Students' Test Motivation in PISA: The Case of Norway

Therese N. Hopfenbeck, Associate Professor, OUCEA

Oxford University Centre of Educational Assessment, Department of Education, University
of Oxford

Corresponding author: therese.hopfebeck@education.ox.ac.uk

Marit Kjærnsli, Associate Professor,

Department of Teaching and School Research, University of Oslo

Students' Test Motivation in PISA: The Case of Norway

Abstract

Do students make their best effort in large-scale assessment studies such as the *Programme for International Student Assessment* (PISA)? Despite six cycles of PISA surveys from 2000 to 2015, empirical studies regarding students' test motivation and experience of the tests are sparse. The present study examines students' test motivation in PISA, and how closely students' achievement correlates with students' reported test motivation. A total of 40 students from eight schools that participated in the PISA study in 2006, 2009 and 2012 were interviewed. In addition, questionnaire data from a total of 9,400 students participating in PISA 2009 and 2012 were collected. Findings of this study indicate that students overall were motivated to do their best in the PISA study despite the fact that these low-stakes tests have no impact on students' grades or future school entrance. In contrast to claims in the media, where it has been suggested that Norwegian students are more relaxed towards the PISA test than students in other countries (e.g. Dagsavisen, December 5th 2013), students report that they are motivated to do their best, but girls report significantly higher test motivation than boys. Despite the policy influence of PISA and the large number of countries participating, very few studies have reported findings from students' perspectives regarding their test motivation and experience of the test. The present study aims to fill that gap.

Keywords: Test motivation, PISA, low-stake tests, student effort, gender.

Introduction

There is an ongoing discussion whether large-scale assessment studies, such as *Trends in International Mathematics and Science Study* (TIMSS), *Progress in International Reading Literacy Study* (PIRLS) and *Programme for International Student Assessment* (PISA) have biased test scores due to students' lack of test motivation (Baumert & Demmrich 2001).

The case reported here uses data from students in Norway who participated in the PISA survey in 2006, 2009 and 2012. PISA is an international comparative standardised

assessment survey that seeks to measure to what extent 15-year-olds 'are prepared to meet the challenges of today's knowledge societies' (OECD, 2009). The first cycle of PISA was carried out in 2000, initiated by the Organisation for Economic Co-operation and Development (OECD) in response to the OECD's member countries' need for comparative data (Heyneman 1993), but it has since then also become a major policy tool for other countries outside the OECD (Grek 2009). PISA is a triennial international survey that measures 15-year-olds' competencies in reading, mathematics and scientific literacy. In PISA the target population is age-based; and include students born in the same year. The students are aged between 15-years 3 months and 16 years 2 months at the time of the assessment. The population includes all students attending school at that age, whether it is a private or a public school and regardless of whether it is in an academic or vocational programme. The schools are selected randomly based on official school statistics in each country. The selection is done by an international center. Then, within the selected schools, the students are randomly sampled from a list with all students in the eligible age at the school. The intention of PISA is to be as inclusive as possible. However, some students may not be able to participate for various reasons. These students can be excluded after they have been randomly sampled. The reason for non-participation must follow the international criteria and the reason must be documented for each student. In Norway 30 students from each school are randomly sampled, and in schools with 30 or fewer students, all of them are sampled. In PISA 2012 almost 4700 Norwegian students from 196 schools took part in the test. The international requirement is minimum 4,500 students from 150 schools. All countries must attempt to maximise the coverage of 15-year-olds students and the countries are not permitted to exclude more than 5% either by excluding schools or by excluding students within schools. All exclusions are documented in detail in the international report (OECD 2013).

In order to study the trends in education, the assessment takes place every three years and each time reading, science or mathematical literacy are included, in which one will be the main domain (OECD 2001).

In Norway, the first PISA results came as a surprise, since it was expected that Norway would perform better. It has been suggested that lack of motivation could be a reason for low results in countries such as USA, Germany and Norway (Haenisch & Nyroos 1998; O'Neil, Abedi, Miyoshi & Mastergeorge 2005; Sjøberg 2007, 2014), when the results have been below what had been expected. Since these tests are considered low-stakes for students, have no consequences for them as individuals, and do not influence students' results, researchers have been concerned that students do not make much effort. As a result, these tests can underestimate students' abilities (Holliday and Holliday 2003, Sjøberg 2007). In contrast to high-stakes tests, such as entrance tests and exams, students do not receive information about their test scores on large-scale tests such as TIMSS, PIRLS and PISA, which in itself can lead to less motivation. If lack of test motivation is not measured and controlled for, it might be a threat to the validity of the interpretation and use of test scores (Thelk, Sundre, Horst, & Finney 2009; Eklöf 2010). Barry et al. (2010) argue that test effort is likely to be high in high-stakes testing when examinees complete tests for which there are personal consequences, but alternatively, they suggest that test effort will be more variable in low-stakes contexts. Given the increasing use of the outcomes of international low-stakes tests such as PIRLS, TIMSS and PISA, it is important to study students' test motivation in these tests. The publication, *Standards for Educational and Psychological Testing*, developed jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), also

