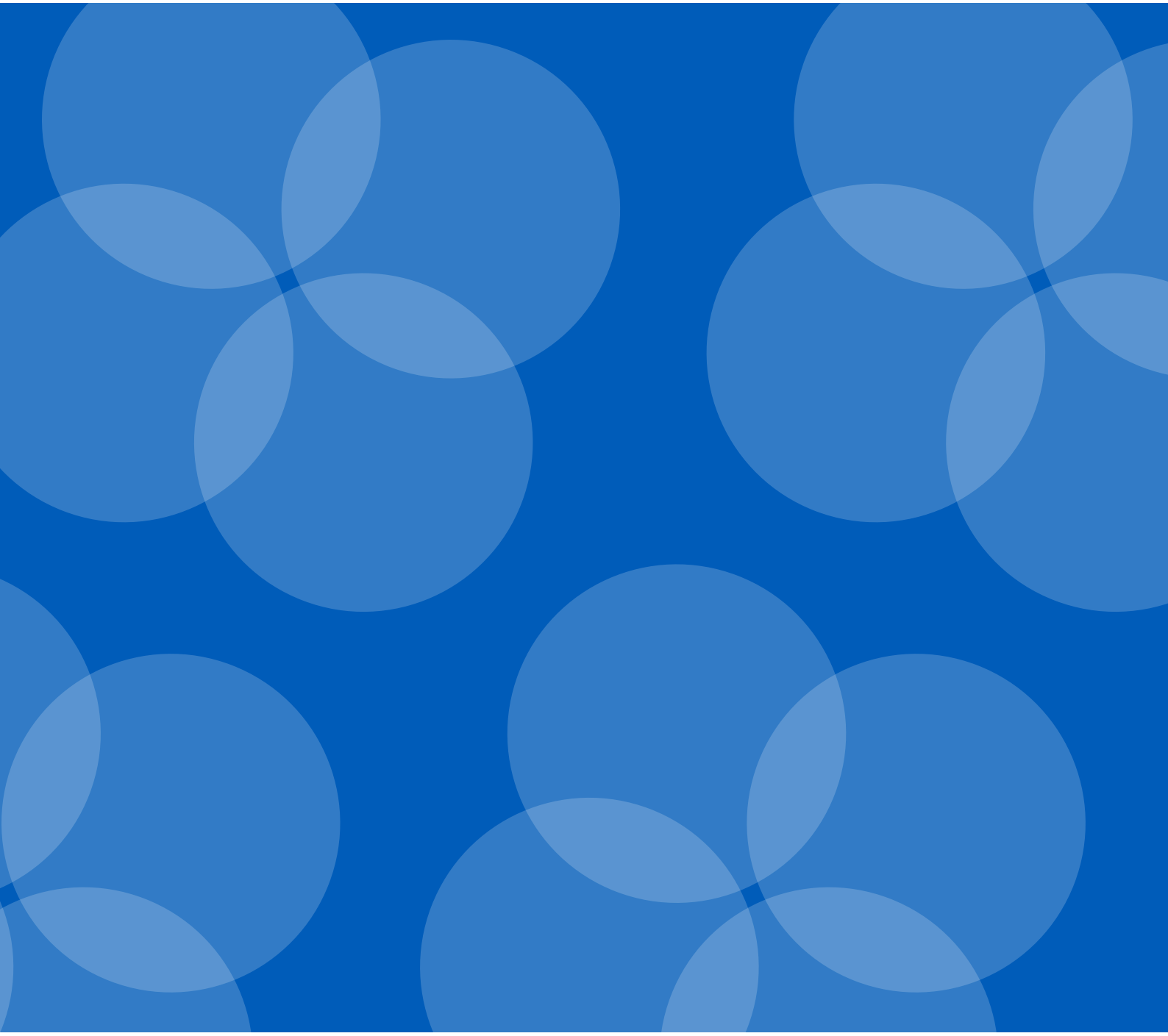
SKOLVERKETS AKTUELLA ANALYSER 2015

To respond or not to respond

The motivation of Swedish students in taking the PISA test



Skolverket
Swedish National Agency for Education

To respond or not to respond

The motivation of Swedish students in taking the PISA test

The publication is available for free download
at National Agencys for education's website
www.skolverket.se

ISSN: 1652-2508
ISBN: 978-91-7559-203-9
Graphic design: AB Typoform

Foreword

The PISA 2012 results demonstrated a sharp decline in all three PISA competencies: reading, mathematics and science. One of the potential partial explanations discussed was the context in which the PISA survey is conducted. The outcome of a PISA test does not have any impact on student grades and is therefore, from the student's perspective, a "low-stakes" test – a test without any direct consequences. It is reasonable to assume that students are more motivated and invest greater effort in tests on which their grades will be based. For this reason, the Swedish National Agency for Education made the assessment that the motivational aspects of international knowledge tests were something that needed to be studied in greater detail.

This report presents two studies on the theme of student motivation, effort and perseverance in taking the PISA knowledge test. Anita Wester of the Agency's Results Evaluation Unit has written the introductory text containing the Agency's assessment.

The first study, *Swedish students' reported motivation and effort in PISA, over time and in comparison with other countries*, was conducted by Hanna Eklöf of the Department of Applied Educational Science, Umeå University.

The second study, *Have Swedish students' perseverance and engagement changed over time? A study of student response patterns in PISA's knowledge test*, was conducted by Matilda Nilsson of the Agency's Results Evaluation Unit.

Stockholm, March 2015

Anna Ekström	*Anita Wester*
Director-General	Senior Advisor

Contents

Conclusions drawn by the National Agency for Education

Background

In 2012, Sweden participated for the fifth time in the OECD's international study PISA (Programme for International Student Assessment). Since the first survey took place in 2000, PISA has been conducted every three years – in 2003, 2006, 2009 and 2012. PISA looks at the knowledge of 15-year-olds in mathematics, reading and science, with each of these subjects taking turns to be the main subject assessed in each survey. Student knowledge is assessed by means of a test. Questionnaires are used to capture student attitudes towards schools and teaching.

In the first rounds of PISA, Swedish students showed very good results in all three competencies, with results that were above the OECD average. These results have since seen a gradual decline throughout the decade and in 2012 they were below the OECD average for the three competencies. Sweden is also the country with the worst results development of all participating countries. The Swedish report from PISA 2012 (2013) presents a number of hypotheses that might be partial explanations for the deterioration in results. One of these explanatory hypotheses is that students maybe do not invest the same level of effort in the PISA test as in previous PISA studies, i.e. the motivation to do their best in the test may thus have decreased over time.

In light of the deterioration in results – as noted not only in PISA, but also previously in PIRLS 2011 (National Agency for Education, 2012a), TIMSS 2011, year 8, (National Agency for Education, 2012b.) and also very recently in the in-depth analysis of PIAAC[1] (SNS, 2014) – the National Agency for Education has, amongst other things, taken the initiative for the two studies presented here. The studies complement each other in the sense that the analysis of Study I is based on student responses to questions about effort and motivation in taking the PISA test, whilst the analysis of Study II is based on the students' actual response patterns in the PISA tests.

Study I. Swedish students' reported motivation and effort in PISA, over time and in comparison with other countries

Aim

The aim of this study is to investigate the responses of Swedish students to questions about effort and motivation to do their best in the PISA test in 2012, and to study reported effort over time, in comparison with other countries and in relation to results in the PISA test. Effort and motivation in taking the PISA test have been assessed by means of two different measures: the "effort thermometer", which consists of a picture of two thermometers with a scale of 1–10, where students have been asked to indicate how much effort they invested in the PISA test and also how much effort they would have invested if

1 SNS (2014). *Lära för livet? Om skolans och arbetslivets avtryck i vuxnas färdigheter.*

the results of the test were going to affect their grades. The effort thermometer is a common measure for all nations participating in PISA 2003, 2006 and 2012. In contrast, the test-taking motivation scale is based on six questions about how motivated students felt in taking the PISA test; this was added as a national supplement to PISA 2012 in Sweden. Both the effort thermometer and the test-taking motivation scale are general and thus do not have an immediate focus on any particular subject. Both are based on the students' own assessment of their effort and motivation.

Results

Swedish students report a relatively low degree of effort in Pisa, internationally and in comparison with countries that are usually considered comparable with Sweden. Students also report a low degree of test-taking motivation in comparison with the same Swedish population of students responding to similar questions in connection with a national test. Furthermore, Swedish students report a relatively large difference between effort in the PISA test and how much effort they would have invested if the test was going to count towards their grade. The results also show that Swedish students' degree of reported effort is somewhat lower now than in previous PISA surveys.

There is a correlation between reported effort and performance in Sweden. This correlation is not negligible and it is stronger than in most other countries. However, we do see correlations of the same size in several of our neighbouring countries. The impact of effort on performance is significant even when we control for other variables. This applies both to Sweden and to most of the countries we compared ourselves with.

The fact that Swedish students report a relatively low degree of motivation and effort, also that reported level of effort has decreased somewhat, and that reported motivation and effort have an impact on performance, are in themselves interesting and important results. Thus, there are a number of indications that the students' test-taking motivation and effort in the test situation represent an important factor to take into account and to continue to monitor.

In summary, it may nevertheless be noted that the decline in reported effort is so small – and the impact of effort is not particularly strong – that it appears unlikely that this decline has contributed strongly to the weak performance of Swedish students in PISA 2012.

According to the statistical analysis, decreased effort could explain a minor part of the deterioration in results (e.g. 4–5 points[2] of 31 points in mathematics in PISA between 2003 and 2012), but the major part of the deterioration is probably explained by other factors, such as students not having sufficient skills in domains measured by the PISA test.

2 The mean is 500 and the standard deviation 100.

Study II. Have Swedish students' perseverance and engagement changed over time? A study of student response patterns in PISA's knowledge test

Aim

The aim of this study is to investigate whether parts of the decline in PISA 2012 results might be attributed to students in the latest survey not having had the perseverance and engagement to answer the test items to the same extent as previously and, by extension, whether the PISA 2012 results can be considered reliable. The main focus is on the mathematics domain, but the analysis also includes reading items.

We examine changes in student response patterns in the test items when they come first in the test booklet (position 1) and when they come last in the test booklet (position 4) in two ways: (1) for each position between two PISA surveys and (2) for the difference between these two positions within one survey (student perseverance). The Swedish students' response patterns are interpreted with reference to the average OECD pattern in order to assess whether the Swedish students' pattern appears to deviate from that of OECD students. However, this analysis makes no comparisons with individual countries. Two hypotheses have been formulated: (1) The decline in PISA results is due to a decreasing level of knowledge in Swedish students. (2) The decline in PISA results is due to decreased student engagement in taking the knowledge test. In this analysis, changes in student response patterns cannot be directly linked to results in the PISA test. It is therefore not possible to give an answer as to how much of the decline in results, expressed in terms of the number or percentage of points, can be explained by a lack of engagement on the part of students.

Results

Student perseverance in mathematics items appears to be lowest in PISA 2003 and highest in PISA 2006, whilst perseverance in PISA 2012 falls somewhere in between. Student perseverance in reading items is generally unchanged in Sweden. The exception is that perseverance, as measured by the share of students answering reading items correctly, decreases between PISA 2003 and PISA 2009.

On the whole, the Swedish students' response patterns suggest that it is primarily the level of knowledge that has decreased prior tothe latest PISA test in 2012. Support for this hypothesis is seen in many of the comparisons made, particularly when we consider the change in the share of students answering the items correctly in mathematics, but also in reading. Between most of the PISA surveys compared, there is a clear, equally large decline in the share of students answering items correctly, regardless of whether the items appear early or late in the test booklet.

There are also changes in the share of students leaving items unanswered that support the hypothesis of a decreased level of knowledge among students, in domains measured by the PISA test. These changes, however, are usually somewhat smaller and somewhat more difficult to assess. Between PISA 2003 and PISA 2009, and to some extent also between PISA 2009 and PISA 2012, there has been a steady rise in the share of students leaving the reading items unanswered. The analysis also indicates an increase in the share of students

skipping mathematics items between PISA 2006 and 2012. Although this increase between these years, when the items come early in the test booklet is not statistically significant, perseverance is unchanged between the surveys. This provides additional, albeit somewhat weaker, support for the hypothesis presented above.

There are also some signs to suggest that test-taking engagement among Swedish students may have decreased somewhat. It is mainly in the comparison of the share of students answering reading items correctly between PISA 2003 and 2009 that we see a response pattern that lends support to the hypothesis of decreasing student engagement. Between these years, the share of students answering items correctly is unchanged when the reading items appear early in the test booklet, but decreases when the items appear late in the booklet, which therefore means that perseverance decreased from PISA 2003 to 2009. However, it should be noted that a similar, though somewhat smaller, change also emerges in the OECD on average.

Although it is not possible to give an answer expressed in terms of the number or percentage of points of the decline in results the overall assessment is that this potential decrease in engagement would not be able to explain any major part of the decline in results shown by Swedish students over the past decade.

The overall picture

The two studies complement each other well, mostly because Study I is based on the students' self-reported motivation and effort, while Study II captures the students' actual behaviour. Even though the results from both analyses should be interpreted with caution, they reinforce each other. Neither analysis supports the idea that it is a lack of motivation and/or effort when taking the test which could explain the majority of the sharp decline in PISA results. Study II shows that no unequivocal pattern of decreasing perseverance over time can be established.

Study I shows that the difference in motivation between taking a high-stakes test and a low-stakes test is great among Swedish students, and international comparison shows that it is among Swedish students that the greatest difference is found. Thus, the motivation of Swedish students is not at its peak when taking the PISA test, but neither is there a maximum of motivation among the students in other participating countries, even though it is somewhat higher than that of Swedish students. The fact that motivation is not at its maximum in any country is a result which seems fairly reasonable, given that students do not benefit from any tangible "win" as a result of their efforts in the PISA test. Neither students nor their teachers are allowed to see the results of individual students.

Taken together, the results of the two studies show that it is hardly changes in motivation, engagement or level of effort that are decisive to the deterioration in results, but that the major part of the deterioration is probably explained by other factors, such as students not having sufficient skills in domains measured by the PISA test.

However, it is important to bear in mind that there are no simple correlations or simple explanations for either the declining results of Swedish students or for how effort and performance interact. In-depth studies from varying perspectives

can help to put pieces of the puzzle into place and so gradually provide a greater understanding of students' knowledge development and of the interaction between socio-emotional and cognitive variables. However, as complex phenomena are involved, the simple answers are only rough descriptions of reality.

Although these analyses have had the ambition of deepening our knowledge of Swedish students' response behaviour and reported motivation and effort in PISA, over time and in comparison with other countries, there are still many aspects that remain to be studied. Two such aspects are differences between boys and girls and what might be the reasons for these differences, as well as more in-depth studies of the students who give no response at all to the questions about effort and motivation. Another aspect is the possible difference between schools, with regard to reported effort. Since PISA is a test that is of no significance to students, it is conceivable that the way in which their school communicates to them the weight of performing well in PISA might be important to their motivation and effort.

Studies of students' response behaviour and reported effort and motivation could also be supplemented by studies on a smaller scale. These might include observations or interviews with students participating in studies similar to PISA in order to gain a better understanding of how the tests are perceived by students, and also of what they actually are responding to when answering questions about the effort thermometer and other questions related to motivation, engagement and perseverance.

STUDY I

Swedish students' reported motivation and effort in PISA, over time and in comparison with other countries

Summary

The aim of the present study has been to look at the degree of motivation and effort in PISA reported by Swedish students, over time and in comparison with a sample of other countries. Our analysis is also intended to contribute knowledge about whether test-taking motivation appears to be an important aspect to take into account, and about whether Swedish students' declining performance in PISA can be explained by their investing less effort, rather than other factors such as a decreased level of knowledge.

In summary, the results from this study indicate that test-taking motivation appears to be an important factor to monitor in the context of PISA but that, at the same time, the declining results in PISA cannot to any greater extent be explained by students investing less effort.

The following results form the basis for the conclusion that test-taking motivation can be an important factor to take into account in the context of PISA:

- Swedish students report a relatively low degree of effort in Pisa, internationally and in comparison with a selected sample of countries. Students also report a low degree of test-taking motivation in comparison with the same population of students taking a national test.
- Swedish students report a relatively large difference between effort in the PISA test and how much effort they would have invested if the test was going to count towards their grade.
- The Swedish students' degree of reported effort is somewhat lower now than in previous PISA surveys.
- There is a correlation between reported effort and performance in Sweden. This correlation is not negligible and is stronger than in most other countries. However, correlations of the same size are seen in several of our neighbouring countries.
- There is a significant effect of reported effort on performance even when we control for other variables. This applies both to Sweden and to most of the countries we compared ourselves with.

Although none of the above points alone signal a crisis with regard to the motivation and effort of Swedish students, together they indicate that test-taking motivation is a factor that should not be ignored.

However, our analysis finds little support for a lack of motivation constituting a decisive variable in the sense that less effort could explain the decline in Swedish results:

- The decline in reported effort is so small – and the effect of reported effort on performance not so strong – that it appears unlikely that this is what explains the sharp deterioration in performance by Swedish students in PISA 2012.
- According to the statistical analysis, the decline in effort could explain a minor part (4–5 points) of the deterioration in results, but the majority is probably explained by other factors, such as students not having sufficient skills in domains measured by the PISA test.

Introduction

The impact and scope of international comparative studies of student knowledge has grown in recent decades, and these studies have sometimes had a significant influence on education systems around the world. The study that has probably had the biggest impact in recent times is PISA. The present study focuses on Swedish students' reported motivation and effort in PISA. In order to provide a background to the analysis, we will first present definitions and assumptions regarding motivation in test situations, how this issue relates to PISA, and why it might be relevant to study this from a Swedish perspective.

The concept of test-taking motivation

Motivation is an important aspect of an individual's development, learning and performance in various contexts. Motivated individuals tend to be more persistent, goal oriented, active and higher performing. (see Pintrich & Shunk, 2002; Wigfield & Eccles, 2000). Motivation can be defined in different ways and measured at different levels. *Test-taking motivation* is measured at a situation-specific level. This type of motivation is a state associated with a given test, and can be defined as "the willingness to work on test tasks and to invest effort and persistence in this work".

There are a few assumptions to indicate that test-taking motivation might be an important factor in some test contexts. One such assumption is that it requires a certain amount of motivation for students to demonstrate their knowledge well. Of course, good test performance primarily requires adequate knowledge. If students lack the knowledge that the test is measuring, high motivation will not help. However, in addition to adequate knowledge, there is also a need for a sufficient degree of motivation to make the effort to demonstrate this knowledge. Another assumption is that the problem of low motivation is greatest in test situations where the outcome is of no consequence for the person taking the test (a low-stakes test). A high degree of motivation often ensues from assigning some kind of value to the task to be performed, so that the task or consequence of the task means something to the individual, directly or indirectly. It can then also be assumed that, when the test result has no consequences, some students will place a low value on the task and thereby not be motivated to do their best. The result might be that students perform below their actual ability. Not taking test-taking motivation into account can thus lead to an incorrect interpretation of the results. In other words, that which is taken to be a pure measure of knowledge may also be a measure of something else (such as motivation).

Might test-taking motivation be an important factor in PISA?

At the political level, studies such as PISA can be viewed as high-stakes. The results are often used by politicians and decision-makers as indicators of the quality of education and they have sometimes been causes contributing to the reform of education systems. Because there is often great media and political interest in the results of international comparative studies, the reporting of

these studies also contributes towards shaping the public image of the state of education in a given country. However, from the perspective of students, PISA is an example of a survey that offers them no incentive to do their best, and it may therefore be important to factor in students' test-taking motivation when interpreting the results. Students must devote hours to a test that for some might be quite demanding, while the result of this investment has no practical significance for them personally. The result does not count towards their grade, and they receive no individual feedback. The results are reported only at the country level long after the study has been conducted. Good performance thus leads to no form of appreciation or reward, and poor performance brings no negative consequences. This could lead to students placing a low value on good PISA performance, investing less effort and performing at a lower level than they would have done if the test results had consequences.

On the other hand, the fact that students are representing their country in an international study could be sufficient to create interest and motivation. The fact that the test does not have any consequences might also make the test situation less stressful, which can be positive for performance, even where motivation is not maximum. We should also not underestimate students' interest in doing well simply for their own sake, because they "always do their best", or in order to "help with the study". Some students may find the opportunity to test themselves interesting and meaningful, but without the grades stress surrounding many other tests.

All these arguments could be true, and there are probably students in all the above categories in a given test situation: those who find the test meaningless and those who think it feels exciting and interesting. There have been relatively few studies of test-taking motivation in the actual context of international studies such as TIMSS and PISA, but the studies nevertheless performed in these and other contexts have generally shown that students report lower motivation in low-stakes tests compared with high-stakes tests and that test-taking motivation tends to have some significance for performance (see for example Eklöf & Knekta, 2014, or Wise & DeMars, 2005 for an overview).

If students in some countries were to be highly motivated to do their best, while students in other countries lack motivation to make an effort, comparisons of results in terms of knowledge levels could be misleading. In discussions about PISA, it is sometimes argued that students in some countries place a greater value on good performance in PISA than students in other countries due to the country's test culture and culture in general, or due to how the test is portrayed to students. The question of test-taking motivation in PISA thus becomes a question of whether the test results are reliable: whether a country's results actually reflect the students' knowledge, and whether comparisons between countries are actually comparisons of knowledge levels and not of motivation levels.

Sweden in PISA

The results from the latest rounds of PISA have been disappointing for Sweden. Since the first PISA survey in 2000, there has been a continuous decline in Swedish results, and Sweden is now the country in the PISA study with the greatest deterioration of all (National Agency for Education, 2013). Could

Sweden's declining results be due to students having become less motivated to invest effort to do their best in PISA and to Swedish students being less motivated to do their best than students in other countries? The assessment climate in the Swedish education system has changed considerably over the past decade; Sweden today has more national tests in more subjects and more school years than we have ever had before, and student grades are allocated earlier etc. In the spring semester in year nine, students take national tests in four subjects. The national tests are used by teachers as a basis for grading and are perceived as important to students. It is conceivable that, with the greater range of national tests currently taken by Swedish year nine students, tests such as PISA are perceived by them as being less important for them to invest their time and energy in.

In light of this, the aim of the present study is to look more closely at the degree of motivation and effort in PISA reported by Swedish students, over time and in comparison with a sample of other countries.

Aim

The aim of this study is to investigate the reported level of test-taking effort and motivation among Swedish students participating in PISA, over time, in comparison with other countries and in relation to PISA results. Test-taking motivation has been assessed by means of two different measures: "the effort thermometer", which is a common measure for all nations participating in PISA since 2003, and a test-taking motivation scale, which was added as a national supplement to PISA 2012 in Sweden. Both of these measures are based on self-report, that students themselves indicate their position on a number of statements concerning motivation and effort.

The study builds on the questions below, and findings will also be presented in the following order:

1. Do Swedish students appear to be motivated to do their best in PISA 2012? Under this question, we will present descriptive results from the test-taking motivation scale and the effort thermometer.
2. Are there differences between countries with regard to self-reported effort on the effort thermometer in PISA 2012? Here, we look at the results from an international perspective and for a sample of countries.
3. Has level of reported effort changed over time, in Sweden and in comparison with a sample of other countries?
4. Is there any correlation between reported test-taking motivation and test performance in Sweden and other countries?
5. Might test-taking motivation explain the decline in Swedish results? Is there any effect of reported effort on performance when other attitude/motivation variables are taken into account?

Before presenting the results, we will give a brief description of the analyses and delimitations made, and of the instruments used.

