

# 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 the test-taking motivation of students in the Swedish TIMSS 2003 context. Swedish Grade 8 students formed the study sample, and the focus was on mathematics. Test-taking motivation was measured using self-report measures (Likert-scale and open-ended questionnaire items), and reported level of test-taking motivation was regressed on test score. The quantitative part of 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, although rather weakly, related to mathematics achievement. Qualitative analysis of an open-ended questionnaire item mainly corroborated obtained findings but added some complexity to the results. In their answers to the open-ended item, most students reported they were well motivated to do their best in TIMSS and that they valued a good performance. Many students 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 mathematics result in TIMSS 2003 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.

## INTRODUCTION

This paper presents parts of a research project being conducted on the Swedish dataset from the Trends in Mathematics and Science Study (TIMSS) 2003. The project is exploring different aspects of student achievement motivation from a measurement perspective and a validity perspective (Eklöf, 2006a, 2006b, 2007).

Student motivation is an important 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, that is, 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 toward a test is often assumed to be a necessary though not sufficient condition for good test performance (Cronbach, 1988; Robitaille & Garden, 1996; Wainer, 1993; Zeidner, 1993). Messick (1988) noted that 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 differ systematically in level of motivation, and if less motivated students are disadvantaged in that they score below their actual proficiency level, then test-taking motivation is a possible source of bias (Baumert & Demmrich, 2001; Mislevy, 1995; O'Leary, 2002; O'Neil, Sugrue, Abedi, Baker, & Golan, 1997; Robitaille & Garden, 1996; Wainer, 1993; Wise & Kong, 2005; Zeidner, 1993) and is 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 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 relatively 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 for the test-taker, that is, low-stakes tests, have been assumed—and in some cases shown to be—associated with a decrease in motivation and performance (Wise & DeMars, 2005; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995). 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, though rarely empirically tested, 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).

The result on the TIMSS test has no impact on student grades in mathematics or science. Also, the results in TIMSS are mainly summarized at a national level and individual results are not given to the students or the schools. Thus, the students and their teachers, parents, and peers never know the result for an individual student. However, 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, even though 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, as well as theoretically and methodologically. The results from earlier studies focusing on test-taking motivation have been somewhat inconclusive, and, in some cases, the link between reported level of motivation and actual achievement has been weak. Studies have found that students are quite motivated even when the test is low stakes for them (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).

However, according to a number of other studies, the stakes of the test do have an impact on motivation and performance (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; Sundre & Kitsantas, 2004; Wise & DeMars, 2005; Wolf & Smith, 1995; Wolf et al., 1995). Brown and Walberg (1993) found that raising the stakes of the test by giving special instructions (emphasizing the importance of doing well and telling the students the test results were to be used for evaluation) raised the typical student's test score from the 50th to the 62nd percentile. Further, in a summary of 12 studies investigating the effects of student test-taking motivation on test performance, Wise and DeMars (2005) found that well-motivated students outperformed less-motivated students with an average effect size exceeding half a standard deviation. Wise and DeMars (2005) were also able to show a near zero correlation between self-reports of test-taking motivation and measures of ability (as measured by SAT scores), a finding that suggests there was no confounding of motivation and ability.

As far as is known, no one has tried to study student test-taking motivation in the actual context of a large-scale study like TIMSS. It is therefore unclear if the validity of the tests used in TIMSS has been compromised by a lack of motivation among the participants. We do not know if (a) the participating students lacked motivation, and/or (b) the rated level of test-taking motivation correlated with test performance. The present study explores these issues in a Swedish TIMSS 2003 context. Note that because only Swedish TIMSS participants are considered in the present study, no comparisons between countries can be made at this point.

## Study Objective

The main objectives of the present study were (a) to investigate the reported level of test-taking motivation and the relationship between test-taking motivation and mathematics test performance, and (b) to explore student perceptions of test stakes and task value by analyzing data from a sample of the Swedish Grade 8 students who participated in TIMSS 2003.