recommends that test developers examine students' test effort when interpreting the test scores (American Educational Research Association, American Psychological Association, National Council on Measurement in Education 1999). Although students' lack of test motivation has been emphasised as a possible explanation for the low test results in PISA in the Nordic countries, such as Sweden and Norway, the claims are often anecdotal or are accusations in newspapers, with little or no empirical grounding. The following study aims to contribute to the research field of test motivation with a unique dataset of empirical research relating to three cycles of PISA in Norway, including both quantitative and qualitative data.

Theoretical background

There are few empirical research studies which have investigated students' test motivation in large-scale assessment studies. In Germany, Brunner, Artelt et al (2007) conducted a study with the aim of investigating whether teachers' coaching of PISA items would influence the achievement scores in mathematics and reading. The researchers investigated the coaching and pre-test effects by using a quasi-experimental design in a sample drawn from German Hauptschule (leading to vocational training) and Gymnasium schools (leading to university), including 72 teachers and a total of 1,323 students. The analysis showed that pre-testing in reading or mathematics did not have a significant positive influence on the PISA scores, and only a combined effect of pre-testing and coaching had a substantial effect on students' performance. Although it was found that the weakest performing students gained the most from coaching and pre-testing, the researchers concluded that they did not consider pre-testing and coaching to be a threat to the validity of the PISA study (Brunner, Artelt et al. 2007).

A second study investigated whether the use of incentives could increase students' effort and performance (O'Neil, Sugrue & Baker 1996). The researchers manipulated various incentives (e.g. financial reward, tasks and instructions) and found that only the financial reward increased performance and reported effort in Year 8s. However, the offer of money did not have an impact on either the performance or reported effort of the Year 12s. It was suggested that the amount of money offered was not enough for the Year 12s compared to the Year 8s.

In a later study, O'Neil, Abedi et al. (2005) reported further investigations using financial incentives in a US study of 415 students from nine schools, including a control group. The researchers provided an increased amount of money (\$10 per correct item, maximum \$20 to each student) for Year 12, who were administered on the third TIMSS mathematics items. The researchers found that, although the incentive group showed a higher level of test-tasking effort, they did not perform better than the non-incentive control group. The speculations about the lack of test motivation in US students as a cause of the poor results on the TIMSS study have been rejected (O'Neil, Abedi et al. 2005). The US study had similar findings to another study conducted by German researchers. Baumert and Demmrich (2001) reported that financial rewards did not improve either performance or reported effort in a study conducted on a total of 467 Year 9 students from eight schools in Germany. They conducted an experimental study with the aim of investigating whether PISA scores could be improved by (i) informational feedback, (ii) grading or (iii) financial rewards. Four student groups were tested, three with the different conditions described, and in the fourth group, students received instructions very similar to what test administrators tell student participants in PISA. The main finding reported was that no differences were found with respect to intended and invested effort or to test performance, and the three different student groups showed similar

levels of effort and motivation as the fourth group which was given the ordinary PISA instruction.

Other studies have shown different results, however. Wise & DeMars (2005) synthesised 12 empirical studies of the effects of examinee motivation on test performance and found that that the average difference between low motivated and highly motivated students were .58 standard deviation. In other words, test motivation could have a significant effect on students' test scores.

Finally, Eklöf et al. (2014) reported an analysis of a cross-country study measuring students' relationships between reported effort and performance on the TIMSS Advanced mathematics test in 2008 in Sweden, Norway and Slovenia. Students in each country reported low motivation, but there was a significant relationship between reported effort and test performance. The researchers recommended that test motivation be monitored in low-stakes assessments (Eklöf et al. 2014).

As is evident from the literature review, there are mixed results from previous studies on test motivation and scores on low-stakes testing, with some studies indicating that test motivation and effort are related to higher performance, while other studies have not found such a relationship. What researchers have found is that females report higher levels of motivation to do their best (Eklöf 2007), while there tends to be a stronger relationship between high motivation and performance for males (Eklöf, 2007; Wise, Kinsbury, Thomason & Kong, 2004).