Implementation

Method and instruments

Our analyses have primarily used data from PISA 2003 and 2012 but also, to some extent, data from PISA 2006. Apart from the national supplements, all data have been downloaded from the freely accessible http://www.oecd.org/pisa/pisaproducts/. Data from student questionnaires, the effort thermometer and the national supplements containing test-taking motivation questions have been used together with test results. Processing and analysis have been carried out in the programmes SPSS and IEA IDB Analyzer. Descriptive statistics, significance tests, correlation and regression analysis as well as analyses of the scales' homogeneity and dimensionality have been produced. Unless otherwise stated, all analyses have been performed with weighted values and with all five plausible values as dependent variables. (OECD 2013a describes why weighted values should be used and why PISA works with results data in the form of several "plausible values").

The focus will be on analyses of the effort thermometer, as this can be used to look at differences over time and in relation to other countries, while results from the test-taking motivation scale only apply to Sweden and only to 2012. Results are primarily presented for Sweden, but some of the analyses of the effort thermometer include a sample of other countries. This sample includes the Nordic countries and other Baltic countries (excluding Russia). In addition, the sample of countries has also included the Netherlands, Australia and New Zealand as these are countries that we sometimes compare ourselves with in these contexts.

Focus will also primarily be placed on mathematics, since mathematics was the main area in both 2003 and 2012, thereby enabling more reliable comparisons over time. The results are thus more stable when placed in relation to mathematics than to the other two PISA subject areas: reading and science. The motivation scales included in the PISA 2012 student questionnaire, and which are also used as variables in some of the analyses in the present study, also mainly concern student attitudes and motivation with respect to learning mathematics. Some of the analyses will however also include reading and science in order to investigate whether effort/motivation has a different impact on the results in different subjects.

The effort thermometer

The effort thermometer was used for the first time in PISA 2000 in three of the participating countries: Norway, Germany and Australia. In 2003, the thermometer was used for the first time in all countries, and it has since then also been used in 2006 and 2012. The effort thermometer comes last in the test booklet and contains two questions that students answer on a scale of one to ten (a thermometer), (see Figure 1). Figure 1 also shows a third scale (in the column on the left), where the number 10 is already marked in order to illustrate an imaginary situation (any at all) that is highly important to the student personally, and where they would therefore do their very best. For the scale in the middle column, the student is then asked to mark how much effort they invested

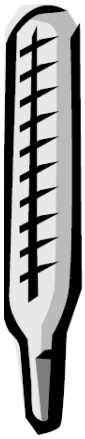
in PISA, *if the student compares their effort in PISA with this imaginary situation*. For the scale in the column on the right, the students indicate how much effort they would have invested in PISA if the test result was going to count towards their grade. It should be noted that the report on the thermometer concerns the PISA test as a whole, i.e. reading, mathematics and science.

Although the effort thermometer has been used in several PISA rounds, relatively few analyses have been performed with a particular focus on the students' self-reported effort, and thus we do not know so much about this. However, based on data from PISA 2000 and 2003 (reading), a thesis has been written containing in-depth analysis of student effort, with a particular focus on "relative effort" (the difference between how much effort they invested in PISA and how much effort they say they would have invested if the result was going to count towards their grade) (see Butler, 2008).

In her thesis, Butler draws the conclusion that a varying degree of motivation among the participating countries does not pose any major threat to the PISA results as the differences between countries were not as large as might have been feared. She also concludes that there was a correlation between effort and performance internationally, and that this correlation was not negligible but also not able to explain much of the differences between countries from an international perspective (Butler, 2008). Based on these analyses, the OECD also draws the conclusion that "Reassuringly, students' self-reports on this subject suggest that the effort they invest in PISA is fairly stable across countries" (OECD, 2007, p. 52). Since this is so far the only major study of the effort thermometer that appears to have been conducted, the results of the present study will be placed in relation to Butler's study where appropriate.

Figure 1. The effort thermometer.

Please try to imagine an actual situation (at school or in some other context) that is highly important to you personally, so that you would try your very best and put in as much effort as you could to do well.

	In this situation you would mark the highest value on the "effort thermometer" as shown below:	Compared to the situation you have just imagined, how much effort did you put into doing this PISA test?	How much effort would you have invested if your marks from the test were going to count towards your school marks?
	☒ 10	☐ 10	☐ 10
	☐ 9	☐ 9	☐ 9
	☐ 8	☐ 8	☐ 8
	☐ 7	☐ 7	☐ 7
	☐ 6	☐ 6	☐ 6
	☐ 5	☐ 5	☐ 5
	☐ 4	☐ 4	☐ 4
	☐ 3	☐ 3	☐ 3
	☐ 2	☐ 2	☐ 2
	☐ 1	☐ 1	☐ 1

The test-taking motivation scale

The test-taking motivation scale consists of six statements (see Table 1 below). Each statement was answered on a four-point scale (from "strongly agree" to "strongly disagree"). The test-taking motivation questions came at the end of the student questionnaire, which students answer after they have completed the test. A similar scale has been used for a number of previous studies in the Swedish context (TIMSS Advanced 2008, TIMSS 2003, National tests, Swedish Scholastic Aptitude Test). The scale contains both questions intended to measure how much value students place on the test (if good performance in the test feels important) and questions concerning how motivated students were to make an effort and how much effort they made. This particular scale has been used by Norway in previous PISA rounds, and has worked well there. Analysis of the scale's characteristics in Sweden's PISA 2012 also suggests that it constitutes a homogeneous and one-dimensional measure.

The effort thermometer and the test-taking motivation scale partially overlap (at least in theory) as they both concern effort. Both these measures are also based on self-report, and we thus only have access to the students' own statements about how much effort they say they invested or how motivated they were. We cannot know how much effort they actually invested, or how motivated they actually were, but have to trust that the students are giving a truthful picture of how they themselves felt and acted. This is a problem with which all self-report scales have to contend, but self-report is used in many studies and there is good support to indicate that results from self-report scales can be useful, though there should of course be an awareness of their limitations.

The test-taking motivation scale is more about how students felt about the test than the effort thermometer is, as it particularly asks about effort and not directly about motivation. As they are answered at different times, it is also conceivable that they are influenced by different errors. The effort thermometer is answered immediately after the test and at this point, there may be a greater risk that students not only respond about how much effort they invested, but also about how they thought it went in the test they had just taken, how difficult the test felt, what imaginary situation they have related their reported effort to (see Figure 1), etc. The test-taking motivation scale is answered at the end of the student questionnaire and at this point, there is a risk that students are tired of answering questions and therefore do not respond at all or do not give completely truthful responses. A student who is unmotivated to invest effort in the test may also be unmotivated to answer questions about their motivation. There is also a risk that they will respond about their entire motivation, both for the test and for the questionnaire. Still, the assumption is that both these measures can provide an insight into how the students perceived PISA, how motivated they were to do their best and how much effort they invested in the test.

Results

Do Swedish students appear to be motivated to do their best in PISA 2012?

The test-taking motivation scale

Table 1 presents the share of students who agreed or disagreed with each statement in the test-taking motivation scale, together with the mean value for each item. Note that in this table, the lower the mean value, the greater the agreement with the statement (where 1 = strongly agree, 4 = strongly disagree).

Table 1. Percentage of Swedish students in PISA 2012 giving each response in the test-taking motivation scale, with mean values at the item level.

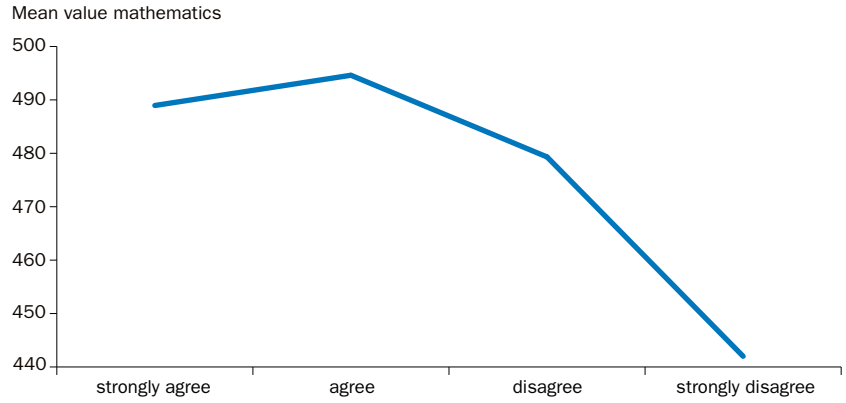
	1) Strongly agree (%)	2) Agree (%)	1) + 2) summarized (%)	3) disagree (%)	4) Strongly disagree (%)	3) + 4) total (%)	M	SD
1. I felt motivated to do my best on the PISA test	16.1	46.8	62.9	27.1	10.1	37.2	2.31	.86
2. I engaged in good effort throughout the PISA test	15.5	58.5	74.0	19.8	6.2	26.0	2.17	.76
3. Doing well on the PISA test was important to me	11.2	34.7	45.9	41.8	12.3	54.1	2.55	.85
4. I worked on the tasks in the test without giving up even if some tasks felt difficult	15.1	47.1	62.2	29.3	8.5	37.8	2.31	.83
5. Doing well on the PISA test meant a lot to me	8.6	28.0	36.6	46.7	16.6	63.3	2.71	.84
6. I did my best on the PISA test	22.0	44.3	66.3	24.5	9.3	33.8	2.21	.89

Combining the percentages for the two positive response options (strongly agree + agree) and the two negative response options (disagree + strongly disagree, the blue columns in Table 1), we see that a majority of students agreed that they felt motivated to do their best in the PISA test (Item 1), that they did their best in the test (Item 6) and worked on the tasks in the test even though they felt difficult (Item 4). Three quarters of students agree that they put in a good performance in the test (Item 2). Less than half the students agreed that it felt important to them to do well in PISA (Item 3), and just over one third of the students

agreed that it meant a lot to them to do well in PISA (Item 5). It should be noted that these percentages apply to the students who actually responded to the items. For each question, there were around 10 per cent missing values.

To exemplify the relationship between responses to the test-taking motivation items and test performance, Figure 2 shows mean results in mathematics per reported value for the first item in the scale, "I felt motivated to do my best on the PISA test".

Figure 2. Results in PISA's mathematics domain in relation to scale increments for the item "I felt motivated to do my best on the PISA test".



Items 2), 4) and 6) in Table 1 above show similar patterns to that in Figure 2. It thus appears that a higher degree of reported test-taking motivation is related to higher performance, with a small dip for those who "strongly agree" with the statement. Those who did not respond to the item at all performed the same as, or worse than, the group which strongly disagreed with the statement. For the two questions concerning whether the test was perceived as important, the pattern looks a little different: the students indicating that they "strongly agree" with the statements that it was important and that it meant a lot to them do their best in the test performed considerably worse than those students indicating that they "agree" with these statements. The fact that students report that they think the test is very important is thus in itself no guarantee for their performing well. On the other hand, previous studies have shown that if students report that the test is important, they are also more motivated to invest effort.

Compared with previous studies in the context of TIMSS, using similar if not identical questions, Swedish students in PISA 2012 report a higher degree of motivation than students in the third year of upper secondary school who participated in TIMSS Advanced 2008 (see Eklöf, Japelj Pavešić, & Grønmo, 2014). For example, in TIMSS Advanced 2008, less than 35 per cent of Swedish students indicated that they felt motivated to do their best or that they worked on all the tasks in the test without giving up. However, the group of students participating in TIMSS Advanced 2008 differs in many ways from the group of students in PISA, which makes it less appropriate to make direct comparisons. Students in TIMSS Advanced are three years older than PISA students and presumably much more tactical in their studies. Compared with students who participated in TIMSS 2003, PISA students appear to report a somewhat lower degree of motivation (Eklöf, 2006), but the design for measuring

test-taking motivation was somewhat different in TIMSS 2003. Compared with findings from a similar scale used in the same student cohort, but in the context of national tests, a smaller share of the students in PISA 2012 indicate that they felt motivated: In a study by Eklöf and Knekta (2014) that included a sample of year nine students taking the national test in science in spring 2012, between 80 and 90 per cent of students indicated that they felt motivated to do their best in the test, that they worked on all the items in the test without giving up and that it was important to them to achieve a good result in the test.

In summary, the Swedish students' self-reports on the test-taking motivation scale in PISA 2012 indicate that only a small percentage of students did not bother at all to do their best in the PISA test, but that a fair proportion still appears to have had a fairly lukewarm attitude to the test, and they report a lower degree of motivation than year nine students from the same cohort taking a national test.

Although comparisons with findings from TIMSS and the Swedish national test give some indication of whether students participating in PISA may be considered to report a high or a low degree of motivation, it is still difficult to draw any real conclusions as we currently do not have any other samples from PISA for particular comparison with the test-taking motivation scale. The effort thermometer used in PISA internationally might then give a better indication in that regard. For this reason, we first present Swedish results for the effort thermometer in PISA 2012, and in the next section we then make a comparison with other countries regarding reports on the effort thermometer.

The effort thermometer – descriptive results PISA 2012

The effort thermometer comes last in the PISA test booklet, and students respond to it immediately after the test, before the student questionnaire. The databases report two questions: how much effort the students invested in the PISA test compared with a maximum performance (see Figure 1, the middle column in the figure) and how much effort they would have invested if the test result was going to count towards their grade (the question to the right in Figure 1). From these two questions a difference variable is also calculated as a kind of measure of relative effort. The following primarily concerns one of these questions: how much effort students indicate they have invested in the PISA test, but the difference variable has also been analysed.

In PISA 2012, Sweden has a mean value of 7.03 on the ten-point scale regarding how much effort students invested in the PISA test. 8.4 per cent of the students marked a ten on the effort thermometer, i.e. maximum effort. The most common value indicated is an eight, with almost one quarter of Swedish students marking an eight on the effort thermometer. The second most common value is a seven, marked by about 20 per cent of students. The third most common value is a nine, marked by about 15 per cent of students. Furthermore, about 20 per cent of students mark an effort of five or less. Table 2 below presents the share of students marking each increment on the effort scale, where "1" stands for minimum effort and "10" stands for maximum effort. It also reports mean results in mathematics, reading and science for each increment on the effort scale. Example: 11.6% of students marked an effort of six on the effort thermometer. These students had an average result of 469 points in PISA's mathematics domain.

Table 2. Percentage of students indicating the values 1–10 on the effort thermometer (ET) in PISA 2012, with average test results for each increment.

Reported value on the ET	% of Swedish students	Results mathematics 2012	Results reading 2012	Results science 2012
1	2.0	387	359	382
2	2.0	431	420	432
3	3.1	446	437	446
4	4.7	443	444	450
5	8.4	448	450	452
6	11.6	469	478	478
7	20.4	490	500	500
8	23.6	502	512	511
9	15.7	511	525	521
10	8.4	498	507	505

The results look very similar for mathematics, reading and science, with a higher average test score for almost every increment from 1 to 9 on the effort scale. Those marking a ten on the effort scale, i.e. maximum effort, perform on average at a slightly lower level than those marking a nine or an eight. The group of students that marked a ten on the effort thermometer was a fairly heterogeneous group with regard to PISA performance. Some had very low test scores, while others had very high test scores. The category is likely to include both students who experienced the test as very difficult and who therefore really made an effort, but were unable to perform better, and students with high levels of knowledge and a high degree of effort, and therefore a high test score. But on average, this group has somewhat lower test scores than the students who marked an effort of nine or eight.

The table above presents the results of the students who actually responded to the effort scale. Just over seven per cent of Swedish students gave no response to the question of how much they effort they invested in the PISA test. The mean mathematics score for these students is 410 points, and similar mean scores were obtained for reading and science. Those not responding to the question of how much effort they invested thus perform quite poorly in PISA, on average somewhere between those marking an effort of one and those marking an effort of two. It is possible that this group – the non-respondents – represents students who were less motivated to take the test. Another explanation could be that this group consists of low-performing students who did not have time to answer all the questions in the test booklet and therefore did not reach the question about how much effort they invested in the test.

The fact that a clear majority of students mark a seven or higher on the effort scale cannot be said to indicate a general lack of self-reported effort among Swedish students. At the same time, there is nevertheless a proportion of students who report a low degree of effort, and those who report a low degree of effort have a poorer result in PISA. The effort scale and the test-taking motivation scale portray similar pictures: a majority of Swedish students report a reasonable (see Butler, 2008) degree of motivation and effort, but between one quarter and one third of the students indicate a fairly low degree of motiva-

tion and effort. As with the test-taking motivation scale, it is a little difficult to comment on whether the results on the effort scale represent high or low reports unless there is something to compare them with. One comparison that can be made is that concerning how the students responded to the second question in the effort thermometer: how much effort they would have invested in the PISA test if it was going to count towards their grade. Such a comparison might give an indication of how the students value the PISA test compared with how they value tests on which their grades will be based. For Sweden, the mean value for the question of how much effort the students would have invested in the PISA test if it was going to count towards their grade is 9.51 on the 10-point scale. Swedish students thus report that they would have invested a great deal of effort if the test was going to count towards their grade. The difference between reported effort in the PISA test and estimated effort if the test was going to count towards their grade is 2.47 increments on the scale. Thus, on average, Swedish students report that they would have invested "a couple more notches" of effort if the test was going to count towards their grade. Distributions of the difference variable for Swedish students, and the average mathematics performance for each value of the difference variable, are presented in Appendix 1, Figures D1 and D2.

Above, we have reported descriptive results for Swedish students' reported effort and motivation in PISA 2012. Comparisons with other countries that participated in PISA and comparisons over time can place the above results in a larger context. The following section will explore whether there are differences between countries with regard to reported effort in PISA.

Are there differences between countries regarding self-reported effort on the effort thermometer in PISA 2012?

Figure 3 presents mean values for reported effort for all countries that participated in PISA 2012. Figure 4 presents mean values at the country level for estimated effort if the PISA test was going to count towards a grade. Figure 5 presents the difference between these two reports.

As shown by these three figures, Sweden is one of the countries that reports the lowest degree of effort in PISA 2012, the highest degree of effort if the test was going to count towards a grade, and the largest difference between these two reports.

However, in most cases, the differences between countries when it comes to reported effort is quite small. Although the differences between Sweden and the other countries are statistically significant in many cases, they are quite small if we look at size differences using measures of effect size (standardised mean differences). The majority of countries have a mean value of reported effort in PISA 2012 that is between 7 and 8 on the ten-point scale (see Figure 3). The countries that report a comparatively very high degree of effort (higher than 8.50 on average) are all low-performing countries, as are several of the countries that report an average effort of more than 8.0. However, this latter category also includes, e.g. Shanghai and Taipei, which are, respectively, the best performer and among the best performers in PISA 2012. The country that reports the comparatively lowest level of effort, Japan, is one of the highest performing countries. Japanese students also come lowest with respect to reporting how

much effort they would have invested if the test was going to count towards their grade, so it is possible that Japanese students generally tend to give low reports on a scale such as the one we are studying.

With some exceptions, it may be possible to discern a pattern that might be culturally conditioned or related to type of education system/level, whereby students in some cultures treat response scales of this type in one way (giving lower, perhaps more realistic reports), while students in other cultures or students in developing countries treat scales of this type in a different way (giving higher, perhaps socially desirable reports). Previous research has shown that students from economically less developed countries tend to respond as they believe they should respond, rather than with what they actually think (King, Murray, Salomon, & Tandon, 2004). In other countries, it might be unacceptable for students to say that they did not bother to invest effort in a test. As pointed out by Butler (2008), it may also be true that in countries where literacy is weak, the meaning of the questions in the effort thermometer might be difficult to understand and thus difficult to answer. It is also assumed that students understand the meaning of a thermometer, and/or the principle of giving a report on a scale of 1 to 10. Although this might seem obvious to some, it need not be so for everyone. It could also be true that very high-performing students in some countries perceive the test as easy and achieve a good result without needing to invest particularly great effort, whilst low-performing students perceive the test as very difficult and therefore demanding great effort, without this resulting in a high test score. Having said this, it is of course still of interest to see how students in other countries responded to the effort thermometer, and this is described in the figures below.

Figure 3. Mean value of reported effort in PISA 2012 by country. Sweden, together with the other countries in the sample in this study, has dark blue bars.

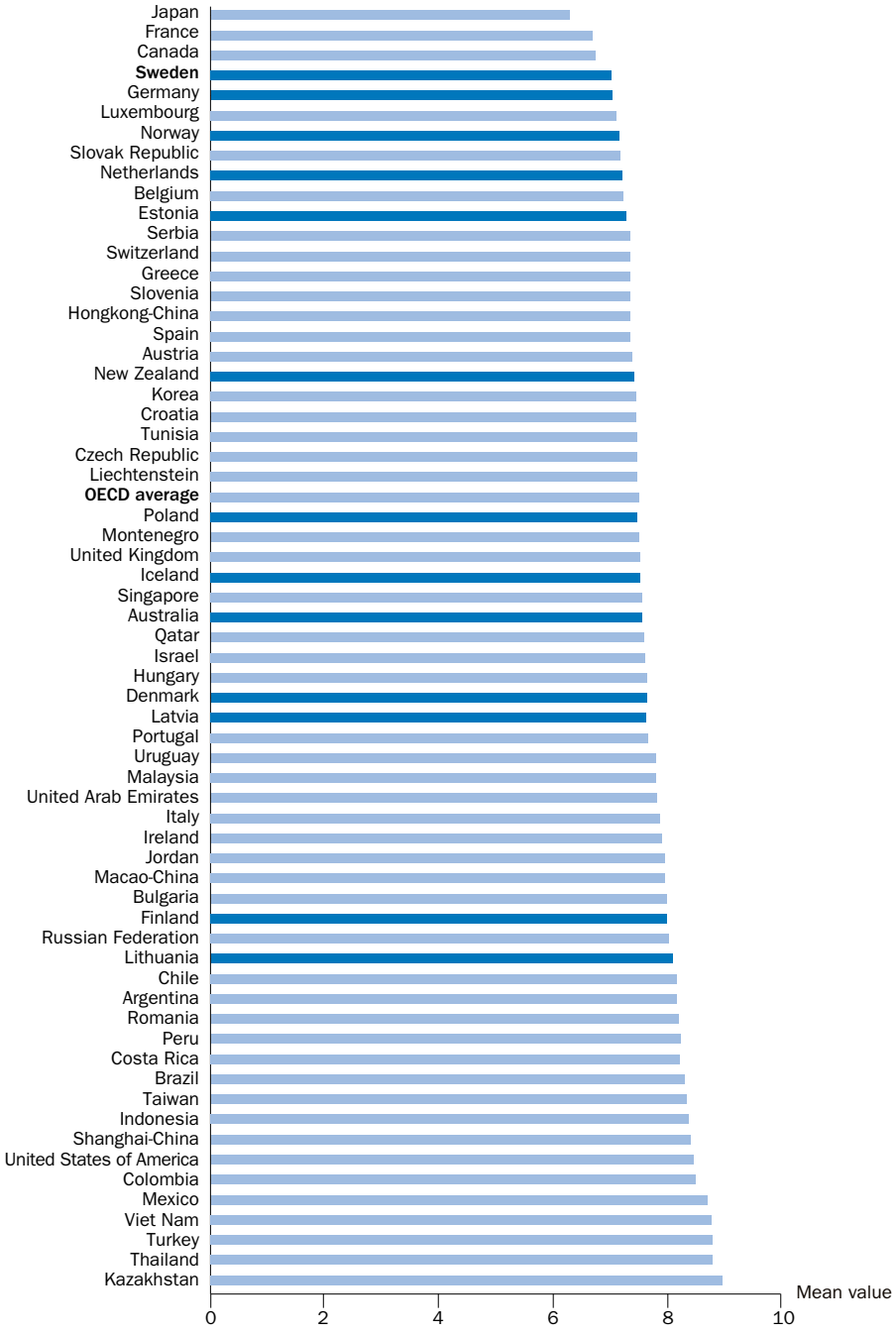


Figure 4. Mean value of estimated effort if the test result in PISA was going to count towards a grade, by country. Sweden, together with the other countries in the sample in this study, has dark blue bars.