## METHOD

### Participants

A sample ( $n = 343$ ) of the Swedish Grade 8 students who participated in TIMSS 2003 took part in the present study. The sample consisted of 174 boys (50.7%) and 169 girls (49.3%). Approximately half the sample was 14 years old at the time of testing; the other half was 15 years old. Students came from 17 classes that participated in TIMSS. The classes sampled to participate were in schools monitored by national quality control observers during the administration of the TIMSS test (see Gonzalez & Diaconu, 2004). The sampling of schools was done to ensure a regional spread and a mixture of urban and rural and large and small schools. 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).

### Measures of Test-Taking Motivation

Quantitative as well as qualitative measures of student test-taking motivation were obtained. The respective measures are presented below (see Eklöf, 2006b, for a detailed study design).

#### ***a) The Test-Taking Motivation Questionnaire***

Because no established measures of test-taking motivation were available, and because the TIMSS test battery had no items that asked about test-taking motivation, I developed a test-taking motivation questionnaire and applied it in the Swedish TIMSS context (see Eklöf, 2006a, for a description of the development and validation of this questionnaire). I used Eccles and Wigfield's Expectancy-value Model of Achievement Motivation (Eccles & Wigfield, 2002; Pintrich & De Groot, 1990; Wigfield & Eccles, 2002) as the general theoretical basis when developing the questionnaire.

The questionnaire was administered before the students completed the TIMSS test booklet. The questionnaire consisted of 24 items, 20 of which pertained to student motivation and general attitudes toward school, mathematics, and science (19 Likert-scale items and one open-ended item; see below). Five items in the questionnaire concerned student mathematics test-taking motivation, and three of these items were associated with a common factor according to factor analysis (see below). It is important to note that because the Test-Taking Motivation Questionnaire is a newly developed instrument, it and the scale used are somewhat tentative and need ongoing development and continued validation.

One of the items in the questionnaire was an open-ended item that instructed the students to write down, in their own words, (a) how motivated they felt to do their best on the TIMSS test, and (b) the reason why they felt the way they did. The open-ended item was assumed to generate answers revealing something about the students' perceptions of task value and of the stakes of the TIMSS test. Using an item like this could thereby allow for a more in-depth understanding of student test-taking motivation in the TIMSS context.

### ***b) Augmentation of the TIMSS Student Questionnaire***

All students participating in TIMSS completed a questionnaire after they had finished the booklet containing the mathematics and science assessment items (see Martin, Mullis, & Chrostowski, 2004). Two items, which asked the students how motivated they had been to do their best on (respectively) the mathematics test and the science test, were added to the Swedish Student Questionnaire as national options. Because I considered a brief post-test measure of test-taking motivation desirable, and because it was not possible to administer the Test-Taking Motivation Questionnaire before as well as after the students completed the TIMSS booklets, I used these two augmented items in the analysis as post-test measures of test-taking motivation.

### ***c) The TTM Scale***

According to exploratory factor analysis on data from the present sample (see Eklöf, 2006a), four items formed a scale that could be used as a measure of mathematics test-taking motivation in the present study. These were:

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

Items 1 to 3 were taken from the Test-Taking Motivation Questionnaire; Item 4 was taken from the TIMSS Student Questionnaire. 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 (1 = *not at all motivated*, 4 = *very motivated*) (see Eklöf, 2007, for a more detailed presentation of the items in the scale).

## **TIMSS 2003 Mathematics Test**

In the TIMSS surveys, each student completes a booklet containing only a sample of the total number of mathematics/science items used in the study. It is therefore impossible, when using raw data, to calculate a total score that can be compared over populations and sub-populations. To allow analysts to obtain comparable achievement scores, each student receives a scaled score that represents an estimation of the student's score if he or she had answered all items (see Martin et al., 2004).

Two different kinds of scores are estimated for each student. One is the national Rasch score (see Eklöf, 2007, for analyses of the relationship between test-taking motivation and test performance using the Rasch score as the dependent variable). The other score consists of five “plausible values” for each student. These values are imputed values obtained through complex item-response modeling. 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 I also used the plausible values in mathematics as the dependent variable in this present study.