The OECD has also made attempts to measure students' test motivation and effort. For the second PISA cycle, in 2003, OECD developed and included a measurement of effort, asking

the 41 participating countries to include it in the booklets given to students (OECD 2001). The ‘effort thermometer’ contained three 10-point scales: a *High Personal Effort scale*, a *PISA Effort scale* and a *School Mark Effort scale* (see Appendix 1). Students were asked to imagine an actual situation that was highly important to them personally, and then report how much effort they would put into that task to do well. They were then asked to report how much effort they put into doing the PISA test compared to the situation they had just imagined; and finally, they were asked to report how much effort they would have invested if their marks from PISA had been counted in their school marks (OECD 2007: 52). After the PISA 2003 cycle, the results showed that overall, students across all participating countries answered that they would have put more effort into the test if it had influenced their school marks, but it was also found that the expenditure of reported effort by students was fairly stable across countries (Butler & Adams 2007).

The Norwegian context

Norway is a country with 5 million inhabitants, and an education system very similar to other Nordic countries such as Denmark, with free schools for all. Even though Norway has been at the top of life quality indexes in recent years, it has only performed around the OECD average in PISA through all four cycles until 2012.

As in Germany (Ertl 2006), Norway experienced the first PISA results as a shock (Elstad et al. 2009), and lack of test motivation has been suggested as a plausible explanation for the lower than expected results. Norwegians generally believed that Norwegian schools were good before PISA, but there were not much data available to inform the policy level of the quality of Norwegian schools. National tests were introduced after the PISA shock and a quality monitoring system with several different indicators has been put in place. The

changes in the Norwegian system can be said to have increased the numbers of tests students are taking in Norway, from having no national tests in 2003, to having national tests from 2003 onwards in years 5, 8 and 9 in science, mathematics, reading in Norwegian and reading in English. The changes in the frequency of tests might also have an effect upon students' test motivation in general, since they risk suffering test fatigue, also known as 'cognitive fatigue' (Ackerman & Kanfer 2009). Recently published studies in Norway also indicate that teachers experience an increased accountability system in Norway with more tests and external pressure (Mausethagen 2013). Such changes could possibly have an impact upon students' motivation to do their best on low-stakes tests which do not have any impact on students' school marks.

The survey and test motivation scale

In order to further explore students' test motivation in the PISA test, Norway included six test motivation items as a national option in the student questionnaire booklet in PISA 2009. These items were developed from already well-known scales described by Eklöf (2007) in her work on TIMSS students in 2003. In addition we adapted some of the items from the *Student Opinion Scale* (SOS) ¹. The Student Opinion Scale instrument is comprised of 10 items using a five-point Likert scale ranging from 1 "Strongly Disagree" to 5 "Strongly Agree." The items were recoded so that the higher scale scores reflected more positive attitude. The instrument measures two factors, *Effort* and *Importance*, with five items in each scale. However, the aim of the research of Sundre and Moore (2007) was to have one total score of test motivation when summarising the ten items of effort and importance, and this resulted in the final scale, the Student Opinion Scale.

¹Sundre and Moore have published articles and scoring guides on their questionnaire, and offer it for free use in research: <http://www.jmu.edu/assessment>.

The test motivation scale used in the Norwegian study consists of six items. Four of the items measure *effort*, two measure *importance*:

The following items measure *effort*:

- *I was motivated to do my best in the PISA test*
- *I put in a good effort throughout the PISA test*
- *While taking the PISA test, I was able to persist even if some of the tasks were difficult.*
- *I made my best effort on the PISA test*

The following items measure *importance*:

- *Doing well in the PISA test was important to me*
- *Doing well in PISA meant a lot to me*

The six motivation items showed an overall alpha of .85, indicating high internal consistency. When we analysed the reliability of the construct containing only the first four effort items the alpha was .84, while when we included the two importance items the alpha increased only to .85. Based upon these analyses, we decided to use the four effort items and the two importance items as a motivation construct in the PISA 2009 test (see Appendix 2 for the Test Motivation Scale in the Norwegian PISA 2009 test). All students who participated in the PISA test in Norway in 2009 and 2012 reported their test motivation in the PISA test, based upon these questions.

Interviews of a sample of participating students in PISA 2006, 2009 and 2012

In order to further explore students' test motivation in the PISA test in a Norwegian context we conducted interview studies with students after they completed their PISA tests in 2006, 2009 and 2012. The first author interviewed 22 students from five schools in 2006, and nine students from two different schools were interviewed in 2009. (Author (s) 2009, 2010). The

study also includes interview results from nine students from one school that participated in the PISA 2012 study. Needless to say, the interviews cannot be generalised to the whole student population which participated in PISA 2006, 2009 and 2012 in Norway; instead their voices provide some perspectives to inform future research studies in students' test motivation.