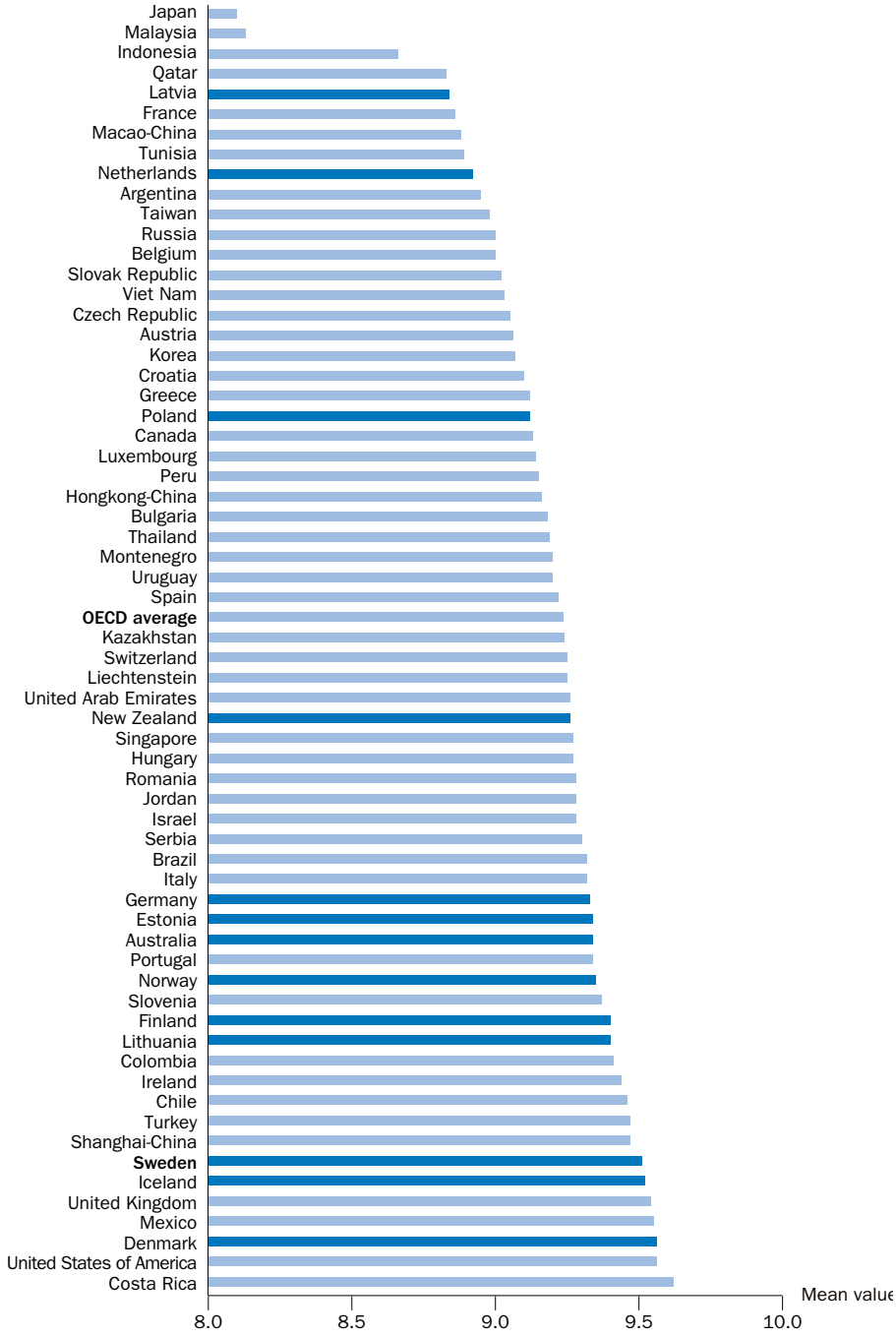
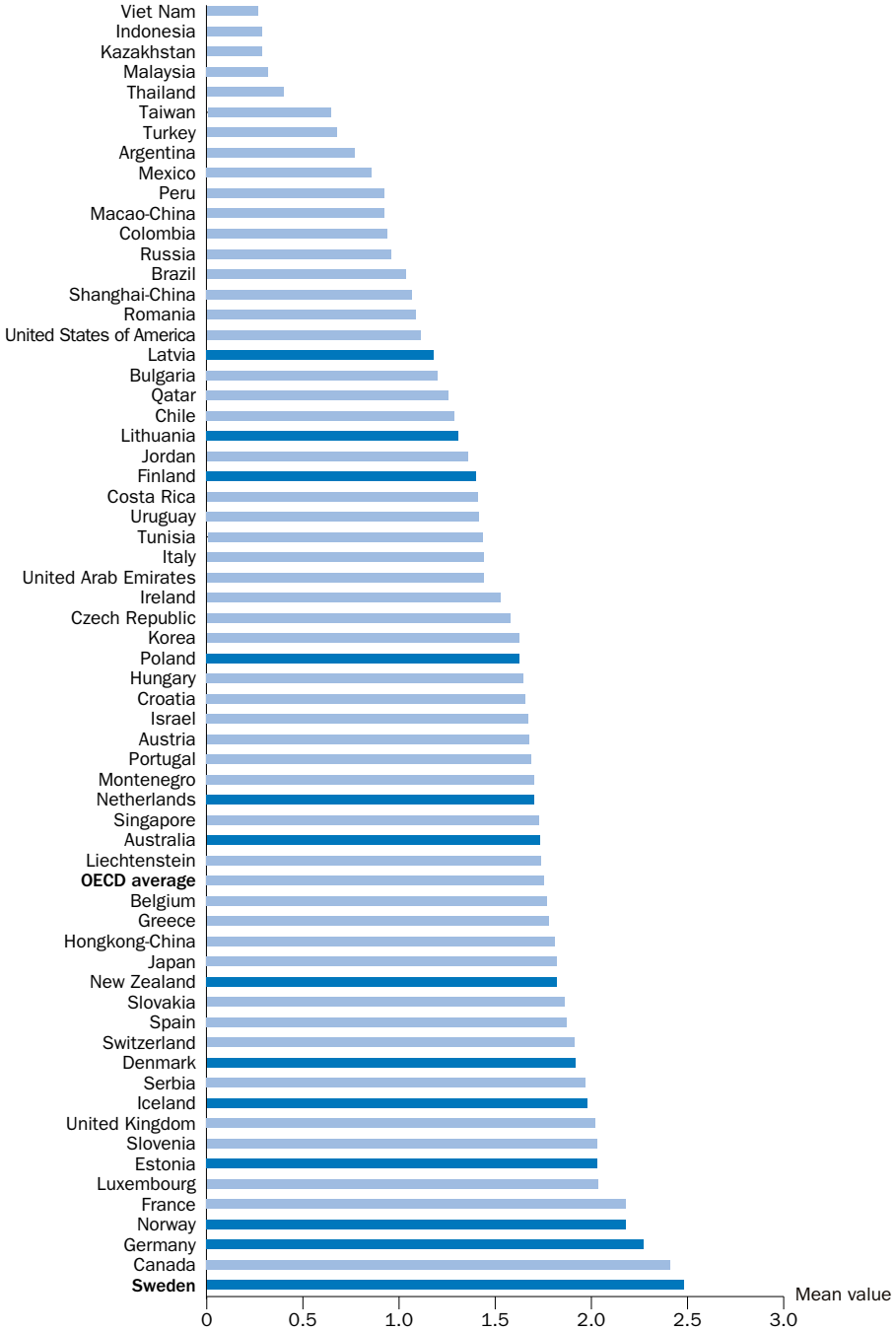


Figure 5. Difference between average reported effort in PISA and average estimated effort if the test result was going to count towards a grade, by participating country. Sweden, together with the other countries in the sample in this study, has dark blue bars.



The figures present the mean value for the OECD countries as this is an accepted standard when reporting PISA results. For the OECD countries, the mean value for reported effort in PISA 2012 was 7.49, whilst the international mean (all countries, both OECD and non-OECD countries) is 7.75.

A slightly lower mean value for the OECD countries, in other words. With regard to reported effort if the PISA test was going to count towards a grade, the OECD average is 9.24, while the international average is 9.20, basically the same. The international average for the difference variable is 1.44, while the OECD average for this variable is 1.75. Thus, on average, students in OECD countries differentiate more between the PISA test and tests on which their grades will be based.

It is possible, and probable, that the effort thermometer does not really function in the same way and does not measure quite the same thing everywhere, and so there should be caution when making comparisons between countries. This also means that comparing, for example, Sweden's result with an international average will be fairly misleading. Here, however, comparisons of the difference variable can be considered somewhat less influenced by response styles, etc. (see Butler, 2008). Still, comparisons across all 65 countries are fairly difficult to interpret. For this reason, the present report has selected a number of countries that are geographically and/or culturally close to Sweden, or which are usually otherwise of interest for us to compare ourselves with, for more detailed comparative analysis. These are the Nordic countries (Sweden, Norway, Finland, Iceland, Denmark) and other countries around the Baltic (Estonia, Latvia, Lithuania, Poland, Germany), as well as the Netherlands, Australia and New Zealand.

Table 3 below presents average reports on the effort thermometer for each country in the sample, how much the country's mean value differs from Sweden's, expressed in standardised mean differences *(d)*[3]; estimated effort if the test was going to count towards a grade; and the difference between these two reports. It also presents each country's results in mathematics in PISA 2012.

3 The standardised mean difference has been calculated using the formula for Cohen's d, and Sweden has been compared in pairs with each of the other countries as follows: d = Sweden's mean value (M1) minus *the country's* mean value (M2) divided by the pooled standard deviation of the two countries' mean values: $\sqrt{[(1_2 + 2_2) / 2}$

Table 3. Average reported effort in PISA 2012, difference (d) compared with Sweden, average reported effort if the test was going to count towards a grade, the difference between reported effort and effort if graded, difference (d) compared with Sweden and mathematics results for the country in PISA. Sample of countries.

Country	Reported effort in PISA	*d*	Reported effort if graded	Difference/ Relative effortg	*d*	Mathematics results
Sweden	7.03		9.51	2.47		478
Germany	7.05	.01	9.33	2.27	.10	514
Norway	7.16	.06	9.35	2.18	.14	489
Netherlands	7.22	.10	8.92	1.70	.40	523
Estonia	7.31	.14	9.34	2.03	.22	521
New Zealand	7.44	.21	9.26	1.82	.33	500
Poland	7.50	.23	9.12	1.62	.42	518
Iceland	7.55	.25	9.52	1.97	.23	493
Australia	7.56	.27	9.34	1.78	.37	504
Denmark	7.64	.32	9.56	1.92	.28	500
Latvia	7.66	.35	8.84	1.18	.67	491
Finland	8.00	.52	9.40	1,40	.58	519
Lithuania	8.10	.59	9.40	1.30	.62	479

Swedish students report a lower level of effort in PISA than the other countries in the sample, and a larger difference between reported effort in PISA and estimated effort if the PISA result was going to count towards their grade. In particular, the relatively large difference between the two reports could possibly be interpreted as a consequence of the national test period that Swedish students may have been in: that they at the end of the spring semester took other tests more important for their grades than PISA, and that they therefore differentiate more greatly between the different types of tests. On the other hand, Swedish students have also in previous PISA rounds had a comparatively large difference between the two effort reports. This will be covered further below.

Germany, Norway, Estonia and the Netherlands, together with Sweden, rank fairly low with regard to reported effort in PISA. Iceland, Denmark, New Zealand, Australia, Poland and Latvia are about half an increment above Sweden on average, while Finland and Lithuania are, on average, about one increment above Sweden. For the Nordic countries, it is particularly Finnish students, but also the Danish and Icelandic, who report a higher degree of effort in PISA than Swedish students. All the Nordic countries report high levels of effort if the test result was going to count towards a grade, while the Netherlands and Latvia are a little lower.

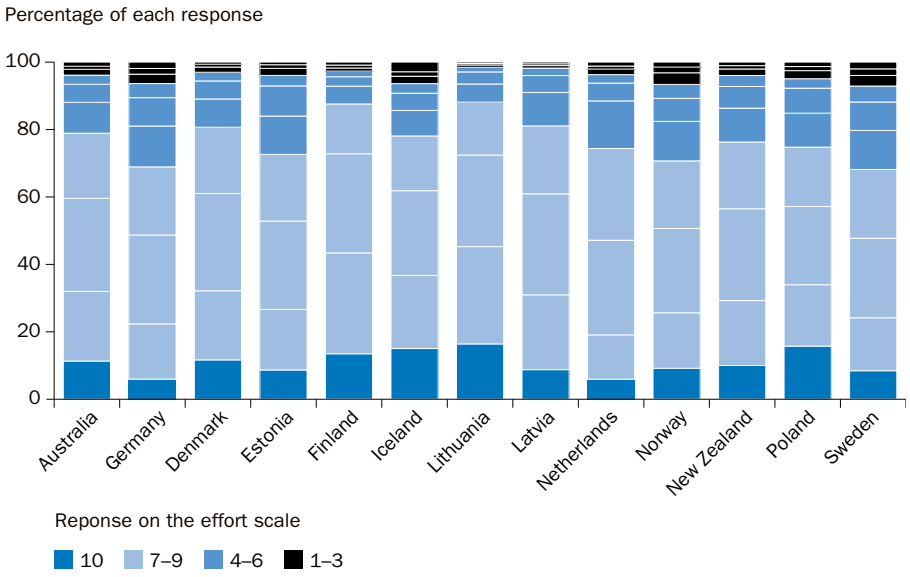
As mentioned, the measure of effect size is Cohen's d. Values of Cohen's d around 0.20 are usually regarded as a small difference in practice, values around 0.50 are usually regarded as a medium difference and values greater than 0.80 are usually regarded as a large difference. However, as noted by for example Hattie (2009), a value that is small in one context might be regarded as large in another, depending on what is being studied. Given that the total variation between countries is fairly limited for reported effort (most have a mean value between 7 and 8 and a standard deviation around 2), a full increment's dif-

ference, as in the comparison between Sweden and Finland, can probably be regarded as quite large, even if the d-value indicates that it is medium in size.

The d-values are larger for the difference variable than for reported effort (see Table 3) when Sweden is compared with the non-Nordic countries in particular. This lends additional support to the notion that Swedish students appear to differentiate more between the PISA test and tests on which their grades will be based than students in several other countries.

Figure 6 presents the distribution of responses for the various increments on the effort thermometer for the countries in the sample. It shows the share of students indicating each value on the effort thermometer (minimum effort = 1, maximum effort = 10). Since the responses are skewed (with few students reporting a very low effort), for the sake of clarity values 1–3 are black, while values 4–6, 7–9 and 10 are various shades of blue. Figure 7 then presents the average mathematics results for each value on the effort thermometer for each country.

Figure 6. Distribution of responses on the effort thermometer for a sample of countries.



Although Sweden, on average, reports lower effort than other countries in the sample, the response pattern does not deviate dramatically from those of the other countries. With regard to the amount of missing data for the effort thermometer, this varies from only about 3 per cent for Finland and Iceland, to about 10 per cent for Australia and the Netherlands. Denmark, Sweden, Norway and Poland have between six and eight per cent missing values.

For Poland and Estonia (see Figure 7), there seems to be no relationship at all between reported effort and performance. Students reporting a low degree of effort perform at least as well as those reporting a high degree of effort. For Germany and the Netherlands, we see similar patterns, while for the Nordic countries, and for Australia, New Zealand and Latvia, it is fairly clear that a higher degree of reported effort is associated with a higher performance – up to a report of nine on the effort thermometer. All countries except Finland have a lower average performance for the group indicating a ten on the effort thermometer.

Figure 7. Mathematics results by response on the effort scale for the sample of countries in PISA 2012. Sweden has a black dashed line, other Nordic countries have dotted lines and the other countries in the sample have solid lines.

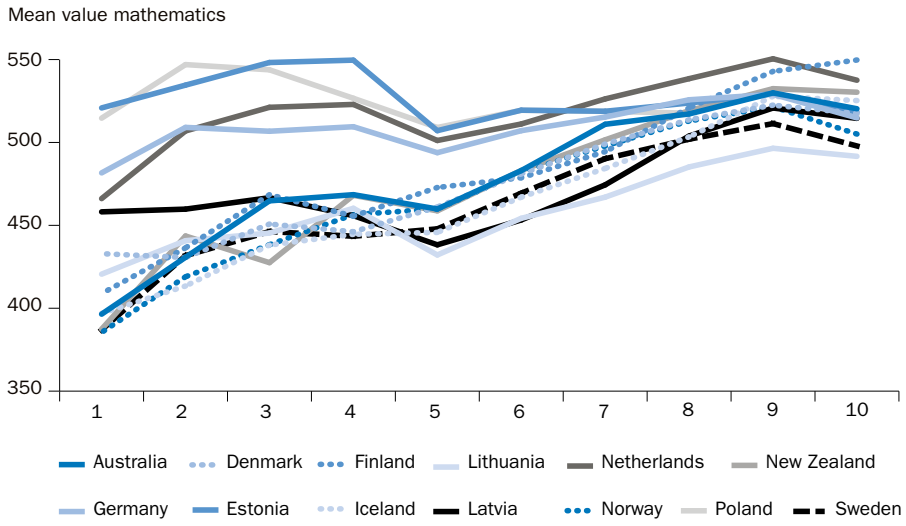
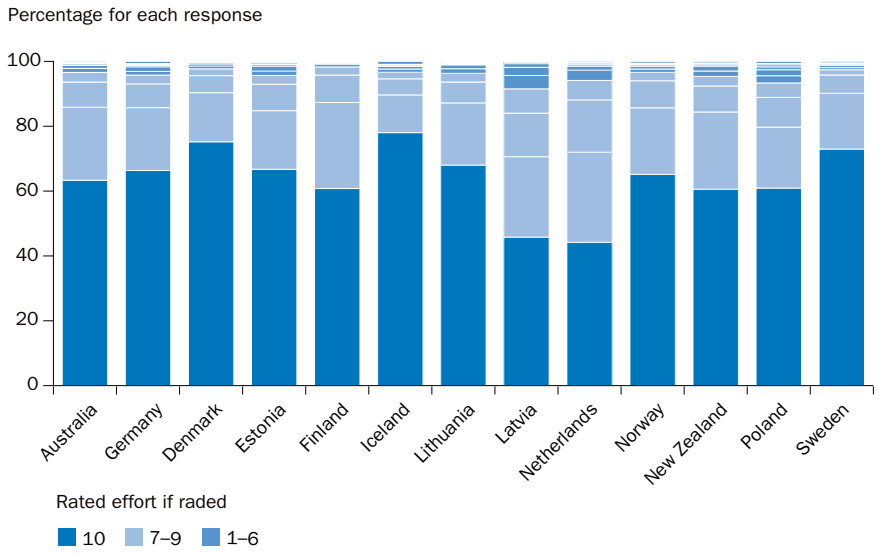


Figure 8 presents the distribution of responses to the question of how much effort students would have invested in the PISA test if it was going to count towards their grade. The bars in the figure represent the same sample of countries as above. Since the responses are highly skewed, for the sake of clarity values 1–6, 7–9 and 10 have been given various shades of blue. In all countries except Latvia and the Netherlands, it is most common by far for students to mark a 10 for how much effort they would have invested if the PISA test was going to count towards their grade. All countries also have a fairly large proportion marking a 9, while other increments have quite small shares.

Figure 8. Distribution of responses to the question of how much effort students would have invested if the PISA test was going to count towards their grade. A sample of countries.



So far, we have presented results for reported test-taking motivation and reported effort for Sweden in PISA 2012. We have also presented results for the effort thermometer internationally and for a sample of countries in PISA 2012. However, the effort thermometer has also been used in previous PISA cycles, so it is possible to study possible changes over time. This is done in the following section.

Has level of reported effort changed over time, in Sweden and compared with other countries?

In this section, we look first at changes over time with regard to Swedish students' reported effort in the PISA test. A large reduction in effort could be one explanation for Sweden's declining results, if reported effort were to have an impact on performance that cannot be explained by other variables. We then look at the selected sample of countries to see whether there has been any change over time in these countries. If other countries were to have stable effort reports over time, while Sweden's were to drop, this could explain Sweden's falling ranking – again assuming that reported effort has an impact on performance in Sweden and also in other countries, and that the effort scale can be assumed to be constant over time.

Compared with PISA 2003 and 2006, the reported effort of Swedish students has dropped somewhat. In PISA 2003, Sweden had a mean value of 7.38 on the effort thermometer. In PISA 2006, the mean value was 7.37, and in PISA 2012 it was 7.03. This is a decrease of 0.35 increments on the ten-point scale from 2003 to 2012 (with mathematics as the main subject in both years), and 0.34 increments from 2006, while there was in principle no difference between 2003 and 2006 (when the main subjects were mathematics and science, respectively). The effort thermometer was not used in 2009. The decline from the previous year to 2012 is statistically significant according to the t-test, but the effect size is quite small (d = 0.18 which is usually considered a fairly weak effect). Reported effort if the PISA test was going to count towards a grade is fairly stable between the years (9.46, 9.57 and 9.51 on the ten-point scale), while the difference between the two reports has consequently increased in 2012.

The distributions of responses to the question of how much effort the students invested in PISA have changed somewhat from 2003 to 2012, as a result of students reporting slightly lower effort, but the pattern is the same. For example, a smaller proportion of students mark an effort of ten in 2012 compared with 2003 (8.4% versus 11.6%). However, in both 2012 and 2003 (and 2006), the students marking a ten perform worse than those marking a nine. Notably, the share of students not giving any response to the question of how much they effort they invested has increased between 2003/2006 and 2012. In PISA 2003, and in PISA 2006, just over three per cent of the students did not respond to the question, while in PISA 2012 the corresponding share was just over seven per cent. As in 2012, the 100 students leaving the question blank in PISA 2003 performed on a par with those indicating an effort of one or two (412 points on average), while the 53 students belonging to the category of "other invalid responses" had an average result of 487 points. In the sample of countries, the proportion of non-responses has increased also in Australia and the Netherlands, while in Germany and Latvia there is a smaller percentage

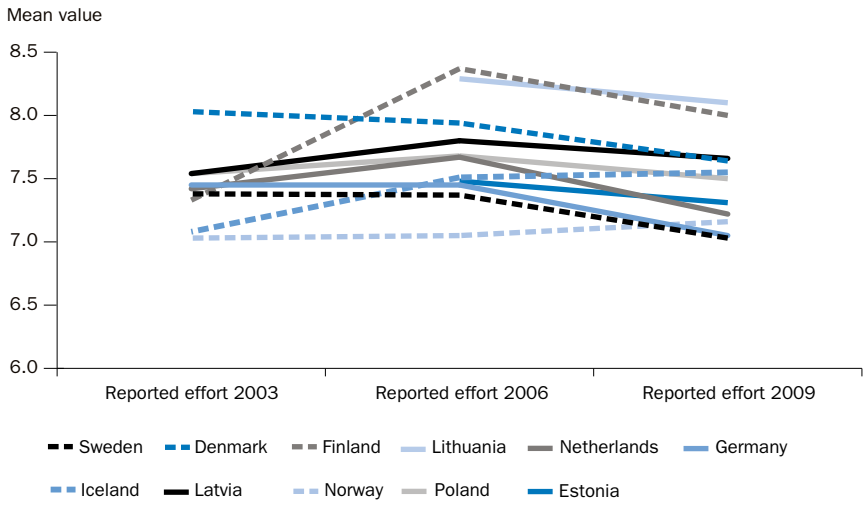
of missing values in 2012 than in 2003. Other countries have about the same amount of missing values now as they did previously. For example, compared with Indonesia, which has 53 per cent missing data on the effort thermometer, the figures are quite small for Sweden and the other sample countries. Nevertheless, it may be worth noting that Sweden's percentage of non-respondents has increased since PISA 2006, as this might say something about the students' motivation. As mentioned earlier, if students are not motivated to invest effort and do their best in the test, they might also lack the energy to go to the trouble of answering questions about how much effort they had invested. Alternatively, the larger amount of missing data might be explained by a larger proportion of low-performing students who do not manage to answer all the questions in the test booklet in time and therefore do not reach the question about effort. The effort thermometer comes at the very end of the test booklet.