### **Quantitative Data Analysis**

My first step was to compute descriptive statistics and correlations between variables using SPSS 15.0. I then regressed student ratings of test-taking motivation on the mathematics score. This regression included two motivational scales used internationally in TIMSS—mathematics self-concept (MSC) and student valuing of mathematics (VoM). These two scales are domain-specific measures of achievement motivation, in contrast to the TTM scale, which measures task-specific motivation. I held the MSC and the VoM scales constant in the regression analysis in order to determine if the test-taking motivation scale explained any variance in the mathematics score not explained by these two motivational variables. The *IEA IDB Analyzer*, a plug-in program for SPSS supplied by the IEA Data Processing and Research Center in Germany, was used to run the regressions. This program for analyzing IEA data uses all five plausible values as dependent variables and makes sure that correct sample weights are used and that standard errors are correctly calculated through the use of the jackknifing method (see Martin et al., 2004). All tests of significance were two-tailed and the alpha level was set to 0.05.

### **Qualitative Data Analysis**

My analysis of the open-ended item in the Test-Taking Motivation Questionnaire followed a two-stage procedure. First, I categorized the student responses to the open-ended item as “positive”, “negative,” or “indifferent,” depending on how the students expressed themselves in terms of their motivational disposition toward the TIMSS test. Second, I analyzed the reasons students gave for their motivation or lack of motivation, identified common themes in the students’ responses, and joined similar responses from students in one category.

For reliability purposes, I engaged a second rater to analyze the responses to the open-ended item. This external rater was blind to my categorization but did receive a scheme containing the main categories and instructions for coding. There was a high degree of convergence between us. We disagreed on only 24 (7%) of the 329 valid responses to the open-ended item.

## FINDINGS

The findings are presented as follows:

1. A description of the students' ratings on the TTM scale as well as their ratings of the individual items.
2. An account of the multiple linear regressions exploring the relationship between the TTM scale and the mathematics score, with students' ratings of mathematics self-concept and their valuing of the mathematics subject held constant.
3. A summary of the results from the open-ended item in the Test-Taking Motivation Questionnaire.

### Reported Level of Test-Taking Motivation

The TTM scale had a score reliability coefficient of  $\alpha = 0.79$ , 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 ( $SD = 0.55$ ), which indicates that the students in the sample reported on average a fairly high level of test-taking motivation.

In regard to 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%) taking the test. A majority of the students said that it was either very important or somewhat 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, 2007, for more detailed results).

### Relationships between Ratings of Test-Taking Motivation and Mathematics Score

For the total sample, the TTM scale positively but rather weakly correlated with the mathematics score ( $r = 0.25$ ,  $p < 0.01$ ). As noted above, investigation of whether the TTM scale accounted for any variation in the TIMSS mathematics score when other relevant variables were held constant required building a regression model in which the TTM scale and the two motivational scales used internationally in TIMSS—that is, mathematics self-concept (MSC) and student valuing of mathematics (VoM)—were independent variables.

According to this model, the three independent variables together explained about 42% ( $R^2$ ) of the variation in the mathematics score for the present sample. Most of this variation was explained by the MSC variable ( $\beta = 71.34$ ,  $SE = 5.91$ ,  $t = 12.06$ ,  $p < 0.01$ ). The TTM variable had a positive and statistically significant, though rather weak, relationship to the mathematics score when the other independent variables were partialled out ( $\beta = 11.76$ ,  $SE = 4.78$ ,  $t = 2.46$ ,  $p < 0.05$ ). The VoM variable was not related to the mathematics score when the effect of the other independent variables was partialled out ( $\beta = -7.60$ ,  $SE = 6.41$ ,  $t = -0.44$ ,  $p = n.s.$ ).

## The Open-Ended Item

There were 329 valid responses to the open-ended item (96% of the total number). During the first phase of analysis, I coded student answers as “positive,” “negative,” or “indifferent” in terms of motivational disposition toward the TIMSS test. According to this analysis, 238 (72%) of the 329 students expressed themselves in positive terms concerning their participation in TIMSS and their motivation to do their best. Forty-eight students (15%) expressed a rather indifferent attitude, and 43 students (13%) reported a negative motivational disposition toward the TIMSS test.