The interviews were semi-structured and the students were asked how the school had prepared them for the test, about their understanding of both the item booklets and the background questionnaire, and whether they tried to do their best. The interview guide is provided in Appendix 3, and in the results section below some student responses are presented and discussed. Parental consent was given for all participating students, and interviews were recorded and transcribed by the first author and a research assistant. Interviews lasted approximately 10 to 15 minutes in 2009 and 2012, while the interview in 2006 lasted almost an hour, since more topics were covered. In order to make it as easy as possible for students to report how they had experienced the test, all interviews took place in the students' schools on the same day as, or on the day after, students had taken their PISA test. The interviewer talked to each student about each item they had completed, and asked questions about their answers using the cognitive interview techniques (Karabenick et al. 2007) to understand more about how students had interpreted the items, and to fully understand their responses.

Results from the survey data on students' test motivation

To address our main research question of whether students in Norway are motivated to do the PISA test, we first present the results of students reporting on the items about test motivation in the questionnaires in PISA 2009, and then the results of the interviews in 2006, 2009 and

2012. Results from the test motivation items in 2009 and 2012 are also presented. In 2006, only the interview study was conducted, and survey data is therefore not included regarding test motivation for these students. Table 1 reports students who answered positive to each statement. We have put the *strongly agree* or *agree* together to show the percentages of students who agreed to the following statements about test motivation.

Please insert table 1 here:

We present the items in descending order from those which received most agreement to those which received the least, and interestingly, the order of the items is the same both in 2009 and 2012. A total of 86% (2009) and 81% (2012) of students reported that they agreed with the statement, *I put in a good effort throughout the PISA test*, while the statement *Doing well in PISA meant a lot to me* which measure the importance of the PISA tests, have the lowest scores in both cycles, with 43% in 2009 and 36% in 2012. In other words, despite reporting that the PISA is not important for them as students, the majority of the 15-year-old students reported to have been motivated to do their very best on the test. A relatively high proportion of the students also reported that they were able to continue, even when some of the tasks were difficult. What is most striking is that there was a statistically significant decline in students' agreement on all the items between 2009 and 2012, indicating that students reported less motivation for the PISA test in 2012 than in 2009. Unfortunately we were not able to explain the reason for this; we did not have the survey results when we interviewed the students. As explained in the discussion section below, we did not obtain any information in the interviews that could shed light on this finding, but there is clearly a need to follow up on students' test motivation in the years to come.

Results in gender gaps in the PISA survey from 2009 and 2012

To investigate possible gender differences, we used Rasch model scaling of the test motivation scale. with a mean of 0, and a standard deviation of 1. In 2009, girls had a test motivation of .13, while the boys' was at -.13. In other words, boys reported having less test motivation than the average (which was set to 0) and the difference between girls and boys was .26 sd in the girls' favour. The difference is as large as a quarter of a standard deviation. The analyses showed that girls were overall more positive about the test, reporting higher test motivation than the boys, with a difference of a quarter of standard deviation, which, it must be said, is substantial.

To further explore these differences, we looked at each item and the percentages of girls and boys who agreed or strongly agreed with each statement. The results are presented in Table 2: the statements are listed in descending order of the size of the gap between boys and girls. The difference is greatest on the statement, *Doing well in the PISA test was important to me*. A far greater proportion of girls than boys think it is important for them to do well on the test. It is also worth noticing that both girls and boys were motivated to do their best, and this meant that they did a good job, although the girls reported that they were more motivated than the boys. However, the smallest difference was between girls and boys on the statement, *While taking the PISA test, I was able to persist even if some of the tasks were difficult*, indicating that both girls and boys struggled to keep up their motivation to continue working on tasks when faced with challenges.

Please insert table 2 here

The construct measuring importance of the PISA test, *Doing well in PISA meant a lot to me* and *Doing well in the PISA test was important*, has the lowest scores from the boys in both 2009 and 2012. In other words, 60% of the boys did not report that doing well in PISA meant a lot to them, yet as many as 84% of the boys reported that they put in a good effort throughout the PISA test in 2009, and 78% in 2012. This item is also the one that girls agreed most with both in 2009 and 2012. In other words, the vast majority of the Norwegian students reported that they worked hard in the PISA test, despite most of them reporting that it was not an important test for them personally. In other words, there is no empirical support for claiming that Norwegian students lack test motivation when participating in the PISA test. We recommend, however, further investigation into the gender differences in reported test motivation, as well as into the reported decline for both genders in their test motivation in 2012.

Based upon earlier research, we would expect to find a relationship between students' achievements and test motivation. Usually, high-achieving students are more motivated than low-achieving students to do well, and we would expect low-achieving students to report that they did not do their best (Weiner, 2007; Boekaert 2010). Table 3 shows the analysis of the correlations between test motivation, reading engagement and achievement scores on the PISA test in 2009.