In PISA 2012, there were, as mentioned, three countries reporting a lower level of effort than Sweden. In PISA 2003, there were eight countries (of the countries also participating in PISA 2012) that reported a lower level of effort (Belgium, Finland, Luxembourg, Austria, Iceland, Norway, France, Japan). Including all countries, even those not participating in PISA 2012, there were 11 countries that reported a lower level of effort than Sweden in PISA 2003. In PISA 2006, there were seven countries reporting a lower level of effort. Sweden has thus "dropped a few places" in terms of reported effort, although the decrease is not particularly large in absolute terms. It is true that Sweden reported the largest difference between reported and estimated effort in PISA 2012, but in 2006 Sweden had the second largest difference (after Norway). Swedish students have thus also in the past "differentiated between tests", and in all PISA rounds that have included the effort thermometer, Swedish students have on average reported a fairly low level of effort.

At the same time, Sweden is the country with the largest results decline of all countries since 2003, -31 points for both mathematics and reading, -21 points for science (the results are more uncertain for reading and science). From 2006, Sweden's point average has dropped by 24 points for mathematics and reading, and 18 points for science. Also for other countries is it possible to discern lower reports on the effort thermometer combined with lower performance (Denmark being one example), but the picture is complex. In Iceland, students report higher effort in 2012 than they did in 2003 and 2006, while mathematics results have declined quite dramatically. In Finland, students report higher effort in 2012 than they did in 2003 (though lower than they did in 2006), and here too the results have gone down. In Germany, on the other hand, students report lower effort in 2012 compared with 2003, while the results have improved. In Australia and New Zealand, reported effort is unchanged on average, but performance has gone down. Looking at all the countries participating in PISA 2003 and PISA 2012, there are also no clear trends. Internationally, the mode (most common value) for reported effort in the PISA test is an eight in 2003, 2006 and 2012. The mode for estimated effort if the results were going to count towards a grade is ten in 2003, 2006 and 2012.

Figure 9 presents the way the mean value for reported effort in PISA has changed over time for the countries with which Sweden is primarily compared in this report.

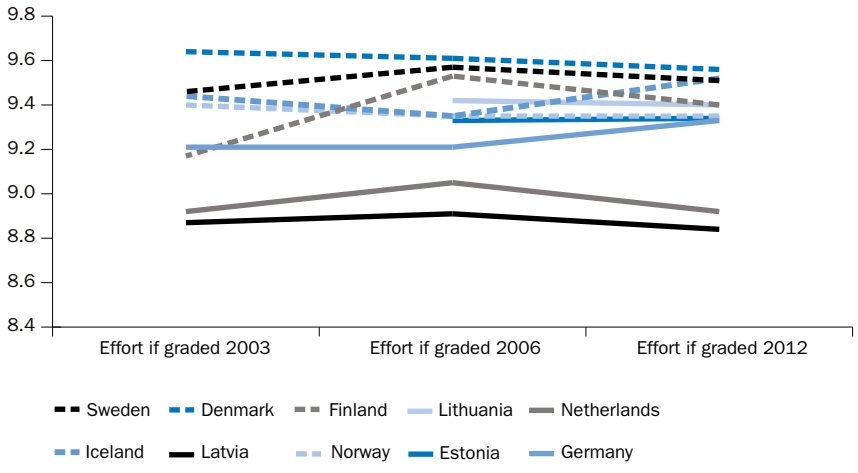
Figure 9. Trends in reported effort in the PISA test for Sweden + sample of countries. The Nordic countries have dashed lines in the figure, while the other countries in the sample have solid lines.



With regard to reported effort, Iceland shows an increase over the years, and Denmark shows a decline. Sweden and Germany are fairly stable between 2003 and 2006, and then go down in 2012. Norway is fairly stable, but shows a small increase in 2012. Finland exhibits a somewhat odd pattern, with a large increase between 2003 and 2006, and then a decline in 2012. Estonia and Lithuania were not part of PISA 2003, which is why their lines are shorter.

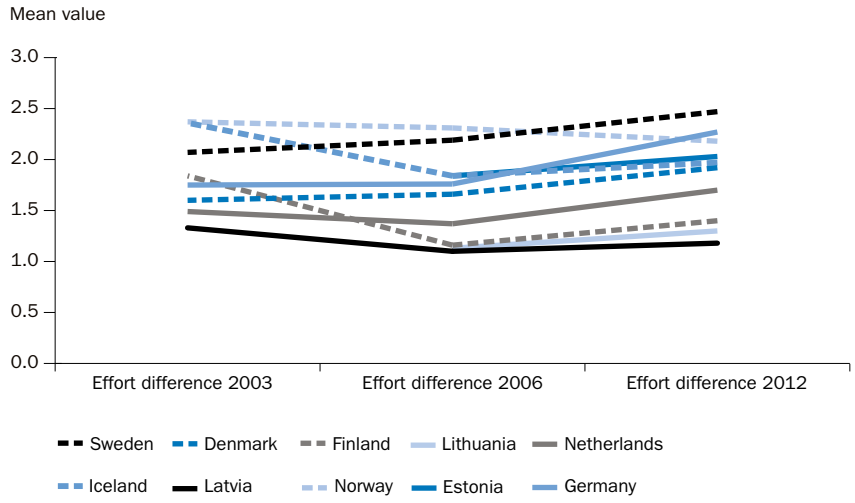
With regard to estimated effort if the test was going to count towards a grade, the reports are fairly stable between the years (see Figure 10, but note that as the scale is fairly compressed, what might appear to be clear rises or declines represent fairly small differences). Changes in the difference variable over time (Figure 11) are therefore mainly due to changes in reported effort in the PISA test.

Figure 10. Trends in reported effort if the PISA test was going to count towards a grade.



Latvia and the Netherlands stand out slightly as they on average indicate a lower degree of effort for tests on which grades are based compared with the other countries in the sample. Finland exhibits a similar pattern as in the previous figure, a fairly large rise between 2003 and 2006, and then a decline in 2012. Otherwise, there have been no major changes if the countries are compared with themselves.

Figure 11. Trends in the difference between reported effort in PISA and the estimated effort if the test was going to count towards a grade.



This section focuses on changes over time with regard to reported effort in PISA, and estimated effort if the PISA test was going to count towards a grade. We have seen that Swedish students' reported effort in PISA has gone down somewhat if we compare 2012 with 2003 and 2006. The estimated effort if the test was going to count towards their grade is fairly stable over the years, and the difference between the two reports has increased.

The following deals with one central question: what is the correlation between reported effort and performance, and can the lower degree of reported effort explain the deterioration in PISA performance? This is discussed in the following two sections, and we begin by looking at correlations between reported effort and performance, from a Swedish and an international perspective.

Is there a correlation between reported test-taking motivation and test performance?

In absolute terms, there is not a very large change in reported effort between PISA rounds, either in Sweden or in the other countries Sweden is compared with here. Nevertheless, it is a decline and should effort be strongly correlated with performance, this decline could be important to take into account when interpreting changes in results. The following presents an analysis of correlations between reported effort and performance.

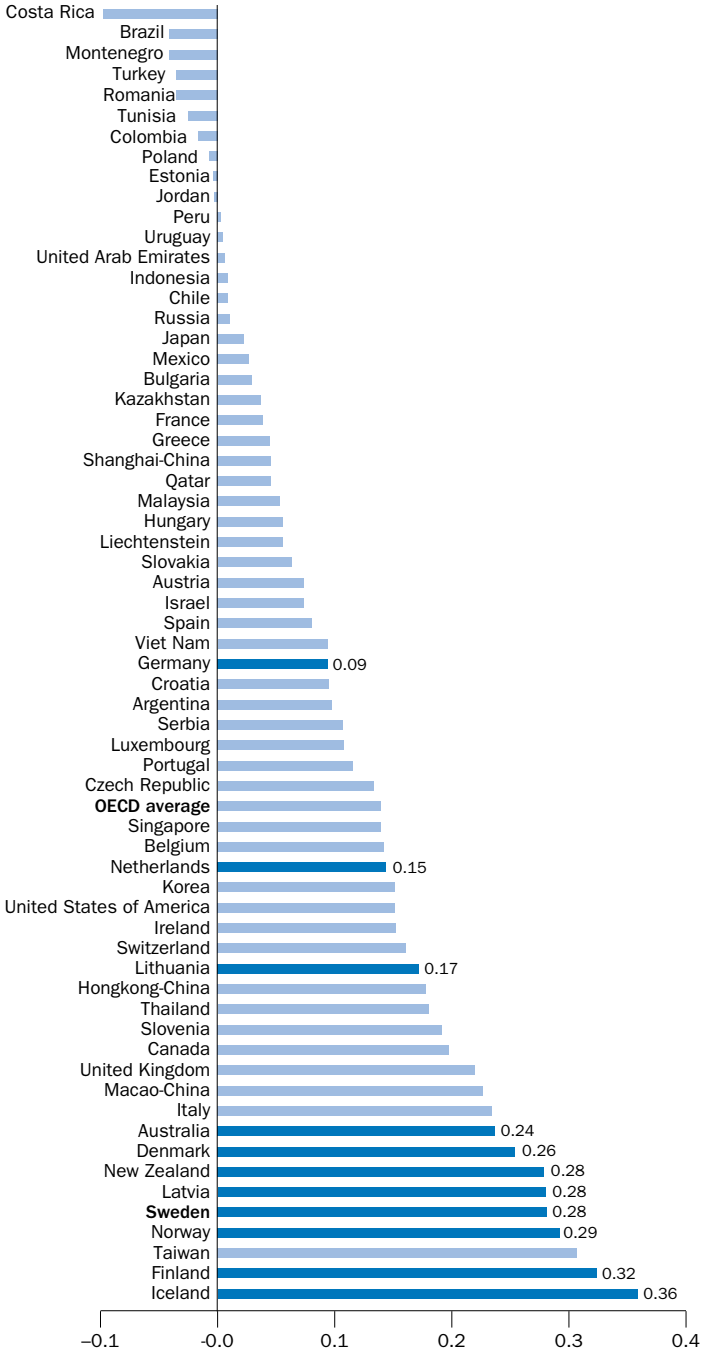
A comparison that includes all countries participating in PISA 2012 (see Figure 12) clearly shows that the correlations are different for different countries. In some countries, there is in principle no correlation between reported

effort and PISA performance, and in some countries the correlation is even negative. Most countries show a positive correlation, but the strength of correlation varies from country to country, possibly due to the effort thermometer functioning a little differently in different countries. Figure 12 presents correlations (expressed as "*r*") between reported effort in PISA and mathematics performance in PISA 2012. The correlations are similar for reading and science. The average correlation between reported effort and performance for the OECD countries is $r = .14$, somewhat higher than the international mean value of $r = .10$. The corresponding OECD mean values for correlation with reading and science are $r = .17$ and $r = .15$.

Of all the participating countries, Iceland reports the strongest correlation (expressed as "*r*", the value above each country in the sample), followed by Finland. In Sweden, the correlation between reported effort and mathematics performance is $r = .28$, and the same or similar strength of correlation is found in Norway, Denmark, Latvia, New Zealand and Australia. In Germany, the Netherlands and Lithuania, the correlation is positive but fairly weak. In Estonia and Poland, there is no correlation between reported effort and performance. For Sweden, the correlation between effort and performance has been fairly stable over the years and for the three different subjects. The correlation coefficient varies from .25 (Science in 2003) to .32 (Reading in 2006).

The correlation between reported effort and performance appears to be particularly strong in the Nordic countries. Of the eight countries with the strongest correlation, five are Nordic. Although this cannot be interpreted to mean that it is only in the Nordic countries that effort has any significance, it perhaps says something about differences in how the effort thermometer is interpreted and responded to in different countries, how the response scale is interpreted, that it is acceptable in some countries to say that you do not invest any effort, that students in some countries differentiate more greatly between tests, or that there are other reasons that we cannot know anything about based on the available data.

Figure 12. Correlation between reported effort and mathematics performance in PISA 2012, all countries. Sweden, together with the other countries in the sample in this study, has dark blue bars. The strength of correlation (r) for the countries in the sample is given above each bar. The higher the value, the stronger the correlation, values between -1 and +1 are possible.



The majority of countries thus have a positive correlation between reported effort and PISA performance, but no country has a very strong correlation. Correlations of the size .20 – .30 can be considered fairly modest, but not negligible. Figure 12 (the correlation figure) shows that it is fairly misleading to talk about correlations between reported effort and performance at the aggregated level, or to compare an individual country with an international mean value, as this says fairly little about the situation in other individual countries. The international mean for the correlation between effort and performance is .10, but if, for example, this were to be used to claim that effort has a positive but only weak correlation with PISA performance, it would not provide a particularly good picture of the situation in the Nordic countries, nor in Costa Rica, Brazil or Montenegro.

There is also no clear link at the country level between the level of reported effort and the correlation between effort and performance. Finland and Taiwan are countries that report relatively high levels of effort, while Sweden and Norway report relatively low levels of effort, and in all these countries the correlation between reported effort and performance is comparatively strong. There are also examples of the opposite. In Russia and Shanghai, students report a high degree of effort, while students in Japan report the lowest degree of effort, and in these countries the correlation between effort and performance is in principle 0.

For example, in Poland and Estonia, countries that have significantly raised their performance in PISA, there is no correlation at all between reported effort and performance, either now or in previous PISA surveys (although Estonia did not participate in PISA 2003).

To exemplify the fact that varying correlations between effort and performance might be due to self-report scales functioning differently in different countries, a correlation analysis corresponding to that in Figure 12 was conducted for the PISA scale which measures self-concept in mathematics ("I learn things quickly in mathematics", "Mathematics is one of my best subjects", etc.) and mathematics performance. The results are presented in Appendix 1, Figure A1. The analysis of this scale similarly shows that there are fairly large differences between countries regarding the strength of the correlation, and that it is in principle the same countries that exhibit the strongest correlations between reported self-concept and performance as between reported effort and performance: all the Nordic countries are in the top eight for both figures.

With regard to the Swedish test-taking motivation scale (see Table 1), the students' report on this scale is also positively related to performance in PISA (r = .21). The individual items in the scale are also positively correlated with performance. However, for the two items concerning whether good PISA performance is perceived as important, the correlation is weak (r = < .10) as the relationships are curvilinear rather than linear, while the correlations are a little stronger for the items concerning the motivation to invest effort (r = .15 – r = .25).

Thus, in summary, there is a positive and significant correlation between test-taking motivation and performance in Sweden, and this correlation is of the same strength as seen in several previous studies. In the other Nordic countries, and in several of the countries that are "similar to us" and that we tend to compare ourselves with in international studies, there are also clear correlations between reported effort and performance, whilst from an international perspective, there are examples of countries where there is no correlation at all, coun-

tries in which the correlation is very weak and countries where the correlation is negative.

The next section presents results from a number of analyses studying the impact of reported effort on performance in a little more detail: whether the impact is constant over the years, whether other motivation variables measured in PISA are more important than effort, and whether the decline in level of reported effort can explain the Swedish decline in PISA.

Might test-taking motivation explain the decline in Swedish results?

We have so far established that Swedish students report a comparatively low degree of reported effort in the PISA test in 2012. We have also seen that the Swedish students' reported degree of effort has declined somewhat since 2006, while there was no difference in reported effort between 2003 and 2006. Furthermore, there appear to be positive correlations between reported effort, test-taking motivation and test performance so that a higher reported motivation and effort is associated with higher performance in the test. These correlations have been fairly constant over the years for Sweden.

We will first take a closer look at how much of the students' performance could be explained by effort/motivation and whether the declining results might be due to the students investing less effort now than previously.

But, even if we were to identify a significant impact of effort on performance, this does not necessarily mean that the impact remains when other variables are entered into the analysis. It could, for example, be the case that the students' reported effort might actually be explained by how good they think they are at mathematics, how persevering they generally are in their school work, or how interested they are in mathematics. Were this to be the case, a potential impact of effort will disappear or decrease when these other aspects are taken into account. For this reason, we have also had the ambition to analyse whether the students' reported degree of effort/test-taking motivation might explain some of the variation in mathematics results in PISA 2012 beyond that explained by other variables, such as intrinsic and instrumental motivation, mathematics self-concept, mathematics self-efficacy, mathematics anxiety, socio-economic background, etc.

Can a lack of effort be the explanation for the Swedish students' declining results?

The regression analysis estimated how much of the Swedish students' performance in PISA is explained by their (self-reported) degree of effort and motivation and how much of the decline in performance in the different subjects could be explained by a decline in effort. The independent variable in this analysis was reported effort in PISA (the effort thermometer). The dependent variable was the result in mathematics, science and reading. The same analysis was conducted with the test-taking motivation scale, but this only includes results from 2012 since the test-taking motivation scale has not been used previously. The results from the analysis are presented in Appendix 1, Table B1.

The B and β–coefficients of the regression analysis

A regression analysis generates a number of values that indicate how great an effect one or more variables has on other variables. Here we want to know whether the effort thermometer (and later also other variables) has an effect on performance in PISA. Some of these values are the B coefficient, the β coefficient and R2, the proportion of explained variance. Here, the B coefficient, or b value, can be interpreted as meaning that an increase of one increment on the effort thermometer yields a performance increase corresponding to the b value. The β coefficient, or β value, is a standardised measure of the effect. The β value is the same as the correlation coefficient if there is only one independent variable. R2 indicates how much of the variation in performance can be explained by reported effort (or other independent variables).

For Sweden, the analysis with the effort thermometer as a predictor of mathematics performance in PISA 2012 resulted in a b value of 12.20. Similar values were obtained for the other subjects and for other years (see Table B1 in Appendix 1). Thus, statistically, an increase of one increment on the effort thermometer yields an increase of 12.20 points in the PISA test, a significant effect. (Example: An increase from 7 to 8 on the effort thermometer would, according to the regression analysis, yield 12.20 more points in the test, e.g. 490 points instead of 478). This also corresponds fairly well with the initial descriptive results for average results in relation to increments on the effort scale: the higher the reported effort, the higher the performance. The regression analysis also shows that effort explains about 8 per cent of the variation in the students' performance in mathematics. This impact is fairly constant within each subject if we compare results from PISA 2003 and PISA 2006 with PISA 2012. The impact is also in principle the same across the subjects, even though a somewhat stronger impact is seen with regard to reading.

The impact of effort on performance in the various PISA subjects is thus about as strong now as it was in 2003, but the students report a lower degree of effort now compared to then. Could this be the cause of the Swedish students' poorer results? According to the regression analysis, this might be a contributing factor, but hardly the main explanation. As we have seen in previous sections, the Swedish students' reported degree of effort has gone down on average by 0.35 increments on the effort scale since 2003. A decrease of one increment (for example, from a value of eight to a value of seven) would mean 12.20 PISA points less. An increment change of 0.35 corresponds to less than this, between 4 and 5 PISA points (taking an average of the impact in 2003 and 2012). However, Sweden's results have fallen significantly more than this, in mathematics by 31 PISA points since 2003. Some of these points (or ~ 15% of the decline) could thus be explained by a lower degree of effort.

We also looked at effects of students' decreased effort in relation to PISA 2006, since it is between 2006 and 2012 that we see a decline in the students' reported effort. Between 2003 and 2006, there was no decline in reported effort (however, the main subjects were different in these years, which could have some effect on the students' reported effort), but a small decline in performance, in all subjects. The decline in performance between 2006 and 2012 is 24 points for mathematics and reading, and 18 points for science. Thus, in theory, 4–5 points of this could be explained by lower effort by Swedish students.

Note that the results trend for science in particular is somewhat uncertain as science has only been a main subject once.

With regard to the test-taking motivation scale, this too is a significant predictor of performance. According to the regression analysis, an increase of one increment on the test-taking motivation scale yields an increase of 26 points in the PISA test. This is a higher value than for the effort thermometer, but then the test-taking motivation scale does not have as many increments and a comparison is therefore not meaningful. The test-taking motivation scale is also a significant predictor of results in Swedish and science (see Table B1 in Appendix 1).

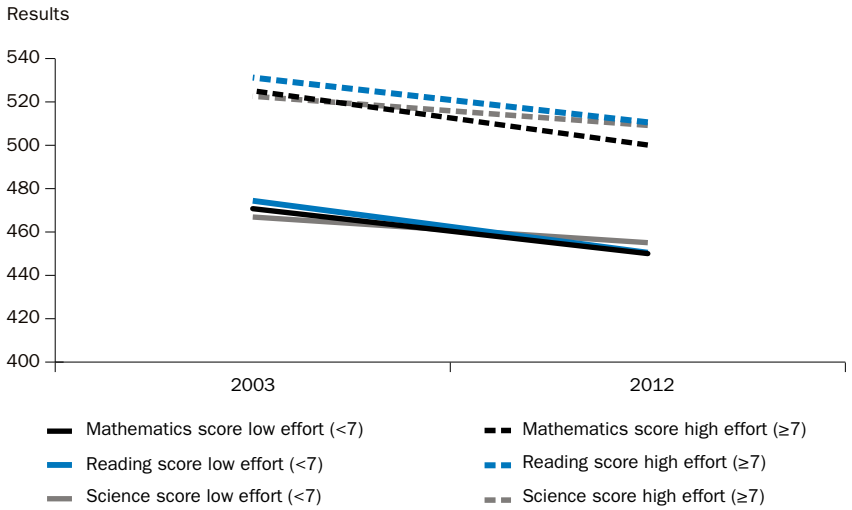
If we were to correct for the lower average degree of self-reported effort (as measured by the effort thermometer) in PISA 2012 and estimate the students' results, the Swedish students would have had an approximately 4.5 points higher result in mathematics, 483 points instead of 478. However, the regression does not actually say anything about causality (if it is the higher degree of reported effort that actually causes the better performance), and we do not know what the situation would have really been if Swedish students were to have reported the same degree, or a higher degree, of effort now compared with previously. There are several examples of countries, such as Finland and Iceland, where there are just as strong correlations between effort and performance as in Sweden, and where students now report a higher degree of effort than they did in PISA 2003, whilst performance has still gone down. It is therefore difficult to draw far-reaching conclusions solely on the basis of the regression analysis.

It can, however, be established that the students' reported level of effort appears to have a significant impact on performance, that a certain portion of the Swedish results can be explained by effort, but that lower level of effort can hardly be a main explanation for Swedish students performing at a lower level now than they did in 2003. The majority of the Swedish decline is likely to be quite simply explained by other factors, such as that the students have poorer skills in domains measured by PISA, which in turn can be due to many different things.

Change in results for students reporting a high/low degree of effort

In PISA 2003, 25 per cent of students reported an effort in PISA corresponding to a six or less on the effort thermometer. In PISA 2012, this share had increased to 32 per cent. Could it be that the results of students reporting low effort have deteriorated more than those of students reporting high effort, and could such a shift thus be an additional explanation for the general decline? The correlation and regression analyses above give no indications of this, but to further illustrate this, Figure 13 presents average results for the years 2003 and 2012 for mathematics, reading and science, according to whether the students reported a low (less than seven on the effort scale) or a high (seven or higher on the effort scale) degree of effort in the PISA test.

Figure 13. Average PISA result for students reporting a low (six or lower on the effort thermometer) or a high (seven or higher on the effort thermometer) degree of effort, respectively. By subject and for PISA 2003 and 2012.



The decline in results is of the same magnitude for those reporting a high degree of effort as for those reporting a low degree of effort in PISA. In mathematics, the decline is even somewhat larger for those reporting a high degree of effort. Thus it is not the case that those reporting high perceived effort in the test perform just as well now as previously and that it is the students who have a low self-reported motivation that give rise to the deterioration in results. The result is most reliable for mathematics, as this was a main area both in 2003 and 2012, but the trends look the same for all subjects.