The second phase involved analyzing the students’ explanations as to why they felt motivated or not. I grouped the students reporting a positive motivational disposition toward the TIMSS test in their answers to the open-ended item into four major categories:

- *Category 1:* The students in this category expressed a *social responsibility/comparative perspective* as the main reason for wanting to do their best. These students wanted to do their best because they had been chosen for this study; they wanted to do their best to help with the research; and they wanted to do their best because so many countries were to be compared.
- *Category 2:* Here, the students mainly gave *competitive reasons* for their motivation to do well. They wanted to show that Sweden is a prominent country, and they wanted to “win this competition.”
- *Category 3:* The students in this category mainly gave *personal, intrinsic* reasons for their motivation to do well. They said they always did their best in order to feel good about themselves, and they wanted to do their best to test themselves—to see how much they knew.
- *Category 4:* In this category, students reported that they felt motivated simply because it was “*fun*” or “*interesting*” to participate in the study.

The first two, slightly overlapping, categories thus consisted of students reporting mainly extrinsic factors as motivating. The latter two categories consisted of students referring to mainly intrinsic factors as motivating.

Among the students who claimed that they were not well motivated to do their best, two further categories could be identified. The first included students who reported the *low stakes* of the test as the reason why they did not feel maximally motivated (the result did not count for their grades; they would never know the results). The second group of students reported that they were not motivated because they *did not like school, the school subjects tested, or tests in general*. Also, a fair number of students could not be placed in any of the above categories, either because they gave reasons that did not easily fit into these categories or because they did not give any particular reason for their motivation or lack of motivation. Table 1 shows the number of students placed within each category.

It is important to note here that 30 students from the sample used in the present study also agreed to be interviewed about their perceptions of the TIMSS test. The

interview results corroborated, to a high degree, the results from the analysis of the open-ended item, which adds credibility to the findings from the questionnaire study (for detailed findings from these interviews, see Eklöf, 2006b).

**Table 1: Student explanations for why they felt motivated/not motivated to do their best on the TIMSS test according to their answers to the open-ended questionnaire item ( $n = 329$ )**

Primary reason given	<i>n</i>	%	Examples of student answers
Personal/Intrinsic	58	18	<i>I want to do my best to see how much I have learned over the years.</i>
Fun/Interesting	39	12	<i>I feel very motivated to do this test. I feel no pressure, but it feels fun to do it.</i>
Comparative/Social Responsibility	84	26	<i>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>
Competitive	38	12	<i>I am fairly motivated to do my best, as it is a competition. And you'd rather win.</i>
Low stakes of the test	31	9	<i>Not particularly motivated, as it doesn't affect my grades.</i>
Dislike school subjects/Tests in general	12	3	<i>Not very motivated, as I don't like math and science.</i>
Various other reasons/No reason given	67	21	<i>I don't have to go to English class.</i>

## SUMMARY AND CONCLUDING REMARKS

The main purpose of the work presented in this paper was to study student test-taking motivation in the TIMSS 2003 context. Few studies, if any, have investigated student test-taking motivation in the actual context of a large-scale, international comparative study like TIMSS. The issue of test-taking motivation on low-stakes tests is an issue about validity and about the trustworthiness of test results. As such, it is an issue worthy of attention in the context of large-scale, comparative studies.

Two main conclusions can be drawn from the obtained results. First, the quantitative as well as the qualitative analyses indicate that the Swedish students participating in TIMSS 2003 were generally 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 have been affected by a lack of motivation

among the participating students for two reasons. First, most of the students reported that they were well motivated to do their best, and second, ratings of test-taking motivation related weakly to test performance. In essence, the obtained results indicate that TIMSS mathematics scores for Swedish students were minimally biased by test-taking motivation and so can be interpreted as representing mathematics knowledge.

The results obtained in the present study do not support findings from a number of previous studies that report test stakes influence motivation and performance, and that low-stakes tests are associated with a low level of test-taking motivation as well as a lower than expected performance. Students in the present sample did not report a low level of test-taking motivation, probably because most students did not perceive the TIMSS test as a low-stakes test in the sense that it was unimportant and not worth spending effort on. Some students reported they were intrinsically motivated to do their best while other students reported they were extrinsically motivated, for example by the fact that they were representing Sweden in a study where countries would be compared with one another. The students who did report the low stakes of the test as detrimental to their motivation constituted a minority of the total sample.

