Lawrence Erlbaum.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Vol. 3, pp. 13-103). New York: Macmillan/American Educational Research Association.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, *50*, 741-749.
- Mislevy, R. J. (1995). What can we learn from international assessments? *Educational Evaluation and Policy Analysis*, *4*, 419-437.
- Nevo, B., & Jäger, R. S. (1993). *Educational and psychological testing: The test taker's outlook*. Stuttgart: Hogrefe & Huber Publishers.
- O'Neil, H. F. Jr., Sugrue, B., Abedi, J., Baker, E. L., & Golan, S. (1997). *Final report of experimental studies on motivation and NAEP test performance* (CSE Technical Report 427). Los Angeles, CA: University of California, CRESST.
- O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment*, *10*, 185-208.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33-40.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (2nd Ed.). New Jersey, NJ: Merrill Prentice Hall.
- Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, *29*, 6-26.
- The Center for Educational Testing and Evaluation (2001). *Student test taking motivation and performance: Grade 10 mathematics and science and grade 11 social studies*. Research Report. University of Kansas, KS: School of Education.
- Törnqvist, B. (1998). *TIMSS: Teknisk rapport* [TIMSS: Technical report] (Report No. 132). Umeå: Umeå University, Department of Educational Measurement.
- Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement*, *30*, 1-21.
- Wigfield, A., & Eccles, J. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp 92-120). New York, NY: Academic Press.
- Wise, S. L., & DeMars, C. E. (2003). *Examinee motivation in low-stakes assessment: Problems and potential solutions*. Paper presented at the annual meeting of the American Association of Higher Education Assessment Conference, Seattle, WA.
- Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, *8*, 227-242.
- Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education*, *8*, 341-351.
- Zeidner, M. (1993). Essay versus multiple-choice type classroom exams: The student's perspective. In B. Nevo & R. S. Jäger (Eds.), *Educational and Psychological Testing: The Test Taker's Outlook* (pp. 85-111). Stuttgart: Hogrefe & Huber Publishers.