Does effort have any impact on performance in the sample of countries?

The regression analysis with all the countries in the sample shows that all the Nordic countries have b values between 11.95 (Denmark) and 16.65 (Finland). Australia, Latvia and New Zealand have values similar to that of Sweden. Germany, the Netherlands and Lithuania have lower, but still significant values. Poland and Estonia have insignificant values (the b coefficients are in principle 0, and the explained variance is 0). Thus, in Poland and Estonia, reported effort has no impact on performance, whilst the impact is strongest in the Nordic countries, together with Australia, Latvia and New Zealand. The explained variance in the other countries in the sample, expressed as R^2, is .01, .02 and. 03 in Germany, Lithuania and the Netherlands, respectively, and between .06 and .13 in the Nordic countries, Australia, Latvia and New Zealand. In other words, according to the simple linear regression, between 0 per cent (in Poland and Estonia) and 13 per cent (in Iceland) of the variation in the mathematics results is explained by the students' reported effort in the PISA test. For Sweden, this figure was 8 per cent.

All the results presented so far apply to the entire sample, without looking at different subgroups. Some previous studies have shown that there are differences between boys and girls with regard to motivation to do their best on low-stakes tests: that girls tend to report a higher degree of motivation and effort, while the

correlation between motivation and performance has sometimes been stronger for boys. However, there are also studies that have not shown any major gender differences. To investigate the situation in PISA for Sweden in particular, analyses were run for boys and girls separately. We now present overview results of this analysis.

Are there any differences between boys and girls?

In PISA 2012, Swedish girls had a mean value of 7.26 on the ten-point effort scale, while for boys this was 6.80. The girls thus report a higher degree of effort than the boys do. The correlation between reported effort and performance is also somewhat stronger for girls, especially in mathematics, while for reading it is in principle the same.

On average, girls also report a somewhat higher degree of test-taking motivation, as measured by the test-taking motivation scale, compared with boys. The mean value for the total sample on the scale is 2.37, for girls it is 2.35 and for boys it is 2.40 (note that here a lower value stands for a higher degree of motivation). It is above all for the items "I felt motivated to do my best…" and "I did my best…" that the girls give higher reports, while the differences are fairly small for other items. Overall, however, the differences between boys and girls are not particularly large when it comes to reported test-taking motivation. The correlation between reports on the test-taking motivation scale and performance is somewhat stronger for the girls.

The analysis of gender differences with regard to reported effort in the sample of countries reveals the same picture as for Sweden: girls report a higher degree of effort than boys do. The smallest gender difference is found in the Netherlands and New Zealand, where this is only around one tenth of a scale point. Several countries have gender differences of about the same size as Sweden, around half an increment, while in Poland, for example, the difference is almost a whole increment.

An analysis of trends in reported effort for boys and for girls in Sweden shows that the decline is larger for girls. In both 2003 and 2006, the gender differences were larger with regard to reported effort. The respective decrease in reported effort for girls is 0.46 and 0.43 increments on the effort scale when 2012 is compared with 2003 and with 2006. The respective decrease for boys is 0.25 and 0.26 increments. The girls thus report relatively lower effort now than previously, but still report higher effort than the boys.

The impact of effort on performance is fairly similar for boys and girls, over the different years and the different subjects. However, there is a difference of one (in reading) or more (in mathematics) points, where the b values are higher for girls. The impact of effort on performance is thus somewhat stronger for girls than for boys, but on the whole, there are no major differences. There are examples of countries where gender differences are much larger. In Germany, for example, reported effort is significantly related to performance for girls, while for boys there is almost no relationship at all.

Gender could attribute around one point more of the change in girls' results to reduced effort, and around one point less of the change in boys' results. We might therefore have hypothesised that the boys' results would have deteriorated somewhat less than for the girls. In practice, however, it is the opposite: it is the

boys' results that have deteriorated most. In mathematics, for example, boys performed better than girls in PISA 2003 and PISA 2006 (even though they reported a lower degree of effort than girls). In PISA 2012, there is no difference between boys and girls in mathematics performance. In reading, girls have always performed better but in PISA 2012, the difference in favour of girls is greater than it has been previously. These results also show that there are probably no simple correlations or simple explanations for the declining results of Swedish students or for how effort and performance interact, but relationships between non-cognitive and cognitive variables are complex phenomena.

This study has only conducted overview gender analyses. More in-depth analysis would need to be done in order to reach a better understanding of the dynamics of gender, effort and performance. The results are in line with previous studies with regard to girls reporting a higher degree of effort and motivation than boys, but unlike previous studies the results do not show that the correlations between effort/motivation and performance are stronger for boys.

Does test-taking motivation have any impact on Swedish students' test performance when we control for other background variables that are usually related to performance?

As mentioned in the introduction, motivation can be measured at different levels. Test-taking motivation is measured at the situation-specific level, in the form of the individual's motivation to do their best in a given test or a given task. It is perhaps most common for attitudes and motivation to be measured at the domain-specific level (motivation to learn mathematics) or more general level (motivation to learn and do better in school in general). It is reasonable to hypothesise that there are correlations between domain-specific and situation-specific motivation, and it might therefore be worthwhile to analyse them together to see if they in fact seem to be measuring the same construct. PISA's student questionnaires include several questions that are in one way or another related to domain-specific motivation. PISA has developed scales, indices, of these questions as measures of mathematics self-concept, intrinsic and instrumental motivation to learn mathematics, self-efficacy, mathematics anxiety, perseverance, attribution of failure, etc. (OECD, 2013b). The results of the Swedish students on these scales have been reported in the national report for PISA (National Agency for Education, 2013), and in more detail in the OECD report "PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs" (OECD, 2013b), for which reason they are not covered further here. By way of background, it will suffice to mention that the Swedish students report a higher degree of positive self-concept, higher intrinsic and instrumental motivation, higher self-efficacy, but also higher mathematics anxiety now compared with 2003. Compared with the OECD average, Sweden does not deviate in any extreme manner with regard to its levels on these scales, but Swedish students do come comparatively low with regard to mathematics anxiety.

The following indices were used as variables in the regression analysis, together with the effort thermometer:

- *Mathematics self-concept* – whether students think they are good at mathematics, that they learn easily, etc.
- *Self-efficacy* – whether students feel confident that they can solve various mathematics tasks
- *Intrinsic motivation* – interest in mathematics
- *Instrumental motivation* – the value of mathematics
- *Mathematics anxiety*
- *Perseverance*
- *Attribution* – how students explain failure in mathematics (internal/external causes)
- *The test-taking motivation scale* (6 items for analysis of only the Swedish sample).

The analysis also used the PISA index of economic, social and cultural status (ESCS) to take into account socio-economic background.

All of these variables have significant correlations with mathematics performance when analysed separately. The correlations are quite strong in some cases; in particular for self-concept and self-efficacy, and are positive for all variables except test anxiety. The patterns in principle look the same in the Nordic countries, while the correlations are somewhat different in some of the other countries. For example, in Germany and the Netherlands, there is a weaker correlation between self-concept and performance, but a stronger correlation between self-efficacy and performance.

This particular regression analysis only used parts of the samples in the different countries. This is due to the fact that PISA 2012 used three different questionnaire versions, which means that not all students answered the same questionnaire items. Instead of estimating responses for the students (one third of the total sample) who did not answer certain questions, we here used only those students who took questionnaire version B, which included all the above indices. It may be noted that the Swedish students who answered version B have a somewhat higher mean value on the effort thermometer compared with the total sample. As this particular analysis does not primarily aim to study the degree of effort, but to more generally see whether effort has significance for performance even when a number of other aspects are taken into account, the analysis has been carried out without further corrections.

The results are presented in more detail in Appendix 1 (Table C1), but it may in summary be noted that the impact of reported effort on performance remains even when a number of other variables and their impact on performance are taken into account. For Sweden, the impact on performance decreases somewhat, from about a 12 points per unit increase on the effort scale to just over 8 points (note that this then only really applies to those students who took questionnaire version B), while the impact of some of the other variables (attribution of failure, perseverance) disappears completely when variables such as self-concept and self-efficacy are entered into the model. Effort has about the same impact on performance as socio-economic background and mathematics anxiety (for mathematics anxiety the impact is negative: the more anxiety, the

poorer the performance). In the previously mentioned thesis on effort in PISA 2003, Butler arrived at roughly the same result (Butler, 2008).

One ambition of this regression analysis was to see if it was possible to "explain away" effort by means of this aspect having a great deal of overlap with aspects more related to domain-specific motivation. However, the effect of reported effort on performance remains and the effort thermometer thus appears to measure something not covered by the existing scales in the PISA questionnaire.

When a corresponding multiple regression was carried out for the Swedish sample using the test-taking motivation scale instead of the effort thermometer, the test-taking motivation scale was also a significant predictor, even when other variables were included in the analysis.

A corresponding analysis was also conducted for the other countries in the sample of countries. For all other countries in the sample except Poland and Estonia, reported effort is a significant predictor of performance, even when controlling for a series of other motivation variables, and for socio-economic background (as measured by the index of economic, social and cultural status) (see Table C1, Appendix 1).

Below, in conclusion, we discuss the results from the analyses presented in this study. Do the results suggest that test-taking motivation is an important factor to take into account when interpreting the results from PISA, or can we perhaps ignore the factor of test-taking motivation?

Summary discussion

A majority of Swedish students say they have been fairly motivated and agree that they did their best in the PISA test. A majority of students also mark fairly high reports on the ten-point effort scale. At the same time, fairly large proportions of students, albeit a clear minority, report a rather low degree of motivation and a low degree of effort.

Is this result to be interpreted as good news or bad news? Are the students generally motivated or not? Is it positive that a majority of students report that they were motivated to do their best, or is it a problem that a fairly large proportion, albeit a minority, reports that they did not feel so motivated, that they did not do their best? Of course, the optimal situation would be if all students were to agree that they felt motivated to do their best in the test and if they indicated maximum effort. However, PISA is a test without consequences for students. Individual results are not disclosed to either students or teachers, and their grade is not affected by their test score. For this reason, it does not seem reasonable to expect that all students would perceive the test as extremely important or that all would invest maximum effort. There is probably no test for which 100 per cent of test-takers report maximum motivation and effort. Even in the context of, e.g. the Swedish Scholastic Aptitude Test, a test whose result can have major consequences for the person taking it, and for which similar motivation questions have been used in a questionnaire study, there is a certain percentage of test-takers who indicate that they did not feel motivated to do their best, even if this percentage is lower than in PISA (see Eklöf & Knekta, 2014).

An international comparison with all participating countries in PISA 2012 shows that Sweden is one of the countries where students report the lowest degree of effort in PISA 2012, the highest degree of effort if the test was going to count towards a grade, and the largest difference between these two reports. This could be serious, but even though Sweden reports a low level of effort compared to most other countries, we do not deviate in any extreme manner; there are quite a number of countries in the same range as Sweden. Similar results in previous PISA rounds have been interpreted by the OECD to mean that motivation is not a major problem since all countries on average report a similar degree of effort (OECD, 2007). However, there are actually some differences between countries in reported effort, and there are some differences in the relationship between reported effort and performance.

The present report has on several occasions discussed why it can be difficult, and perhaps also inappropriate, to compare an individual country's results for a variable such as reported effort with all the other participating countries, or indeed with the international mean value. Just because it is possible to make comparisons does not always mean that it is appropriate to make all the comparisons that are possible, so to speak. This might, at times, instead yield misleading results.

For this reason, a sample of countries – consisting of the Nordic countries, countries around the Baltic, the Netherlands, Australia and New Zealand – was used for the comparative analyses in this report. Even when we compare ourselves with this sample of countries, Sweden reports the lowest degree of effort,

although here too the differences are very small in comparison with countries such as Norway and Germany.

However, compared with Finland, for example, Sweden has a significantly lower level of reported effort.

The analysis of reported effort over time shows that there has been a decline for the Swedish students of approximately 0.35 increments on the ten-point effort thermometer since PISA 2003. Although this is not a massive decline, it may still be important to consider. It might say something about the changes that have taken place in the Swedish assessment system over the past decade. In previous PISA surveys, the reforms of the Swedish national assessment system had not yet been implemented, and so the students did not have the quantity of national tests that they have today, and as they had in 2012. It is possible that the increasing number of high-stakes tests is a cause that contributes to their reporting a somewhat lower degree of effort in PISA, which in 2012 coincided with the national test period and which then may have emerged as a less important test, at least for some of the students. However, if this hypothesis were to have any relevance, it appears even more important to continue to monitor the students' reported degree of effort and motivation in PISA. Students in forthcoming PISA rounds will have even more experience of tests with stakes attached to them than the students in previous PISA rounds had. With more experience of tests and grades, students can be expected to be more aware of where it is worthwhile to invest effort or not. Schools, teachers and others will need to really communicate the importance of students investing effort in studies such as PISA.

For all the years that effort has been measured in PISA, a significant impact of effort on performance is seen in all three subjects surveyed in PISA. This impact is about equally strong for all subjects and all years. However, the impact of effort on the performance of Swedish students is not so strong, and the decline in reported effort not so great, that there is a basis for drawing the conclusion that lower effort is the cause of the Swedish students' lower performance. According to the statistical analysis, an increase of one increment on the effort thermometer yields a performance increase of just over 10 points. However, the change in reported effort is significantly less than a full increment, indeed, less than half an increment. By this logic, the lower degree of effort could explain a few points of the decline that has taken place since 2006, while the majority of the decline has other explanations.

On a couple of occasions, the present report has made comparisons with the same population of students in a national test context, where the students taking a national test reported a higher degree of motivation and effort than the students participating in PISA. Can it then be assumed that students' performance on a national test is closer to their "personal best"? Yes, perhaps, but as usual, reality is complex and there are generally multiple reasons for a given performance in a given situation, motivation being one of them.

It is also, in a more general sense, difficult to determine what a "true" performance is. The state of the Swedish education system is often interpreted by means of grade averages, national test results and results in international comparative studies. If, for example, we were to compare results in national tests with results in PISA, it is not obvious, however, that a given individual would attain the same result in both tests. Neither do national tests, grades, nor

PISA results present the truth; they represent different measures with different strengths and weaknesses. It is reasonable that students with a solid knowledge of mathematics, for example, should be able to do well both in a national test and in PISA's mathematics domain, but were the results to differ, it is not necessarily the case that one result is "more correct" than the other. This is a further reason for also attempting to measure socio-emotional skills in the test situation, such as motivation and test anxiety, in order to be in a better position to form an idea of how different variables come into play in different test situations.

There are no simple correlations or simple explanations for either the Swedish students' declining results or for how effort and performance interact. In-depth studies of different aspects from varying perspectives can help to put pieces of the puzzle into place and so gradually provide a greater understanding of students' knowledge development and of the interaction between non-cognitive and cognitive variables. However, as complex phenomena are involved, the simple answers are rarely exact representations of reality.

With regard to the effects of effort/test-taking motivation, consideration could be given to whether there might also, besides the direct impact on test performance illustrated in this study, be an indirect impact, in so far as lack of motivation to invest effort in PISA could also be an indication of how much energy students have for investing effort in school in general, with homework, assignments, etc. It is conceivable that the students' reported effort in PISA says something about the students' attitude beyond the PISA test: that there might be an indirect impact that goes from reported effort in PISA via a more general reduction in effort at school, which in turn leads to poorer results in school. This is a hypothesis worth consideration in future studies.

Limitations and further studies

The estimate of how much of the students' performance can be explained by effort is based on a statistical model that looks at linear relationships and which is also based on self-report. The method of self-report has its limitations: it assumes, first of all, that everyone understands the question in the intended manner and, secondly, that they respond to that particular question without weighing in other aspects and, thirdly, that everyone responds honestly and in accordance with (the perceived) reality.

Although this analysis has had the ambition of deepening our knowledge of Swedish students' reported motivation and effort in PISA, over time and in comparison with other countries, there are still many aspects that remain to be studied in more depth.

Two such aspects are differences between boys and girls and what might be the reasons for these differences, as well as more in-depth studies of the students who do not respond to the questions about effort and motivation.

Another aspect is the possible differences between schools with regard to reported effort. Since PISA, as we know, is a test that is of no significance to students, it is conceivable that the way in which their school communicates to them the weight of performing well might be important to their motivation and effort. The fact that the students are sampled at the school level also means that there is a dependency between students, for which reason it is correct also from a purely statistical standpoint to analyse variables divided into different levels.

However, preliminary multilevel analyses indicate that the vast majority of the variation (94%) is at the individual level.

Studies using the effort thermometer and the test-taking motivation scale could also be supplemented by studies on smaller scale. These might include observations or interviews with students participating in studies like PISA in order to gain a better understanding of how the tests as well as the self-report items are perceived by students, and how they reason when they respond to them.

Concluding remarks

There are a number of indications that the students' test-taking motivation and effort in the test situation represent an important factor to take into account and to continue to monitor. The reported effort of Swedish students has changed somewhat over time. We are relatively low in reported effort compared with countries that we usually compare ourselves with; reported effort and motivation have an impact on test performance and this impact persists even when other attitude/motivation variables are taken into account.

At the same time, the decline in reported effort is so small and the impact of effort also not particularly strong that it appears unlikely that this decline has contributed strongly to the weak performance of Swedish students in PISA 2012. According to the statistical analysis, effort could explain a few points of the deterioration in results (e.g. 4–5 points of 31 points in mathematics in PISA between 2003 and 2012), but the majority is probably sooner explained by other factors, such as students not having sufficient skills in domains measured by the PISA test.

References

Butler, J. C. (2008). *Interest and effort in large-scale assessment: the influence of student motivational variables on the validity of reading achievement outcomes*. PhD thesis, University of Melbourne, Faculty of Education.

Eklöf, H. (2006). *Motivational beliefs in the TIMSS 2003 context: Theory, measurement and relation to test performance*. Doctoral Dissertation, Umeå University, Department of Educational Measurement.

Eklöf, H., Japelj, B., & Grønmo, L. S. (2014). A cross-national comparison of reported effort and mathematics performance in TIMSS Advanced 2008. *Applied Measurement in Education, 27,* 31–45.

Eklöf, H., & Knekta, E. (2014). *Different stakes, different motivation? Studies of test-taking motivation in different assessment contexts*. Paper presented at the annual meeting of the American Educational Research Association (AERA), Philadelphia, April 2014 (In symposium: The Impact of Test-Taking Motivation and Test Consequences on the Validity of Test Score Inferences).

Hattie, J.A. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

King, G., Murray, C., Salomon, J., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1), 191–207.

OECD (2013a). *PISA 2012 results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science* (Volume I). Paris: OECD.

OECD (2013b). *PISA 2012 results: Ready to Learn: Student Engagement, Attitudes and Motivation* (Volume III). Paris: OECD.

OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World* (Volume 1). Paris: OECD.

Pintrich, P. R., & Schunk, D. H. (2002). Motivation in education: Theory, research and applications (2nd Ed.). New Jersey, NJ: Merril Prentice Hall.

Swedish National Agency for Education. (2013). *PISA 2012. 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap*. Stockholm: Swedish National Agency for Education

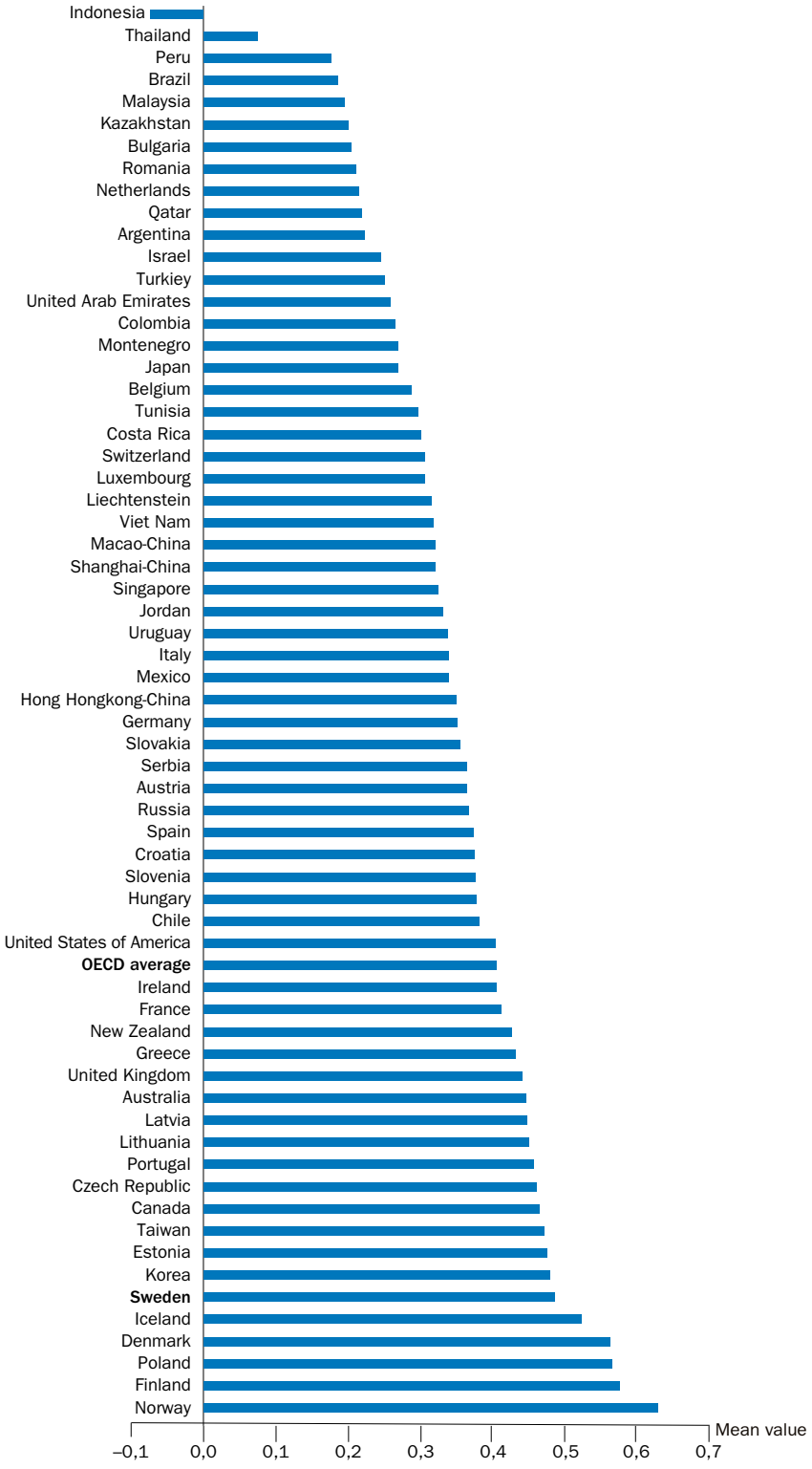
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1–17. doi:10.1207/s15326977ea1001_1

Appendix

A. Correlations between the scale measuring mathematics self-concept and mathematics performance

Figure A1. Correlations between the scale measuring self-concept in PISA and mathematics performance, all countries.