It is noteworthy that children in Grade 8 in the current Swedish education system have little experience of extensive testing in school. They are therefore not used to lengthy, standardized tests like those used in TIMSS. Further, these students do not receive grades for their schoolwork until they are in Grade 8. Therefore, Swedish eighth-graders are not “fed up” with testing and assessment, and this might explain their positive attitude toward doing the “non-consequential” TIMSS test. Also, the importance of doing well on a test like TIMSS has to be accentuated by someone, and it seems that the Swedish teachers and school leaders had succeeded in informing the students about TIMSS and in motivating the students to do their best.

Even if the test-taking motivation scale used in the present study seems psychometrically sound (see Eklöf, 2006a), it is important to acknowledge that the scale is a newly developed self-report scale, that the properties of the scale have not been extensively validated, and that the results depend on the honesty of the respondents. The low correlation between test-taking motivation scores and mathematics achievement may possibly be due to an inability of the scale to capture relevant aspects of student test-taking motivation. The restriction of range in the TTM variable, and the resulting ceiling effect, may also have contributed to the weak association between test-taking motivation and achievement. Future studies are needed to further the development of the TTM scale.

An obvious limitation of the present paper is that the study includes only Swedish TIMSS participants and that it therefore is not possible to study potential bias in the international comparisons due to varying levels of test-taking motivation. Nothing is known about student test-taking motivation in other countries participating in TIMSS and in other large-scale, international studies. Level of test-taking motivation may well differ between countries and cultures. Cross-country comparisons of test-taking

motivation and of the effect of test-taking motivation on test achievement therefore constitute an important area of future research. Other systematic group differences in test-taking motivation such as age differences, gender differences, and differences between ethnic and social groups in national and international contexts are also worthy of systematic investigation.

The present paper also illustrates the possibility for each participating country to add (or adapt) national options to the TIMSS test battery. The findings of this present study and the lack of research overall relative to the issues discussed in this paper strongly suggest that more nations, or possibly collaborating clusters of nations, should take advantage of this possibility. Moreover, because nations participating in TIMSS have unique characteristics, adding national options to the anchor instruments administered by TIMSS could allow mirroring of these unique characteristics and enable large-scale investigation of questions of particular interest. It would also be possible to include a measure that combines student effort, perceived importance of a good performance, and/or level of motivation in the TIMSS test battery. Including such a measure could contribute not only to our understanding of score meaning but also to the validity of score-based inferences.

## 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.
- Brown, S. M., & Wahlberg, H. J. (1993). Motivational effects on test scores of elementary students. *Journal of Educational Research, 86*, 133–136.
- 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). Lawrence, KS: School of Education, University of Kansas.
- 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, 66*, 643–656.
- Eklöf, H. (2006b). *Motivational beliefs in the TIMSS 2003 context: Theory, measurement and relation to test performance*. Unpublished doctoral dissertation, Department of Educational Measurement, Umeå University, Umeå.
- Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*, 311–326.

- Gonzalez, E., & Diaconu, D. (2004). Quality assurance in the TIMSS 2003 data collection. In M. Martin, I. V. S. Mullis, & S. J. Chrostowski (Eds.), *TIMSS 2003 technical report* (pp. 142–171). Chestnut Hill, MA: Boston College.
- Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: Boston College.
- 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. (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'Leary, M. (2002). Stability of country rankings across item formats in the Third International Mathematics and Science Study. *Educational Measurement: Issues and Practice*, *21*, 27–38.
- O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment*, *10*, 185–208.
- O'Neil, H. F., 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: CRESST, University of California.
- 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.
- Robitaille, D. F., & Garden, R. A. (1996). *Research questions and study design* (TIMSS Monograph No. 2). Vancouver: Pacific Educational Press.
- 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.
- 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: Academic Press.
- Wise, S. L., & DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and possible solutions. *Educational Assessment*, *10*, 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*, 163–183.

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.