Please insert table 3

The results in table 3 show that the correlation between test motivation and the four variables varies between $r = 0.2$ and $r = 0.26$. This indicates a correlation, such as in science, of $r = 0.2$, meaning that 4% of the variance can be explained by students' test motivation.

Another analysis of test motivation can be found in Figure 1, where the covariance between the reading scores and test motivation is illustrated. The Figure shows a weak relationship between these two variables, and it is further noticeable that students who have scored high, as well as those who have scored low on the PISA test, have all reported low test motivation. Similarly, both high- and low-performing students have reported high test motivation. In other words, it is not possible to find a pattern where certain student groups indicate higher test motivation than other groups, based upon their scores on the PISA test.

Please insert figure 1.

PISA 2003 and 2006 contained an 'effort thermometer' that measured students' perceptions of their own effort and motivation during the test (Appendix 1). The effort thermometer correlated with the $r = .2$ PISA scores and the students' self-reported grades in science. This result could indicate that test motivation in PISA also reflects the effort students put into ordinary schoolwork. Similar results were found in the study conducted by Eklöf on Swedish students, with correlations of 0.3 on similar questions (Eklöf 2006).

Results from the interviews in 2006, 2009 and 2012

The interviews with students participating in the PISA test were conducted by the first author in each cycle. The interviews were recorded and transcribed by the first author, while in

2012, a research assistant transcribed the interviews. In 2006, the interviews were part of a larger study (Author 2009), and lasted 30 minutes, while the interviews in 2009 and 2012 were solely about test motivation and students' experiences as test takers, and lasted between 9 and 12 minutes. An interview guide was developed by the first author in 2006, and slightly adjusted for the 2009 and 2012 cycles. In the final two cycles, students were asked to explain how they understood the test motivation questions and why they had responded in the way they had. In 2006, only general questions were asked, since the test motivation items did not form part of the national options that year. Table 4 shows the number of participating schools and students in both the interviews and the survey. Schools were selected from the sample which had already agreed to have test administrators present to check that the PISA test was carried out according to the international procedures. The selected schools were from three different counties in Norway in both rural areas and cities.

Please insert table 4

Results from the interviews 2006

In the Norwegian PISA test in 2006, 22 students were interviewed about their experience of the test directly after they had participated in it. The majority of the students (15 out of 22) said that the PISA test was "OK" or "interesting", while five of the students thought the test was boring. One girl said: "I think it was rather boring actually, I had expected much more difficult questions, and I thought it would be more difficult, such as algebra and so on...". Only one of the students claimed he did not do his best: "It was ok, but it is not the same when we are not given any grade, and some of the questions were strange" (Author 2009). Overall, the majority of these students talked more positively about the test and the tasks in the test, than critiques of the PISA test had suggested they would do. On the contrary,

students were overall positive about the PISA test itself, but more critical about the student questionnaire, which, some of the students found, asked too personal questions about their family background, such as parents' education, current job and socio-economic status. As one girl explained: "...not all students are proud of their parents". The questionnaire is always given to the students after they have participated in the PISA test, so there is no reason to believe that these questions would have influenced students' attitudes to the test, or their motivation to do well.

Results from the interviews 2009

From the sample of 4700 students from 197 schools, nine students (six girls and three boys) from two schools were interviewed about their experience of taking the PISA test. The interviews were conducted in their school, and recorded and transcribed by the researcher. The interviews in 2009 were approximately 12 minutes long. The interview guide built upon similar questions to those from 2006, asking students (i) how the school prepared them for the PISA test, (ii) how they experienced taking the PISA test, (iii) how they experienced answering the student questionnaire booklet, and finally (iv) to what extent they were motivated to do the PISA test. In 2006, 2009 and 2012, the interviewer used PISA tests to remind the students about what they had done, showing some of the tasks. In addition, students were asked about the test motivation items in 2009, to further investigate how students had interpreted the different items. All students explained that they had been given information about the PISA test through an information letter, and they were positive or neutral about this question. Eight students explained their experience of the PISA test in positive wordings, such as "fun", "alright", and one girl said: "It was ok, it was just like some other test". When asked about the tasks, one of the girls said she did not want to take the PISA test, but when she started, she thought it was okay. One of the boys said he thought the

tasks were “stupid” because he did not like the texts, and did not think the questions matched the texts. In addition, he experienced the test differently to what he was used to, and therefore he did not like it. Just as in 2006, when 15 of the 22 students were positive about the PISA test, most of the students used words such as “Alright” and “ok” when describing their experiences. Questions about the questionnaire reveal different patterns. Just as in 2006, students were more critical of the questionnaire and the questions than of the PISA test itself, and particularly of the questions about their parents.