B. Regression analysis – effects of reported effort and motivation on test performance

Table B1. Effect of reported effort and motivation on test performance in PISA 2003, 2006 and 2012

Test	Effect on test score per increment on effort thermometer	Standardi-sed effect	Decline reported effort	Decline explained by effort (points)	Decline performance PISA (points)
	B	β	2003–2012/ 2006–2012	2003–2012/ 2006–2012	2003–2012/ 2006–2012
Ma 2003	13.65	.29			
Ma 2006	12.59	.28			
Ma 2012	12.20	.28	-0.35/ -0.34/	4,5/ 4,2/	31/ 24/
Läs 2003	13.80	.29			
Läs 2006	15.72	.32			
Läs 2012	15.18	.31	-0.35/ -0.34/	5/ 5,2/	31/ 24/
NO 2003	13.26	.25			
NO 2006	13.10	.28			
NO 2012	13.28	.29	-0.35/ -0.34/	4,6/ 4,5/	21/ 18/

	B Test-taking motivation scale	β
Ma 2012	26.08	.20
Läs 2012	29.27	.20
NO 2012	27.32	.19

C. Multiple regression analysis – effects of reported effort on test performance when other variables are also included in the model

All variables were entered into a regression model as independent variables, and the five plausible values in mathematics were entered as dependent variables. How much each variable contributes to explaining the results after the other variables have been taken into account is shown in Table C1, which presents the results from the multiple regression for the sample of 13 countries, the samples that answered student questionnaire version B. Note that here, use has been made of the ready indices available in the database; PISA has transformed these into a mean value of 0 (the OECD average) and a standard deviation of 1. For this reason, the effort thermometer has also been rescaled to have a mean value of 0 and a standard deviation of 1.

Table C1. Effects (b + beta as well as standard error and t values) of a number of variables on mathematics performance in PISA 2012. A sample of countries.

Land	EqVar	b	beta	b.se	beta.se	b.t	beta.t
Australia	(CONSTANT)	497.78					
	Mathematics anxiety	-8.67	-0.09	2.35	0.02	-3.69	-3.74
	ESCS (SES)	23.63	0.20	1.80	0.02	13.10	13.15
	Attribution	0.80	0.01	1.86	0.02	0.43	0.43
	Instrumental motivation	-1.39	-0.01	2.10	0.02	-0.66	-0.67
	Intrinsic motivation	-11.46	-0.13	2.36	0.03	-4.86	-4.95
	Self-efficacy	39.23	0.44	2.08	0.02	18.83	19.13
	Perseverance	-5.53	-0.06	1.96	0.02	-2.82	-2.78
	Self-concept	21.26	0.21	3.09	0.03	6.89	6.93
	Effort	15.60	0.16	1.49	0.02	10.49	9.75
Denmark	(CONSTANT)	493.76					
	Mathematics anxiety	-8.08	-0.10	3.53	0.04	-2.29	-2.30
	ESCS (SES)	22.60	0.23	2.27	0.02	9.98	10.29
	Attribution	1.18	0.01	2.39	0.03	0.49	0.49
	Instrumental motivation	-0.85	-0.01	2.14	0.02	-0.40	-0.40
	Intrinsic motivation	-12.49	-0.15	2.32	0.03	-5.39	-5.28
	Self-efficacy	26.22	0.29	3.01	0.03	8.72	8.57
	Perseverance	-4.57	-0.05	2.56	0.03	-1.78	-1.78
	Self-concept	29.81	0.37	3.54	0.04	8.42	8.21
	Effort	18.71	0.21	2.18	0.02	8.59	9,01
Estonia	(CONSTANT)	520.01					
	Mathematics anxiety	-18.53	-0.22	3.78	0.04	-4.90	-4.99
	ESCS (SES)	18.79	0.20	2.32	0.02	8.10	8.15
	Attribution	4.48	0.05	3.45	0.03	1.30	1.32
	Instrumental motivation	-2.69	-0.03	2.68	0.03	-1.00	-1.00
	Intrinsic motivation	-12.46	-0.14	3.23	0.04	-3.86	-3.86
	Self-efficacy	30.08	0.33	2.77	0.03	10.84	11.28
	Perseverance	-5.44	-0.06	1.91	0.02	-2.86	-2.88
	Self-concept	21.33	0.24	4.02	0.05	5.30	5.23
	Effort	4.98	0.06	2.47	0.03	2.02	2.00

Land	EqVar	b	beta	b.se	beta.se	b.t	beta.t
Finland	(CONSTANT)						
	Mathematics anxiety	-7.23	-0.08	2.74	0.03	-2.63	-2.64
	ESCS (SES)	14.43	0.14	2.06	0.02	7.01	7.17
	Attribution	1.25	0.01	1.55	0.02	0.80	0.80
	Instrumental motivation	0.33	0.00	2.39	0.03	0.14	0.14
	Intrinsic motivation	-12.67	-0.15	2.47	0.03	-5.14	-5.16
	Self-efficacy	19.60	0.22	2.81	0.03	6.98	7.07
	Perseverance	1.28	0.01	2.09	0.02	0.61	0.61
	Self-concept	33.50	0.43	2.35	0.03	14.27	13.93
	Effort	18.09	0.18	1.96	0.02	9.23	9.45
Germany	(CONSTANT)						
	Mathematics anxiety	-18.44	-0.22	3.32	0.04	-5.56	-5.55
	ESCS (SES)	25.77	0.26	2.24	0.02	11.53	12.26
	Attribution	12.52	0.13	2.77	0.03	4.53	4.47
	Instrumental motivation	-5.74	-0.07	3.09	0.04	-1.86	-1.86
	Intrinsic motivation	-3.85	-0.05	4.04	0.05	-0.95	-0.95
	Self-efficacy	38.04	0.39	3.10	0.03	12.26	13.48
	Perseverance	-4.53	-0.05	3.11	0.03	-1.45	-1.44
	Self-concept	10.39	0.13	4.49	0.05	2.32	2.30
	Effort	11.21	0.12	2.25	0.02	4.99	4.85
Iceland	(CONSTANT)						
	Mathematics anxiety	-10.94	-0.12	3.28	0.04	-3.34	-3.33
	ESCS (SES)	10.60	0.10	2.86	0.02	3.71	3.81
	Attribution	3.48	0.04	2.35	0.03	1.48	1.48
	Instrumental motivation	7.34	0.08	2.87	0.03	2.56	2.55
	Intrinsic motivation	-11.94	-0.13	3.36	0.04	-3.56	-3.59
	Self-efficacy	23.76	0.29	3.11	0.04	7.64	7.67
	Perseverance	-5.09	-0.06	2.92	0.03	-1.74	-1.75
	Self-concept	27.50	0.32	4.07	0.05	6.76	6.84
	Effort	19.76	0.24	2.28	0.03	8.65	8.20
Latvia	(CONSTANT)						
	Mathematics anxiety	-14.00	-0.13	4.47	0.04	-3.13	-3.19
	ESCS (SES)	22.89	0.25	2.59	0.03	8.85	9.59
	Attribution	0.16	0.00	3.32	0.03	0.05	0.05
	Instrumental motivation	-8.25	-0.09	3.03	0.03	-2.72	-2.74
	Intrinsic motivation	-10.64	-0.11	4.08	0.04	-2.61	-2.63
	Self-efficacy	24.71	0.25	4.69	0.05	5.27	5.45
	Perseverance	-1.71	-0.02	2.57	0.03	-0.67	-0.67
	Self-concept	28.69	0.29	4.93	0.05	5.82	5.76
	Effort	20,71	0.22	2.93	0.03	7.07	6.82

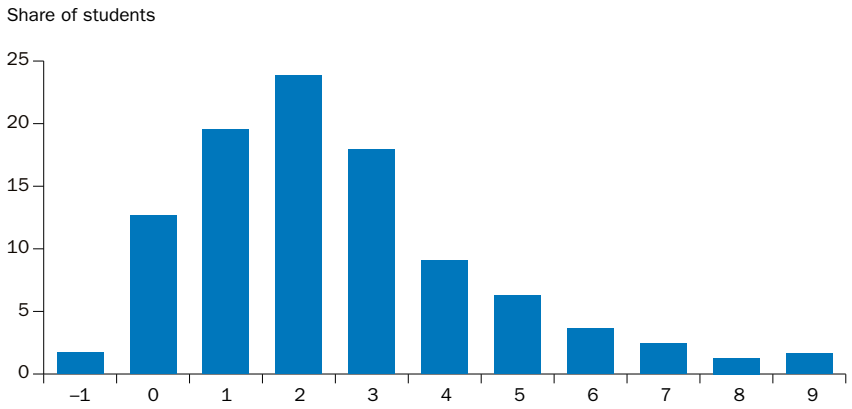
Land	EqVar	b	beta	b.se	beta.se	b.t	beta.t
Lithuania	(CONSTANT)						
	Mathematics anxiety	-18.65	-0.21	2.49	0.03	-7.49	-7.54
	ESCS (SES)	23.84	0.25	2.31	0.02	10.33	10.61
	Attribution	-0.01	0.00	2.02	0.02	-0.01	0.00
	Instrumental motivation	0.67	0.01	1.92	0.02	0.35	0.35
	Intrinsic motivation	-17.36	-0.21	2.37	0.03	-7.33	-7.61
	Self-efficacy	31.19	0.34	2.83	0.03	11.02	12.75
	Perseverance	-5.89	-0.06	2.92	0.03	-2.01	-1.98
	Self-concept	21.94	0.25	3.26	0.04	6.73	6.77
	Effort	12.67	0.12	2.58	0.02	4.92	5.07
Netherlands	(CONSTANT)	529.20		3.63		145.94	
	Mathematics anxiety	-12.75	-0.13	3.71	0.04	-3.44	-3.44
	ESCS (SES)	28.59	0.27	3.39	0.03	8.44	8.86
	Attribution	-1.94	-0.02	3.89	0.04	-0.50	-0.50
	Instrumental motivation	0.59	0.01	3.51	0.04	0.17	0.17
	Intrinsic motivation	-4.81	-0.05	4.46	0.05	-1.08	-1.09
	Self-efficacy	38.17	0.41	3.81	0.03	10.03	11.81
	Perseverance	0.88	0.01	2.98	0.03	0.29	0,30
	Self-concept	-8,32	-0.09	3.88	0,04	-2.14	-2.17
	Effort	8,46	0,09	2.43	0.03	3.48	3,46
Norway	(CONSTANT)	491.71		2.79		176.52	
	Mathematics anxiety	-9.85	-0.11	3.47	0.04	-2.84	-2.86
	ESCS (SES)	15.80	0.13	3.13	0.02	5.04	5.12
	Attribution	1.81	0.02	2.68	0.03	0.67	0.67
	Instrumental motivation	0.37	0.00	2.27	0.02	0.16	0.16
	Intrinsic motivation	-10.16	-0.12	3.06	0.03	-3.32	-3.32
	Self-efficacy	25.01	0.31	3.04	0.04	8.22	8.69
	Perseverance	-1.29	-0.02	2.49	0.03	-0.52	-0.52
	Self-concept	32.97	0.40	3.10	0.04	10,63	10,53
	Effort	15,73	0.19	1.94	0.03	8.11	7.48
Poland	(CONSTANT)	520.89		2.86		181.92	
	Mathematics anxiety	-19.20	-0.20	3.07	0.03	-6.25	-6.29
	ESCS (SES)	19.80	0.20	2.38	0.02	8.31	9.06
	Attribution	-2.30	-0.03	2.00	0.02	-1.15	-1,14
	Instrumental motivation	1.84	0.02	3.33	0.03	0.55	0.56
	Intrinsic motivation	-15.67	-0.16	3.83	0.04	-4.10	-4.14
	Self-efficacy	34.09	0.38	2.64	0.03	12.93	14.76
	Perseverance	-1.92	-0.02	1.96	0.02	-0.98	-0.98
	Self-concept	23.08	0.25	3.37	0.04	6.85	6.73
	Effort	0.87	0.01	1.78	0.02	0.49	0.49

Land	EqVar	b	beta	b.se	beta.se	b.t	beta.t
Sweden	(CONSTANT)	476.11		2.66		179.03	
	Mathematics anxiety	-17.28	-0.18	2.77	0.03	-6.23	-6.29
	ESCS (SES)	18.57	0.16	3.13	0.03	5.93	6.11
	Attribution	-0.99	-0.01	2.86	0.03	-0,34	-0.35
	Instrumental motivation	-7.91	-0.09	2.77	0.03	-2.85	-2.88
	Intrinsic motivation	-8.55	-0.10	3.04	0.04	-2.82	-2.79
	Self-efficacy	22.70	0.25	3.65	0.04	6.21	6.37
	Perseverance	0.20	0.00	2.40	0.03	0.08	0.09
	Self-concept	29.31	0.32	3.97	0.04	7.39	7.17
	Effort	16.03	0.19	1.79	0.02	8.95	8.56

D. Analyses of the difference variable (the difference between reported effort in PISA and estimated effort if the test result was going to count towards a grade)

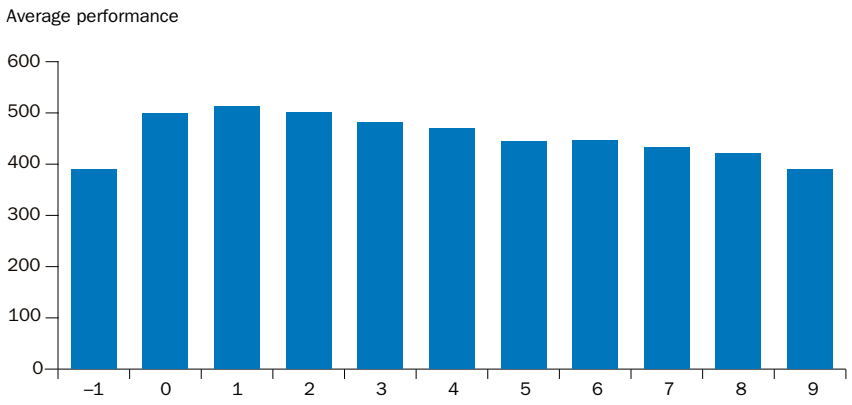
The figures, which cover Swedish data for PISA 2012, merge all categories of "unrealistic raters" (Butler, 2008: students who reported higher effort in PISA compared with if the test was going to count towards their grade) into one category (-1) as these categories contained very few students. The group with "2" on the difference scale is the students who have a difference of two increments for reported effort and effort if the test was going to count towards their grade (typically, an eight in the first and a ten in the second). Those belonging to the category of "9" in the figure below are the students who have a difference of nine increments: a one in effort in PISA, but a ten if the test was going to count towards their grade. This group also contains very few students.

Figure D1. Share of students with each value of the difference variable (-1 = all students reporting more effort in PISA than in graded tests, 0 = students reporting the same effort in PISA as in a normal test, 1 = students reporting one increment's less effort in PISA than in graded tests ..... 9 = students reporting minimum effort in PISA, but maximum effort if it was going to count towards their grade).



Students who marked that they had invested one increment's less effort in PISA than they would have invested if the test was going to count towards their grade are the highest performing group in PISA, followed by students who have a value of 0 or 2 in the figure above (see Figure D2 below).

Figure D2. Average performance per "increment" on the difference scale, Sweden, Mathematics, PISA 2012



STUDY II

Have Swedish students' perseverance and engagement changed over time

A study of student response patterns in PISA's knowledge test

Summary

Sweden has participated in PISA since the first survey in 2000. PISA looks at the knowledge of 15-year-olds in reading, mathematics and science, and some surveys also look at various forms of problem-solving ability.[4] In the first rounds, Swedish students performed at a relatively high level in all the areas assessed, but in the latest study Sweden instead has a position below the average for the OECD countries. One hypothesis regarding this decline in results is that it is the students' engagement in taking the PISA knowledge test which has decreased, rather than a deterioration in their knowledge. This study investigates whether parts of the results decline in PISA can be said to be due to students in the latest survey not having had perseverance and engagement in answering the test items to the same extent as previously.

This in-depth analysis does not claim to provide an exhaustive answer to the question of whether it is a decline in the engagement or the knowledge level of students that has led to a deterioration in PISA results. However, it does represent one piece in the puzzle of student motivation and effort. The analysis gives no indication that the results decline in PISA is not a fair picture of the students' knowledge in the areas PISA measures.

This analysis examines whether there appears to have been a change in how Swedish students answer the items in the latest knowledge test in PISA 2012 compared with previous studies, with a focus on mathematics and reading. Inspired by an Irish model, PISA's design of rotating test items is used to compare student response patterns. These are analysed on the basis of two possible hypotheses. The first hypothesis says that the results decline in PISA is due to a decreasing level of knowledge in students, while the second hypothesis says that the results decline is due to decreased student engagement in taking the knowledge test. Both these scenarios can occur simultaneously and even reinforce each other, and it can be difficult to separate the two completely. In order to nevertheless attempt an examination of which of these two explanations is most likely, this report presents several ways of illustrating student response patterns. In this analysis, changes in student response patterns cannot be directly linked to results in the PISA test. It is therefore not possible to give an answer as to how much of the decline in results, expressed in terms of the number or percentage of points, can be explained by a lack of engagement on the part of students.

The idea behind the first hypothesis – a decreasing level of knowledge – is that students find it more difficult to answer the items correctly to the same extent as previously, regardless of whether the items come early or late in the test. This means that no change in the students' perseverance takes place between the years. Perseverance here means whether students become more tired towards the end of the test and the test time. Diminishing perseverance means that the share of students answering items correctly towards the end of a test is lower than the share of students answering them correctly at the beginning of the same test. Similarly, the share of students leaving a particular item

4 Problem solving was tested in PISA 2003 and digital problem solving was tested in PISA 2012

unanswered is expected to be lower for those answering that item early on, as opposed to those answering it late in the test.

The second hypothesis – decreased engagement in taking the knowledge test – is instead based on the idea that students initially answer the test items they can, just as students with the same level of knowledge have done in previous years, but to a higher degree than previously do not invest effort all the way through. The students manage equally well at the beginning of the test, which suggests that they possess the same knowledge. However, at the end of the test, they perform lower than the corresponding students in previous years, which is an indication that they give up when it gets difficult, tedious or when they get tired. This means that there is a decrease in the students' perseverance between the years.

The Swedish students' response patterns suggest that it is the level of knowledge that has decreased during the 2000s. This conclusion is based on the fact that the patterns supporting this hypothesis are unequivocal; there is a clear decline in the share of students answering the items correctly in both mathematics and reading, regardless of whether the items come early or late in the test booklet, between most of the PISA surveys compared. Changes in response patterns among the share of students leaving items unanswered also support this hypothesis. These changes, however, are usually somewhat smaller and somewhat more difficult to assess.

There are also some signs to suggest that test-taking engagement among Swedish students may have decreased somewhat. These results are not unequivocal, however. Although it is not possible to return an answer expressed in terms of the number or percentage of points of the results decline, the overall assessment is that this potential decrease in engagement would not be able to explain any major part of the results decline demonstrated by Swedish students over the past decade.

Introduction

Sweden has participated in the OECD project – PISA (Programme for International Student Assessment) – since the first survey in 2000. PISA looks at the knowledge of 15-year-olds in reading, mathematics and science, and some surveys also look at various forms of problem-solving ability.[5]

The latest PISA survey, PISA 2012, confirms and reinforces the picture previously demonstrated by PISA and that is confirmed by other international knowledge surveys such as PIRLS and TIMSS. The skills of Swedish compulsory school students in reading, mathematics and science have deteriorated since the turn of the millennium.[6] Overall, Sweden is the country that has seen the greatest decline in results of all countries participating in PISA (National Agency for Education 2013a). In order to reverse this development, an increased understanding of the possible reasons behind the results decline is essential. As already pointed out in previous reports by the National Agency for Education, it is not possible to single out any isolated explanatory factor for the development that has taken place with regard to the results of Swedish students in recent decades. There are many factors that come into play, and it is difficult to establish a causal relationship between individual changes and this development of results (National Agency for Education 2013b).

The National Agency for Education report on PISA 2012 (National Agency for Education 2013a) discusses possible hypotheses regarding what might be behind the deterioration in results, and it emphasises that further analysis is needed before it is possible to gain a clearer picture of why this development is as it is. One aspect discussed concerns the reliability of the PISA results. This is based on the fact that the students taking the PISA test are anonymous and that the students do not receive any feedback on their results. The test cannot therefore be used as a basis for grades or assessment. This type of test is often called a low-stakes test. This is in contrast to a high-stakes test, which is of significance to the student.

PISA's knowledge tests are taken during the spring semester, at which time most 15-year-olds are in year nine of compulsory school. This coincides with a period when these students are also taking several national tests. Unlike the PISA test, these national tests in year nine can affect the student's final grade.

In Sweden, the number of national tests has increased from three in year nine in 2003 to four in 2012.[7] On the one hand, it could be argued that more national tests may have led to students becoming more accustomed to taking tests and that the students' test anxiety should be less in the PISA test than other tests because it does not affect their grades. This could contribute to a performance that better reflects the students' knowledge. On the other hand, it

5 Problem solving was tested in PISA 2003 and digital problem solving was tested in PISA 2012

6 With the exception of science in TIMSS Year 4 between the surveys in 2007 and 2011

7 In 2003, year nine students took national tests in Swedish, English and mathematics. When PISA 2012 was conducted, a national test in the science subjects had been introduced (with every student taking a test in one of the subjects biology, physics or chemistry). In 2013, national tests were also introduced in the social science subjects. In the same way as for the science subjects, every student takes a test in one of the social science subjects (geography, history, religious education or social studies). Thus, there are now five national tests for year nine students, but when PISA 2012 was conducted, the students took four national tests.

could instead be argued that an increased number of tests during this period has made the students tired of taking tests. Since, for the individual student, it does not make any difference whether they invest any effort or not, it can lead to students not doing their best in the PISA test. The students' performance would then be a poorer reflection than previously of the students' knowledge and thus lower the relevance of the results as a means of measuring knowledge. However, if this explanation were to be correct, it is probably only a partial explanation for the declining results.

PISA 2009 established that Ireland was one of the countries that had dropped most in reading of all participating countries, just as Sweden had done in mathematics in PISA 2012. In an article from Ireland written by Cosgrove and Cartwright (2014), the authors examine two potential, and possibly overlapping, explanations for the development of Irish results based on student response patterns in PISA. The basic concept of this in-depth study is based on these two hypotheses, but the approach has been further developed.

Aim

This study examines whether there appears to have been a change in how the students answer the items in the latest knowledge test in PISA 2012 compared with previous studies, with a focus on mathematics and reading. It is also examined whether such possible changes mainly signal a decline in knowledge or whether the results decline might at least partly be attributed to the students' lack of engagement in taking the PISA knowledge test. The aim is to investigate whether the students' engagement in taking the PISA knowledge test appears to have changed between the years and, by extension, whether the Swedish students' PISA results are reliable. This analysis is undertaken in order to understand more about student perseverance and engagement in the test situations that the international large-scale studies participated in by Swedish students are based on.

Two hypotheses form the basis for the analysis of student response patterns. The first hypothesis is that the results decline in PISA is due to a decreasing level of knowledge in students. The second hypothesis is that the results decline is due to decreased student engagement in taking the knowledge test. Both these phenomena can occur simultaneously and even reinforce each other. It can also be difficult to separate the two completely. In order to nevertheless examine which of these two explanations is most likely, we present several ways of illustrating student response patterns. In this analysis, changes in student response patterns cannot be directly linked to results in the PISA test. It is therefore not possible to give an answer as to how much of the decline in results, expressed in terms of the number or percentage of points, can be explained by a lack of engagement on the part of students.

The next section presents a brief review of PISA's design, with a focus on the knowledge test and the characteristics of it used in our analysis, as well as an overview of the analysis concept, design and method. We then present and analyse the results. The report concludes with a summary discussion.

Implementation

This study analyses student response patterns based on the share of students answering the items correctly and the share of them leaving the items unanswered, depending on when in the test the item comes. We investigate whether these patterns change between the years. Since mathematics is the main subject in PISA 2012, we will focus on mathematics items, but also supplement this with reading items, as we want to be able to draw as broad conclusions as possible and not only focus on a specific subject area. The mathematics items are assessed to be more reading-intensive than the mathematics items which Swedish students are generally accustomed to (National Agency for Education 2009 and 2009). A student's reading ability may thus have an effect on a student's propensity to answer mathematics items. It is therefore of interest to supplement the mathematics items particularly with reading items. We have chosen not to include science in this analysis.