Results from the interviews 2012

The interviews in 2012 were conducted in one school only, and with nine students, the total number of students participating in the PISA test in Norway being 4700. Analysis of the interviews did not result in any new information building on the previous years’, students used similar words describing their attitudes to it, such as ‘exciting’, since it was an international test, but the comments were less strong emotionally than in 2006 and 2009. Both in 2006 and 2009, some students argued that they needed to perform well on the PISA test, since they did not want Norway to look stupid, while in 2012 some students argued for the importance of doing well in general when you are asked to participate.

One girl explained why she had ticked that she had done her best in the PISA test:

‘If you have been chosen to do a test, you should do your best, since there is no reason not to. If there is anything too challenging, you can skip the item, but when you are there (taking the test) you should just go ahead and try to do your best’.

She further explained that since her school participated in the PISA test in May, just before the final exams in schools in Norway, she thought it was helpful for practising for her exams:

‘I find the test valuable for me even though I know that it does not count towards my marks’

One boy suggested that he would have taken more time on the test if he knew he would be given a grade, but he still decided to do as best as he could. He gave the following explanation:

‘It is fun do to well, no matter what!’

When the researcher prompted a question why he would do his best, the boy answered:

‘Why not?’

When the researcher asked another girl why she had responded she was motivated during the test, she answered she just thought about PISA as a normal school test, and she was usually motivated.

The general message from 2012 was again that the majority of the students had done their best, and their arguments were related to the importance of doing your best if you are asked to participate in a research study. Analysis of the interviews showed that 8 out of 9 students reported that they were motivated to do their best, even if the test did not influence their future grades. When prompted for an explanation, one of the boys explained:

‘I did my best, even though the test was not important. If I had had the (PISA) test back marked with a lot of mistakes, I would not worry, since it was not important to me personally. But if it was a mathematics test which mattered to me, I would be sad if I did badly’.

When asked if the teachers had prepared him for the test, he said that they had explained it was an international test, and this had made him feel more excited about the whole test. One girl explained that the responses given in the questionnaire should be as accurate as possible, and that is why she wanted to do her best for the PISA project. Four of the students explained they were motivated to do well since it was an international test and therefore it was more exciting. One girl also claimed the whole class had concentrated well because students

thought the international perspective were particularly motivating. She commented that some of the students gave up after a while, because it was too much for them, but she generally thought they worked hard. Another girl explained that it was fun to do her best, it could be helpful before an exam, and the idea of participating in an international test was interesting. Some of the girls explained that they thought it was okay to do the PISA test because they were not afraid that bad results would have an impact on their grades.

When asked about how the teacher had motivated them, the students explained that the teacher had a Norwegian flag in the classroom and after the test students were rewarded with pizza and soft drinks.

One boy explained that he thought the PISA test was strange, and when asked why, he commented upon the questionnaire and wondered why he had been asked how many books his family had at home. The test in itself was okay, but he reacted to personal questions about his family. Similarly, a girl reported that she did not quite know what to answer about her parents' education, since she did not know about it. As in 2006 and 2009, students who commented upon anything they did not like about the PISA test mentioned the background questionnaire for students. Overall, there was no indication in the interviews that students lacked motivation to do their best.

Discussion

The construct measuring importance of the PISA test, *Doing well in PISA meant a lot to me* and *Doing well in the PISA test was important*, has the lowest scores from the boys. More than half of the boys (53% in Table 1) did not report that doing well in PISA meant a lot to them, yet as many as 84% of the boys reported that they put in a good effort throughout the

PISA test. The findings suggest that, even though it is often claimed that young people in Norway are spoilt and do not care about school, such media attacks do not have any empirical support in the present study. On the contrary, the majority of boys in Norway can be said to have the ability to work hard, despite being given a test that has no value for them personally. On the other hand, there is a clear gender difference and girls report higher values on all items, with the largest difference found on the first item, *Doing well in PISA meant a lot to me*. Stoet & Geary (2014) demonstrated in their study of three PISA cycles that for overall achievement across reading, mathematics, and science literacy girls outperformed boys in 70% of participating countries. As we have pointed out previously, there tends to be a stronger relationship between high motivation and performance for males than for girls (Eklöf, 2007; Wise, Kingsbury, Thomason & Kong, 2004), test motivation could therefore be a variable which needs to be further looked into across the participating OECD countries.

Although a total of only 40 students were interviewed in the three PISA cycles from eight different schools, their comments are useful for informing future interview studies on test motivation. There are few interview studies around the world which have taken students' voices into account, and since students are the ones who take these tests, their experiences should be valued for improving the test situation, informing the policy level of the validity of the results and understanding more about test motivation.