Another question raised when discussing the PISA knowledge test is how the items are designed and how accustomed Swedish students are to this type of item. One aspect of the item design is the way in which students are expected to give their answers. We therefore also examine how student response patterns change when we divide the items into the two item types of multiple-choice items and items for which the students themselves formulate the answer.

This in-depth study focuses on changes in answers among Swedish students over time. In order to capture any general trends and gain an idea of what might be said to be deviations, we have used average response shares from the students in participating OECD countries as reference points.[8] [9]Sometimes, these are referred to in abbreviated form as the OECD average. This is not to be seen as an analysis of the OECD countries, but is an aid in interpreting the answers of Swedish students. This analysis makes no comparisons with individual countries. In this section, we give first a brief introduction to the design of the knowledge test that students take in PISA and how this design makes our analysis possible. This is followed by a discussion of the delimitations we have made. The section concludes with a closer review of the two different approaches used to analyse student response patterns.

8 The countries contributing to the reference values may differ from cycle to cycle as some countries only participate in certain cycles.

9 The Swedish students' response shares are weighted using Total Student Weight. The response shares for the OECD countries are weighted using Senate Weight. The latter causes all countries to contribute equally to the average. This means that countries with a very large number of 15-year-olds do not contribute more to the OECD average than countries with fewer 15-year-olds.

The design of PISA and the knowledge test

There are primarily three competency areas tested in PISA: mathematics, science and reading. Each time the study is conducted, that is in each survey, one subject is the main subject and the other two are minor subjects. Each competency area is the main subject in every third survey, and thus recurs as the main subject every nine years. Table 1 presents a list of the years in which PISA has been conducted and the subject that was the main subject in each survey. All documentation about PISA, how PISA is implemented and databases from all PISA surveys are available at the OECD website.[10]

Table 1. Main subject in PISA's knowledge test in each survey.

Year/Survey	Main subject in the knowledge test
2000	Reading
2003	Mathematics
2006	Science
2009	Reading
2012	Mathematics

In PISA, each student only answers a sample of all the test items, meaning that the items in different students' test booklets differ. Every PISA survey has both new test items and items from at least one previous survey. The test items that recur in more than one survey are called trend items.

All test items are divided into blocks. A test item belongs to only one block, and each block contains about 12–14 test item. A block consists only of items within the same subject area. When a subject is the main subject, there are usually 7 blocks containing items on that subject. The number of blocks per minor subject varies from survey to survey by 2–4 blocks.

There are 13 different test booklets in each survey.[11] A test booklet contains 4 blocks, and each block can thus be placed in four different ways in a test booklet. This means that each block comes in one of four different positions in the test booklet, see Table 2. Each test booklet consists of blocks containing items from at least two different subject areas. Since PISA 2003, a block comes once in each position, and two different blocks do not occur together in more than one test booklet. This is called a balanced rotation design.

10 http://www.oecd.org/pisa/pisaproducts/

11 There are also test booklets with items that are a little easier, but these are not used in Sweden. In the 2012 survey, some students took the test on a computer. From the 2015 survey onwards, all students will perform the items on a computer. This analysis only uses student responses from printed test booklets.

Table 2. Example of block position in the different test booklets. The example is taken from PISA 2012.

The mathematics blocks are in bold type when they come in position 1 and in position 4.

Test booklet number	Block			
	Position 1	Position 2	Position 3	Position 4
B1	**PM5**	PS3	PM6A	PS2
B2	PS3	PR3	PM7A	PR2
B3	PR3	PM6A	PS1	**PM3**
B4	**PM6A**	PM7A	PR1	**PM4**
B5	**PM7A**	PS1	PM1	**PM5**
B6	**PM1**	PM2	PR2	**PM6A**
B7	**PM2**	PS2	PM3	**PM7A**
B8	PS2	PR2	PM4	PS1
B9	PR2	PM3	PM5	PR1
B10	**PM3**	PM4	PS3	**PM1**
B11	**PM4**	PM5	PR3	**PM2**
B12	PS1	PR1	PM2	PS3
B13	PR1	PM1	PS2	PR3

P = Test given on paper, M = Mathematics, R = Reading, S = Science

All items are divided into **blocks**. An item belongs to only one block, and each block contains about 12–14 items. All items in a block belong to the same subject area.

There are 13 different **test booklets** in each PISA survey. A test booklet contains 4 blocks. Each test booklet consists of blocks containing items from at least two subject areas.

A block can be placed in four different ways in a test booklet. The blocks thus have four **positions** in the test booklets.

Since PISA 2003, each block comes once in each position, and two different blocks do not occur together in more than one test booklet. This is called a **balanced rotation design**.

The analysis – concept and hypotheses

This study makes use of the fact that the rotation design allows us to observe how different students answer the same item both when it comes early and when it comes late in a test booklet, both within one year and between two years. This means that we can investigate whether the students' behaviour when taking the test has changed between the years when PISA has been conducted and whether such a change gives support to the hypothesis that their engagement in taking the PISA knowledge test has decreased over the years, or whether the pattern instead suggests that it is the level of knowledge that has decreased.

Which response patterns support each hypothesis?

Table 3a presents the response patterns that support each hypothesis, and Table 3b presents the response pattern that demonstrates diminishing student perseverance within one survey. This analysis covers student response patterns for correct answers and unanswered items. Some items in the knowledge test

yield one or two points for partially correct or completely correct answers. These shares are combined and categorised as a correct answer. The items left unanswered by students may either be ones that are skipped in the middle of a test booklet or ones that the student does not reach at the end of the test. These two categories of unanswered items are analysed together.

Table 3a. Response patterns according to the two hypotheses.

Changes in response patterns that lend support to the hypothesis of a decreasing level of knowledge and of decreased engagement in taking the PISA knowledge test.

		Change between year 1 and year 2	
		P1	P4
Decreasing level of knowledge	Share of correct answers	Decreasing	Decreasing
	Share unanswered	Increasing	Increasing
Decreasing test-taking engagement	Share of correct answers	Unchanged	Decreasing
	Share unanswered	Unchanged	Increasing

Table 3b. Response patterns for diminishing perseverance within one survey.

Response patterns that demonstrate a diminishing student perseverance during the test situation. If the students' perseverance decreases between the years, this suggests that the students' engagement has decreased.

		Change between P1 and P4	
		Within year 1	Within year 2
Diminishing perseverance in the test situation	Share of correct answers	Decreasing	Decreasing
	Share unanswered	Increasing	Increasing

The following pages illustrate the concept of comparing response patterns and the three combinations of the two hypotheses which are possible. This is done by means of Figures 1–3. The example used is that of the average share of students answering the items correctly. In detail, this means that we compare student response patterns for items in position 1 (the first items in the test booklet) and the corresponding items in position 4 (the last items in the test booklet) in two ways: (1) for each position between two surveys and (2) for the difference between these two positions within one survey. The dots mark the average share of students answering the items correctly by year and position. The dashed line illustrates the comparison between the two positions within each year, that is, whether the students' perseverance diminishes in the relevant survey. The difference between two dots in the same position illustrates the difference between two surveys. A combined assessment of the two types of changes is made in order to assess whether there is support for either of the hypotheses.

What is meant by the students' perseverance is the effect that arises when they might begin to lose focus and become tired towards the end of extended periods of concentration and effort. What we expect in any test situation at all is that the share of students answering items correctly towards the end of a test is lower than the share of students answering them correctly at the beginning of the same test. Similarly, the share of students leaving a particular item unan-

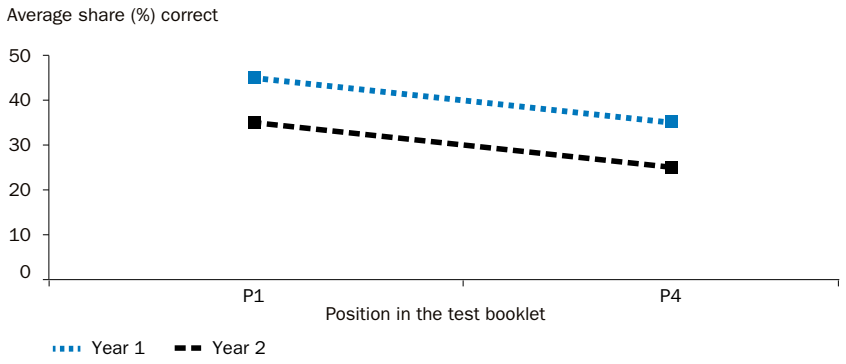
swered is expected to be lower for those answering that item early, as opposed to those answering it late.

Response pattern that lends support to the hypothesis of a decreasing level of knowledge

Figure 1 shows an example of what students' average shares of correct answers in position 1 and position 4 for two different survey rounds might look like when it is the level of knowledge in students that has led to the decline in results.

Figure 1. Generally decreasing level of knowledge.

Example of a possible response pattern for average shares of students who answered the items correctly for the two positions in two different PISA surveys. The lines illustrate the comparison between the two positions within each year, that is, that the students' perseverance diminishes towards the end of the test.



The idea behind the hypothesis of a decreasing level of knowledge is that students find it more difficult to answer the items correctly to the same extent as previously, regardless of whether the items come early or late in the test. They either try to answer the item, but answer correctly to a lesser extent even early in the test, or do not bother giving an answer to more items than students in previous surveys.

This means that there has not been any change in perseverance between the years, as demonstrated by the two dashed lines being parallel, but that the share of students answering items correctly decreases between the years, and that the share of students leaving items unanswered increases, both in position 1 and position 4.

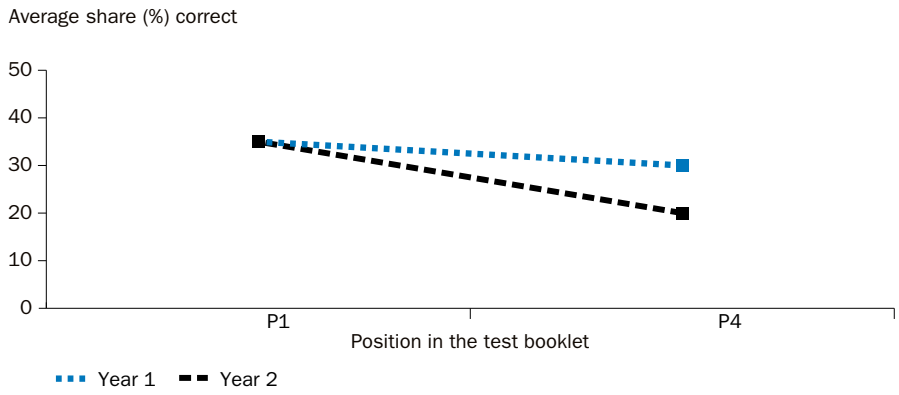
This response pattern could also arise if a higher proportion of students in a later survey bothers significantly less about trying to answer the items, but that all maintain an even, low level and do not become less interested or stop answering the items the longer the test continues. Such a scenario is deemed less likely, and whether they in that case bother less because they have less knowledge and choose not to invest effort or whether they bother less for some other reason is something this analysis cannot answer.

Response pattern that lends support to the hypothesis of decreased engagement

The hypothesis of decreased engagement in taking the knowledge test is instead based on the idea that the students in a survey initially take the test as in previous years and answer the items they actually can, but to a greater extent than previously do not invest effort all the way through. This response pattern is illustrated by Figure 2. The students manage equally well at the beginning of the test as students have done in previous surveys, which shows that they have the same actual ability. However, at the end of the test, they perform lower than compared to corresponding students in previous years, which is an indication that they give up when it gets difficult, tedious or when they get tired.

Figure 2. Decreased engagement.

Example of a possible response pattern for average shares of students who answered the items correctly for the two positions in two different PISA surveys. The lines illustrate the comparison between the two positions within each year, that is, that the students' perseverance diminishes towards the end of the test.



In order for there to be support for this hypothesis, we can expect the share of students answering items correctly and the share of students leaving items unanswered to be stable in position 1, while the share of students answering items correctly decreases and the share of students leaving items unanswered increases in position 4 between the years. The change in the perseverance of students is shown by the dashed lines in Figure 2 no longer being parallel, but instead showing an increasing "gap". This means that perseverance between the two years is to be considered significantly different from each other.

This study cannot answer whether a possible decreasing engagement has to do with the test not meaning or contributing anything to the students, i.e. the low-stakes character of the test, whether it is a general attitude to academic work, and success is less important to students today, or whether it is based on something else.

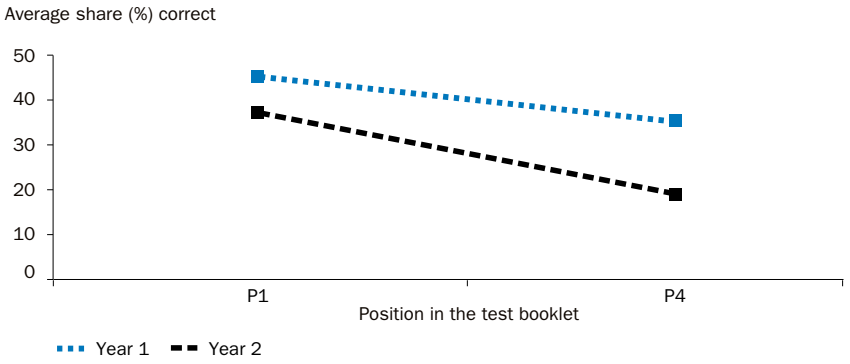
Decreasing level of knowledge and decreased engagement can occur simultaneously

These two scenarios can occur simultaneously. They can also reinforce each other. Such a situation is illustrated in Figure 3. The figure shows that there is a difference between the years both in position 1 and in position 4, but with a lower perseverance in year two. This is manifested by the difference between the

average shares of students answering the item correctly in position 4 between the years being greater than the corresponding difference in position 1.

Figure 3. Decreasing level of knowledge combined with decreasing engagement.

Example of a possible response pattern for average shares of students who answered the items correctly for the two positions in two different PISA surveys. The lines illustrate the comparison between the two positions within each year, that is, that the students' perseverance diminishes towards the end of the test.



A lower perseverance in a later survey might be due to less motivated and engaged students not having the energy or not bothering to get through the entire test with the same extent of focus, but also that they might become comparatively more tired if the test is experienced as being more difficult than it was in previous years because of poorer knowledge. This could then affect the student's opportunities to answer items correctly throughout the test, or it could be reflected in more items being left unanswered towards the end. Similarly, those who have a lower engagement when taking the test might choose to leave more difficult items rather than answering them correctly to a lesser extent, perhaps from the very beginning of the test.

It might therefore be difficult to separate the two hypotheses completely, and interpretations and conclusions must be drawn with a certain caution. It is nevertheless reasonable for some differences in student response patterns to be attributable to each hypothesis. We present several ways of illustrating the hypotheses to increase our opportunities to draw conclusions about the students' changed behaviour.

Multiple-choice items or open-ended items – does it make a difference?

In PISA, the items can be categorised according to the kind of answer the student is expected to give in response to the item. We call this item type. All analyses made of student response patterns will also be made by item type in order to gain further information about the students' response patterns and their possible engagement in taking the PISA knowledge test.

The precise designations or definitions of the various item types differ somewhat between surveys, but two categories of item type are consistently used for all years. The first category contains questions where students choose between fixed response options, known as multiple-choice questions. The second category contains items where students are expected to formulate their own answers, with or without explanation or reasoning for how they arrived at their answer. This latter group of item is called open-ended items.

This part of the analysis thus attempts to say something further about student engagement based on any differences in response patterns between these item types. Our hypothesis is that students find the threshold for answering a multiple-choice item with predefined response options lower than the threshold for beginning to formulate an answer on their own. The correct answer to an open-ended item therefore signals that the student actually possesses the knowledge that the item is designed to test to a higher degree than the correct answer to a multiple-choice item does. Therefore, we might also expect there to be a certain difference in the students' response rates depending on item type.

In order for there to be support for the hypothesis that students do not have the same engagement, and therefore do not invest the same effort, in taking the knowledge test in recent years' surveys, we can expect students to answer open-ended items correctly in later surveys to a lower degree than multiple-choice items. There is always a chance of students guessing the right answer to a test item, but this chance is less for an open-ended item than for a multiple-choice item. If students know the correct answer to a lower degree than previously, and instead guess more on multiple-choice items, the share of students answering the multiple-choice items correctly will also decrease, but to a somewhat lesser extent than the corresponding share for open-ended items.

If comparatively more open-ended items than multiple-choice items are left unanswered by students in more recent studies, this may indicate that they are no longer investing effort to as great an extent as previously. In that case, this means that the share of students leaving open-ended items unanswered should increase over time. We therefore particularly compare whether there are differences in student perseverance between the two item types between two surveys.

Delimitations and selection of surveys for comparison

We have chosen to prioritise subjects for which the same test items as in PISA 2012 also occur in a survey as far back in time as possible, because we want to make a comparison with a round of PISA when the Swedish result was significantly better than it is today. There must also be trend items to compare between the surveys, or two surveys with the same main subject.

Main focus on student response patterns for mathematics items

A review of the test items occurring in more than one survey, and which are thus what we call trend items, has been conducted for the years 2003–2012. PISA 2000 has not been included in this compilation because that study did not use a balanced design.[12]

This report primarily examines mathematics items because mathematics was the main subject in PISA 2012. Student response patterns in mathematics in the 2012 survey are compared with student response patterns both in PISA 2003 and in PISA 2006. Since mathematics was the main subject both in 2003 and in 2012, there are more mathematics items overall in these studies than in others, and it is possible to base a comparison between these two years on more information than for the years when only trend items can be compared.

12 In PISA 2000, reading was the main area, which means that reading has also been the main area in two PISA surveys. However, a balanced design was not used in PISA 2000 and comparisons of position effects are therefore not possible.

We also compare only trend items between 2003 and 2012, although these items are spread across different blocks in the two surveys. Finally, we compare trend items in mathematics between the 2006 and 2012 surveys. These are the same trend items that are included in the 2003 survey, but in the 2006 and 2012 surveys, they are together in the same block.

Supplementary analysis of student response patterns for reading items

We have also chosen to analyse response patterns for reading items because we are not primarily interested in a subject-specific result. The mathematics items in PISA are assessed to be more reading-intensive than the mathematics items to which Swedish students are accustomed. Reading comprehension may therefore have an effect on the propensity to answer mathematics items in PISA. The analysis of reading items is based solely on trend items.

The comparison of reading items makes use of the surveys in 2012, 2009 and 2003. In reading, a new set of items became trend items due to this subject being the main subject in the 2009 survey. This means that for reading in PISA 2012, there are only trend items from 2009, and that another set of items are the same between the 2009 and 2003 surveys. Comparisons of student response patterns for trend items in reading are therefore made partly between the 2012 and 2009 surveys and partly between 2009 and 2003. Note that there are no opportunities to compare trend items directly between 2012 and 2003 since none of the test items are the same between these years.[13]

For science, a large part of the items was replaced in 2006. There are few trend items from PISA 2003 in later studies, and those that remain are not together in the same block between the years. Because of this, and because the mathematics items in PISA have been assessed as being more reading- intensive than those with which Swedish students are accustomed, we have chosen to limit ourselves to looking at items in mathematics and reading, but not in science.

A **general comparison** of student answers to **mathematics** items between **2012** and **2003** is made. In both these surveys, mathematics is the main subject, which means that there are more mathematics items to analyse.

Only **trend items** in **mathematics** are also compared between **2012** and **2003**. However, these items are not together in the same block in both surveys, which means that this analysis is unable to take into account the order of the items.

The same **trend items** in **mathematics** are also compared between **2012** and **2006**. In these two surveys, the trend items are together in the same block in both surveys.

In reading, a new set of items became trend items due to this subject being the main subject in the 2009 survey. For this reason, we compare **trend items** in **reading** between **2012** and **2009**.

In **reading**, the first set of **trend items** is also compared between **2009** and **2003** because we want to make a comparison with surveys as far back in time as possible. Note that there are no opportunities to compare trend items directly between 2012 and 2003 since none of the test items are the same between these years.

13 The blocks containing reading items in the 2009 survey are therefore divided into two groups, one that is a comparison group for the students' answers in reading in 2012, and one that is a comparison group for the students' answers to these items in 2003.

Methods used: two different approaches

The analysis of student response patterns makes use of two different approaches. Both approaches examine how the students' response shares for the equivalent or same test items have changed on average over time, but they have different starting points. The first approach is more general and is based on a general picture of student response patterns. It might therefore contain more information and be less sensitive to deviations in individual items. The second approach is more narrow and is a complement to the general approach. It is based on direct comparisons between the students' response shares for individual items, and can therefore only be based on trend items. Taken together, they provide a good picture of how student response patterns change within the test booklets and between studies.

General comparison of all test items within a subject area

The more general analysis is based on a linear regression analysis. This approach requires a larger quantity of test items and it is therefore best suited to years in which a subject is the main subject. This approach can thus only be used for the comparison between mathematics items in 2003 and in 2012, for which we have access to more items than for other years and for other subjects. The approach is general in that it does not require all test items for the two years to be the same, but is able to take into account comparisons of two test items that are considered equivalent. The analysis handles this by taking into account the items position within the block (if it comes in first, second or third position, and so on, called sequence) and the items' level of difficulty.[14] Besides this, just over 40 per cent of the mathematics items in PISA 2012 are trend items from the 2003 survey, which further enhances comparability between the years.

Analysis of trend items within a subject area

The second approach in the analysis is based only on the items that are trend items. This approach makes a direct comparison of the students' response shares for a particular item on two different occasions or between two positions in the test booklets on the same occasion. There are fewer trend items than the total number of items in a subject when it is the main subject, which means that the comparison might be more influenced by the students' answers to individual items than the general analysis. On the other hand, the analysis is based solely on the actual differences in the share of students who dealt with the same item in a certain way at different times or in different positions.

When the trend items occur together in the same block on both occasions, it means that the items are in the same position in the test booklet, and the students encounter the same items just before and just after the item, two factors that increase comparability.

However, the trend items in mathematics in the 2003 study are spread out over several different blocks, which means that the item's position in the test booklet and the surrounding items cannot then be taken into account. The comparison must therefore be interpreted more cautiously than that of trend

14 Information regarding how the items' level of difficulty is calculated can be found in the PISA Technical Report (OECD 2014).

items between 2006 and 2012. Since the general analysis takes into account the level of difficulty and the position in the test booklet for the mathematics comparison between these years, the two approaches together provide a good picture of student response patterns.

In the **general analysis** performed by means of regression analysis, we examine *differences in the students' average response shares* between years and positions in the test booklet. If we want to analyse student perseverance in a particular year based on the share of students answering the items correctly, we analyse the regression coefficient from a variable that divides the students' response shares according to whether the item had been in position 1 (coded as group 0) or in position 4 (coded as group 1). This coefficient yields the difference between the two groups and is interpreted as "The difference between the share of students answering the items correctly when they are in position 1 compared with in position 4 is on average XX percentage points in the relevant year". If this difference is not statistically significant, we are thus unable to prove that there is a difference between the students' average response shares by position within a year.

In the **analysis of trend items**, we instead examine the average of the differences in the students' response shares between years and positions in the test booklet. For the example of perseverance in a particular year based on the share of students answering the items correctly, we first calculate the difference between the share of students who answered the respective trend item when the item is in position 1 and in position 4 in the relevant year. We subtract the respective share in position 4 from position 1. When we have these differences, their mean value is calculated. If this mean value is negative, we are able to establish that there has, on average, been a decline in the share of students answering the item correctly when the item comes last in the test booklet compared with when the item comes first in the test booklet. We test whether the difference is statistically significant by means of a paired t-test. If this difference is not statistically significant, we are thus unable to prove that there is, on average, a difference between the students' response shares according to position within a year.

Results

The sections below present a compilation of the response patterns that emerge in the analysis. The results chapter is divided into three sections. First are two sections reporting student response patterns by the subject areas of mathematics and reading. Each of these subsections concludes with a summary of student response patterns within the subject area, and the hypothesis that the response patterns support. This is followed by a section reporting the analysis of student response patterns according to item type.