Further, even though the sample is limited, it is worth noting that the majority of the students reported that they put in a good effort in the three PISA cycles. They also found the tasks to be appropriate, while in each cycle they were progressively more negative about the questionnaire study. Comments on the questionnaire are related to not understanding the question, not knowing their parents' education (so they cannot report it), or feeling

uncomfortable reporting on personal matters for an international test. Students are rarely asked about how they experience assessment situations (Elwood, 2013). But we argue that taking students' experiences into account when setting out to improve large-scale assessment studies has the potential to increase both the validity and reliability of the studies. It is also important to investigate how test situations in countries develop over time, and studies such as PISA can offer opportunities to do so, by including test motivation items as national options.

Limitations to the study

The three interview studies were conducted on a limited number of students, and the results can only be interpreted as reflections from the students participating. Even though the researcher requested to interview a mixed-ability group of students, all students were selected by their class teacher and it is probable that less-motivated students were not selected for interviews. Further, when asked about their grades, the whole sample was biased towards high-performing students.

Although we have responses from a total of more than 13,000 students in the three years of the cycle, there are a number of limitations found in survey research. One aspect is the tendency for respondents to tick more positive responses because that is seen as more socially acceptable (Nunally and Bernstein 1994). Another limitation is whether students understand the items according to the test makers' intention (Author 2011) and bias occur in child reports such as the student questionnaire in PISA (Jerrim, J. & J. Micklewright, 2014). Engzell and Jonsson (2015) also reported from a study in England, Germany, the Netherlands and Sweden (Children of Immigrants Longitudinal Survey in Four European Countries) where children, at the age of 14, were able to identify parents occupation, but struggled to

report on parental education. The PISA students are one year older, but still, we cannot be sure whether they all are able to report on the self-reports, the way the test-makers intended them to do. For the present study, the interviews have so far shown that students who participated in these studies understood the items very well, and made explicit distinctions between the items measuring *effort* and those measuring *importance of the test*. It is yet to be seen whether different findings would be shown in a bigger sample involving a more varied student group.

Large-scale assessment studies such as PISA, TIMSS and PIRLS are used in an increasing number of countries around the world. Yet we still do not know much about how students experience these tests, or how motivated they are to participate in them. Recent discourses in the media in Scandinavian countries show that there is a lack of data gathered from students' experiences, and future research should encourage designs involving students' voices. Even though the interviews in this study included few students, the results give a clear message that students are far more motivated to work hard in low-stakes tests than what is currently reported in the media in Norway. Whether we would have found similar results in other countries, is yet uncertain, since we do not know enough about how test motivation differs across countries. Hence, we still do not know enough about the impact students' testmotivation has upon PISA scores in other countries. Considering the current attention PISA receives from policy makers around the world, it would be useful to investigate students' test-motivation across more countries, also outside Scandinavia.

References

- Ackerman, P. L. & R. Kanfer (2009) Test Length and Cognitive fatigue: An Empirical Examination of Effects on Performance and Test-Taker reactions. *Journal of Experimental Psychology: Applied*, Vol. 15, No. 2, 163 – 181.
- American Educational Research Association, American Psychological Association, National Council for Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: APA.
- Barry, C.L., Horst, S.J., Finney, S.J., Brown, A.R., & Kopp, J.P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low- stakes testing. *International Journal of Testing*, 10, 342–363.
- Boekaerts, M. (2010), Motivation and self-regulation: two close friends, in Timothy C. Urdan, Stuart A. Karabenick (ed.) *The Decade Ahead: Applications and Contexts of Motivation and Achievement (Advances in Motivation and Achievement, Volume 16 Part B)* Emerald Group Publishing Limited, pp.69 – 108
- Baumert, J. & A. Demmrich (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education* **XVI**(3): 441 - 462.
- Brunner, M., C. Artelt, et al. (2007). Coaching for the PISA test. *Learning and Instruction* **17**(2): 111 - 122.
- Butler, J. & R. Adams (2007). The Impact of Student Effort on the Outcomes of International Studies. *Journal of Applied Measurement* **8**(3): 279 - 304.
- Eklöf, H. (2006). Student motivation on low-stakes tests: An example from TIMSS 2003, Umeå University: 1 - 14.
- Eklöf, H. (2007). Self-Concept and Valuing of Mathematics in TIMSS 2003: Scale structure and relation to performance in a Swedish setting. *Journal of Educational Psychology* **51**(3): 297 - 313.

- Eklöf, H. (2010). Skill and will: test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice* **17**(4): 345 — 356.
- Eklöf, H., Pavesic, B. Japelij & Liv Sissel Gronmo (2014) A Cross-National Comparison of Reported Effort and Mathematics Performance i TIMSS Advanced, *Applied Measurement in Education*, 27:1,31-45.
- Elstad, E, Nortvedt, G A & Turmo, A (2009) The Norwegian Assessment System: An Accountability Perspective, *CADMO* 17 (1), 89–103
- Engzell, P. & J.O.Jonsson (2015) Estimating Social and Ethnic Inequality in School Surveys: Biases from Child Misreporting and Parent Nonresponse, *European Sociological Review*, 31 (3): 312 - 325.
- Ertl, H (2006) Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany, *Oxford Review of Education*, 32 (5), 619–634
- Grek, S. (2009) Governing by Numbers: the PISA 'effect' in Europe, *Journal of Education Policy*, 24 (1), 23-37.
- Haenisch, H. & Nyroos, M (1998) warum TIMSS nicht geeignet ist , etwas uber de Leistungsfarhigkeit deutscher Schulen auszuzagen.: GGG Aktuell (Gemeinnutzige Gesellschaft Gesamtschule e.V.)
- Heyneman, S. P. (1993) Quantity, Quality and Source. *Comparative Education Review*, Vol 37, No4, pp: 372 – 388.
- Holliday, W. G. & B. W. Holliday (2003). Why using international comparative math and science achievement data from TIMSS is not helpful. *The Educational Forum* **67**(Spring): 250 - 257.
- Jerrim, J. & Micklewright, J. (2014). Socioeconomic gradients in children's cognitive skills: are cross-country comparisons robust to who reports family background? *European Sociological Review*, 30, 766–781

- Karabenick, S.A., Wooleym M. E., Friedel, J. M., Ammon, B.V., Blazeovski, J. Bonney, C.R., et al (2007) Cognitive processing of self-report items in educational research: do they think what we mean? *Educational Psychologist*, 42 (3), 139 - 151.
- Kjaernsli, M. & Olsen, R.V. (2013) Fortsatt en vei aa gaa. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012, Oslo: Universitetsforlaget.
- Mausethagen, Sølvi (2013). Accountable for what and to whom? Changing representations and new legitimation discourses among teachers under increased external control.. *Journal of educational change*. 22 s. doi: [10.1007/s10833-013-9212-y](https://doi.org/10.1007/s10833-013-9212-y)
<http://hdl.handle.net/10642/1665>
- Nunnally, J.C. & Bernstein, I. H. (1994) *Psychometric Theory* (3rd edition) New York: McGraw Hill.
- OECD (2001) *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD Publications.
- OECD (2007) *PISA 2006: Science Competencies for Tomorrow's World*, Vol. 1, Paris: OECD Publications.
- OECD (2009) *PISA 2006 Technical Report*, Paris: OECD Publications.
- OECD (2013) *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science*, Volume I. Paris: OECD publishing.
- O'Neil, H. F., J. Abedi, et al. (2005). Monetary Incentives for Low-Stakes Tests. *Educational Assessment* **10**(3): 185 - 208.
- O'Neil, H. F, B. Sugrue, & E. L. Baker (1996) Effects of Motivational Interventions on the National Assessment of Educational Progress Mathematics Performance. *Educational Assessment*, Vol 2, No. 2, pp. 135 - 157.
- Sjøberg, S. (2007). PISA and Real Life Challenges: Mission Impossible? In S. Hopman (Ed) *PISA according to PISA. Does PISA Keep What it Promises?* Wien, LIT Verlag.
- Sjøberg, S. (2014). Pisa-syndromet. Hvordan norsk skolepolitikk blir styrt av OECD. *Nytt Norsk Tidsskrift* 2014 (1) s. 30-43

Stoet, G. & D. Geary (2014) Sex differences in academic achievement are not related to political, economic, or social equality, *Intelligence* Vol 48, pp. 137–151

Sundre, D. L. (2007). The Student Opinion Scale (SOS) A measure of examinee motivation
Harrisonburg, VA, The Center for Assessment and Research Studies: 6.

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

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.

Weiner, (2007) Examining Emotional Diversity in Classroom: An Attribution Theorist Considers the Moral Emotion. In P. Schultz, R. Pekun and G. Phye (Eds), *Emotion in Education*, (75 - 88) San Diego, CA: Academic Press.

Wise, S. L., Kinsbury, G. G., Thomason, J., & Kong, X. (2004) *An investigation of motivation filtering in a statewide achievement testing program*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego.

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

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

Author (2009, 2010, 2011)

Aftenposten: <http://www.aftenposten.no/meninger/debatt/PISA-er-ikke-fyrtarn-7489705.html>
<http://www.dagsavisen.no/innenriks/ber-foreldre-stille-tøffere-krav-1.2969>

<http://www.vg.no/nyheter/utenriks/skole-og-utdanning/matteforsker-elevene-maa-ikke-faa-skylden/a/10148378/>

<http://www.tu.no/karriere/2012/11/12/-norsk-skole-lar-seg-styre-for-mye-av-pisa-undersokelsen>