All differences reported in this analysis are such differences as are statistically significant and thus cannot be asserted to have arisen solely from the natural variation found in sample surveys.[15]

Student response patterns for mathematics items

The following section presents student response patterns for mathematics items. We first present the comparison of PISA 2003 and PISA 2012, followed by the comparison of PISA 2006 and PISA 2012.

Overall, the general analysis of all mathematics items for these years, and the analysis of the trend items alone, provide a consistent picture of the response patterns and changes examined. The results from the analyses of mathematics items are found in Table 4 and Table 5 in the appendix.

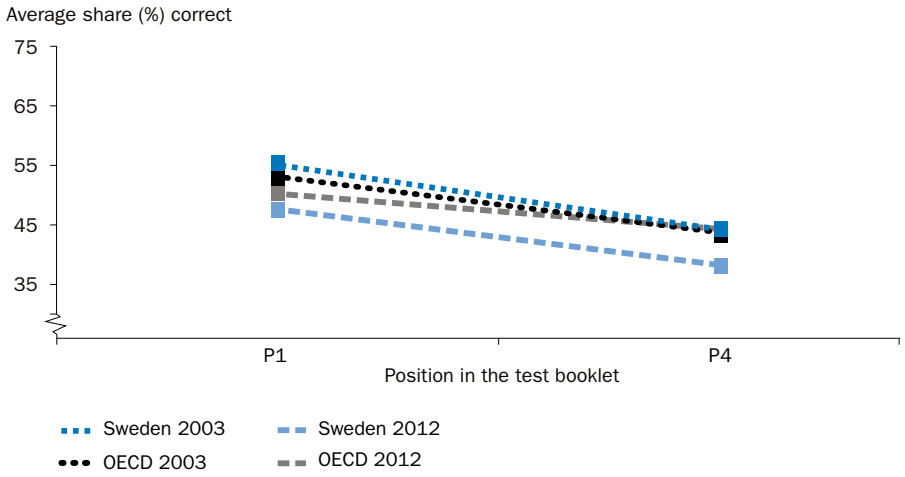
Student response patterns for correctly answered mathematics items in PISA 2003 and PISA 2012

The share of Swedish students answering mathematics items correctly has decreased between 2003 and 2012 in both positions, according to both methods of analysis. Figure 4 presents these response patterns based on the values from the general analysis. The figure shows that the two dashed lines illustrating student perseverance are in principle parallel for the Swedish students when we examine the share of students answering the items correctly. They are similar to the scenario described in Figure 1 regarding support for the hypothesis of a generally decreasing level of knowledge.

15 A significance level of 5 per cent is used. Technical expression meaning that we claim that the mean is different from zero even after we have taken statistical uncertainty into account.

Figure 4. Results from the general analysis of mathematics items in PISA 2003 and PISA 2012: Correctly answered items.

Average share of students answering mathematics items correctly when the items come first and last in the test booklet in the respective survey.



According to the general analysis, the share of students answering the items correctly when the items come first in the test booklet has decreased by about 7 percentage points between 2003 and 2012. The corresponding change when the test items instead come last in the test booklet, i.e. in position 4, is for Sweden a decline of about 5 percentage points. The decline in the share of Swedish students answering the items correctly between position 1 and position 4 according to the two different methods of analysis is around 11–12 percentage points in PISA 2003, while the same share in PISA 2012 is about 9–10 percentage points. These percentage points are not significantly different from each other, which means that student perseverance has not changed between these years.

According to the general analysis, the share of students answering the items correctly in the OECD on average has decreased when the items come first in the test booklet, but is unchanged when the items come last in the test booklet. This means that perseverance has improved on average in the OECD between PISA 2003 and PISA 2012. The decline in the share of students answering the items correctly between 2003 and 2012 is therefore greater in Sweden than in the OECD on average.

Student response patterns for unanswered mathematics items in PISA 2003 and PISA 2012

The response pattern for the share of students leaving mathematics items unanswered is relatively equal between PISA 2003 and PISA 2012 in Sweden. This is also seen in Figure 5, which shows student response patterns for skipped items for 2003 and 2012 according to the general analysis. According to the general analysis, no differences in student response patterns for skipped items are statistically significant. Perseverance, i.e. the change between when an item is answered early in a test booklet compared with late in a test booklet is about 13–14 percentage points according to this analysis.

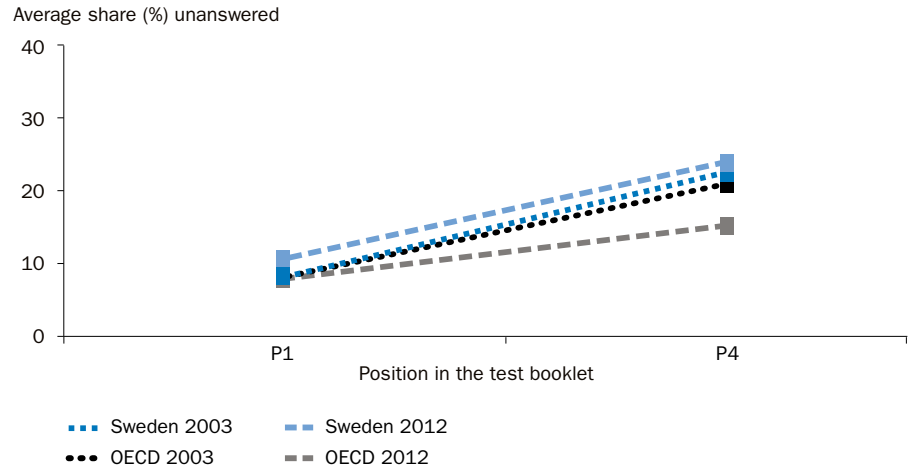
The analysis of trend items demonstrates a small increase in the share of students skipping items when they come first in the test booklet, but a decline

when they come last in the test booklet. This means that perseverance according to this analysis has increased somewhat from PISA 2003 to PISA 2012.

Figure 5 reveals that perseverance based on unanswered mathematics items in PISA 2012 is less among Swedish students than for the OECD countries on average.

Figure 5. Results from the general analysis of mathematics items in PISA 2003 and PISA 2012: Unanswered items.

Average share of students skipping mathematics items when the items come first and last in the test booklet in the respective survey.

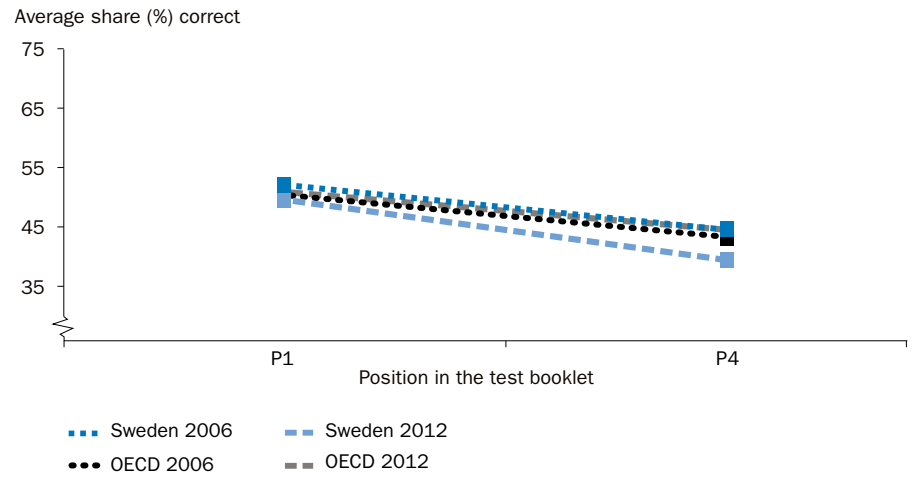


Student response patterns for correctly answered trend items in mathematics in PISA 2006 and PISA 2012

Student response patterns for mathematics items in PISA 2012 are also compared with PISA 2006. This comparison is made only for trend items, which are fewer in number than the mathematics items used in the general analysis of response patterns between 2003 and 2012. Figure 6 presents student response patterns for correct answers.

Figure 6. Results from the analysis of trend items in mathematics in PISA 2006 and PISA 2012: Correctly answered items.

Average share of students answering mathematics items correctly when the items come first and last in the test booklet in the respective survey.



When we compare trend items in mathematics between PISA 2006 and PISA 2012, we can see a decline in the students' share of correctly answered items in both positions. Perseverance according to correctly answered items appears to be just as extensive in these two studies, and these measures are not statistically different from each other. This pattern is similar to what we see in the comparison between PISA 2003 and PISA 2012, thus lending support to the hypothesis that it is the level of knowledge that has decreased prior to the 2012 survey.

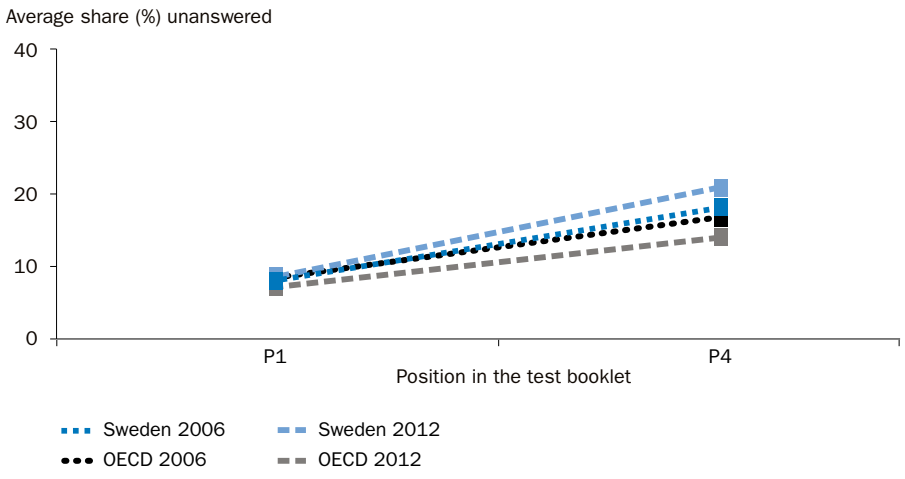
For position 1, the share of students answering the items correctly has decreased by about 3 percentage points, and the corresponding share in position 4 is about 5 percentage points. The difference between position 1 and position 4 in the relevant survey for both these years is around 8–10 percentage points. The Swedish development deviates somewhat from the average development in the OECD countries, where it can hardly be said that any change has taken place in the share of students answering the items correctly between these two years.

Student response patterns for unanswered trend items in mathematics in PISA 2006 and PISA 2012

There is a small difference when comparing the share of students leaving items unanswered in PISA 2006 and PISA 2012, both when the items come first in the test booklet and when they come last in the test booklet. There is a small increase in the share of students leaving items unanswered in these two years. However, when the items come early in the test booklet, the increase between the years is not statistically significant, while the corresponding share when the items come last is about 3 percentage points. Such a pattern could lend support to the hypothesis of decreasing engagement. However, since perseverance is unchanged between the years, the overall picture provides no such support. Instead, there is weak support for the hypothesis of a decreasing level of knowledge.

Figure 7. Results from the analysis of trend items in mathematics in PISA 2006 and PISA 2012: Unanswered items.

Average share of students skipping mathematics items when the items come first and last in the test booklet in the respective survey.



The change in the share of students leaving items unanswered in position 1 compared with position 4 is therefore about as great in Sweden in 2012 as in 2006 (around 10–12 percentage points), as shown by the two dashed lines closely following each other in Figure 7. However, perseverance in the 2006 survey appears to be greater than in the 2003 survey, as shown in Table 5.

The average of the share of students leaving items unanswered for the OECD as a whole appears instead to have decreased somewhat between these years. These years also appear to indicate that perseverance based on unanswered mathematics items is somewhat less among Swedish students than for the OECD countries on average.

Summary of student response patterns for mathematics items

The Swedish students' response patterns for mathematics items support the hypothesis that it is the level of knowledge that has decreased prior to PISA 2012.

The analysis has shown that the share of students answering mathematics items correctly decreased both when these items come first and when they come last in the test booklet between both PISA 2003 and 2012 and between PISA 2006 and 2012. It also gives indications that the share of students skipping items increases between PISA 2006 and PISA 2012, and that perseverance measured as the share of students leaving items unanswered is unchanged between these surveys. These changes are somewhat less than those for the share of students answering items correctly.

Student response patterns for trend items in reading

The following section presents student response patterns for reading items. We first present the comparison of PISA 2009 and PISA 2012, followed by the comparison of PISA 2003 and PISA 2009.[16] All differences regarding student response patterns for reading items are presented in Table 6 and Table 7 in the appendix.

Student response patterns for correctly answered trend items in reading in PISA 2009 and PISA 2012

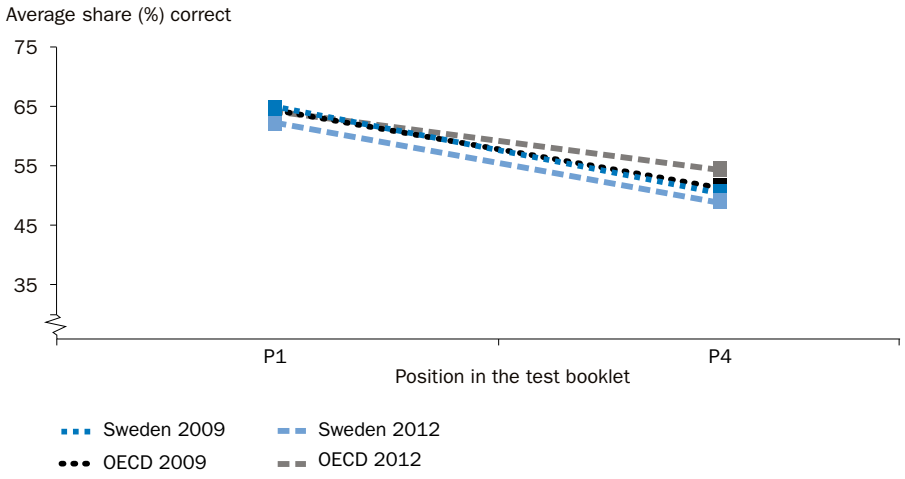
Figure 8 presents student response patterns for the share of students answering reading items correctly in PISA 2009 and PISA 2012. There is no difference in perseverance measured in the average share of students answering reading items correctly between these two years. In the 2012 survey, the difference between the positions in the share of students answering items correctly is about 13 percentage points, and the corresponding figure for the 2009 survey is about 14.5 percentage points.

The share of Swedish students answering these items correctly is on average about 2–3 percentage points lower in the 2012 survey than in 2009, regardless of position. Overall, this lends support to the hypothesis of a decreasing level of knowledge.

The changes between 2009 and 2012 in the Swedish students' response patterns for the share of correct reading items appear to deviate somewhat from the pattern seen for the average of the students in the OECD. There, we see instead an increase in the share of correctly answered items of about 3 percentage points when the items come last in the test booklet and no change when the items come first in the test booklet. This means that perseverance expressed as the share of students answering reading items correctly is somewhat greater in the OECD on average in 2012 than in 2009. It suggests that neither the level of knowledge nor engagement has decreased in the OECD on average.

Figure 8. Results from the analysis of trend items in reading in PISA 2009 and PISA 2012: Correctly answered items.

Average share of students answering trend items in reading correctly when the items come first and last in the test booklet in the respective survey.



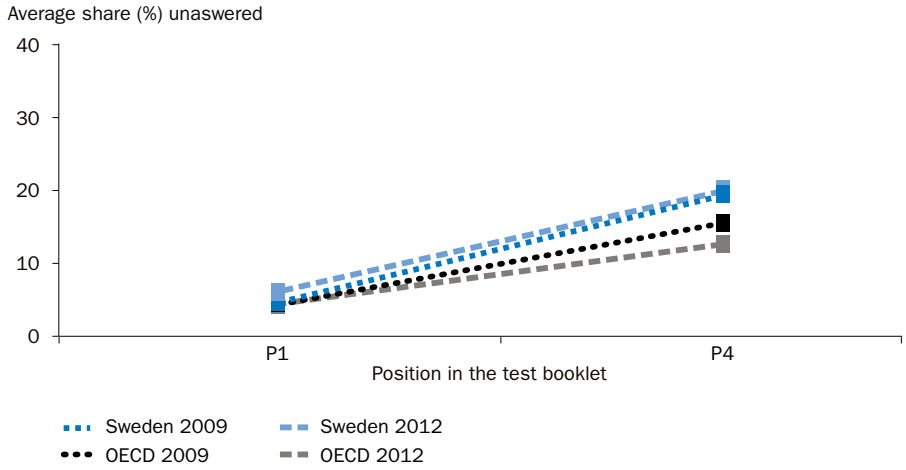
16 Bear in mind that the reading items for 2009 are divided into two groups depending on which year we are comparing with since all trend items in reading are different between 2006 and 2012.

Student response patterns for unanswered trend items in reading in PISA 2009 and PISA 2012

The share of Swedish students leaving reading items unanswered increases somewhat from 2009 to 2012, but primarily in position 1. These response patterns are presented in Figure 9. These changes too are small, only around one percentage point. The perseverance of Swedish students expressed as the change in the share of students leaving reading items unanswered is about 14–15 percentage points in both surveys, thus similar to the share of students answering the items correctly. Overall, this lends some support to the hypothesis of a decreasing level of knowledge. The perseverance of Swedish students deviates somewhat from perseverance in the OECD on average, which appears to be somewhat greater, and also increases somewhat between the 2009 survey and the 2012 survey.

Figure 9. Results from the analysis of trend items in reading in PISA 2009 and PISA 2012: Unanswered items.

Average share of students skipping trend items in reading when the items come first and last in the test booklet in the respective survey.



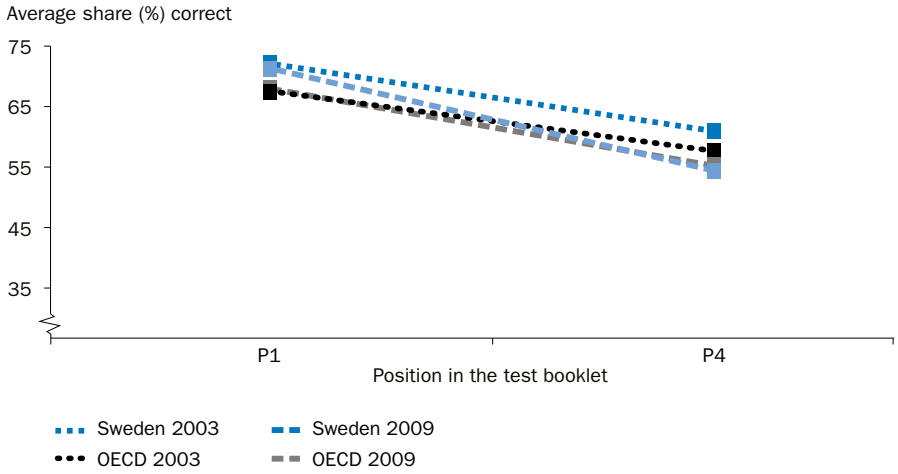
Student response patterns for correctly answered trend items in reading in PISA 2003 and PISA 2009

The change in student response patterns for correctly answered reading items between these surveys is a little different to that in the comparison between 2012 and 2009, and also when compared with the analyses of mathematics items. In Figure 10, the dashed lines indicate a decreasing perseverance between the 2003 survey and the 2009 survey, but from the same level for the average share of students answering the items correctly at the beginning of the test booklet. That is, the share of students answering items correctly is unchanged when the reading items come first in the test booklet, but this share decreases when the items come last. This pattern is similar to that presented in Figure 2, namely the scenario of a decreasing engagement in taking the knowledge test. The difference between position 1 and position 4 is about 11 percentage points in the 2003 survey and about 17 percentage points in the 2009 survey.

A similar pattern is also seen for the OECD average, but with a somewhat smaller decline in student perseverance prior to the 2009 survey.

Figure 10. Results from the analysis of trend items in reading in PISA 2003 and PISA 2009: Correctly answered items.

Average share of students answering trend items in reading correctly when the items come first and last in the test booklet in the respective survey.



Student response patterns for unanswered reading items in PISA 2003 and PISA 2009

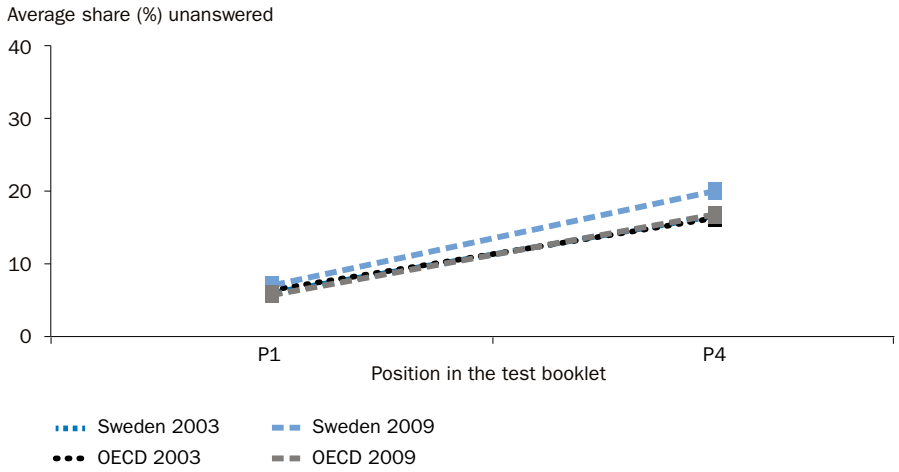
The share of students leaving reading items unanswered appears to increase somewhat between PISA 2003 and PISA 2009, and here the differences are statistically significant both when the items come early and late in the test booklet. These response patterns are presented in Figure 11. These changes too are relatively small; an increase of around 1 percentage point when the items come first in the test booklet and around 3 percentage points when the items come last.

The perseverance of Swedish students expressed as the change in the share of students leaving reading items unanswered is about 11-13 percentage points in both surveys. This pattern lends support to the hypothesis of decreasing knowledge.

Overall, in the OECD on average, no changes appear to have taken place in student response patterns for the share of unanswered items between PISA 2003 and PISA in 2009.

Figure 11. Results from the analysis of trend items in reading in PISA 2003 and PISA 2009: Unanswered items.

Average share of students skipping trend items in reading when the items come first and last in the test booklet in the respective survey.



Summary of student response patterns for reading items

The Swedish students' response patterns for reading items mainly support the hypothesis of a decreasing level of knowledge in the later surveys. The share of students answering reading items correctly decreases between the surveys in PISA 2009 and 2012. These changes are in principle equal in size regardless of whether the items come first or last in the test booklet. Perseverance has thereby not changed between the years. The same pattern also emerges when looking at the share of students leaving items unanswered between PISA 2003 and PISA 2009, and also to some extent between PISA 2009 and PISA 2012. The average share of students leaving items unanswered increases somewhat regardless of whether the items come first or last in the test booklet, together with unchanged perseverance. However, these changes are somewhat smaller than the corresponding changes in the analysis of mathematics items. This pattern differs somewhat from what we see in the OECD on average, where the results instead indicate an increased level of knowledge or increased engagement between these two years.

Reading also reveals a partially different picture of student response patterns than that for mathematics items. The change in the share of students answering items correctly between PISA 2003 and 2009 indicates a decreasing engagement in taking the knowledge test between these years. Between these surveys, the share of students answering items correctly is unchanged when the reading items come first in the test booklet, whilst it decreases when the items come last. This supports a decrease in perseverance from PISA 2003 to 2009. However, it should be noted that a similar, though somewhat smaller, change also emerges in the OECD on average.

The picture is not unequivocal, however. The changes in the share of students leaving items unanswered do not follow this pattern. Perseverance according to this measure is unchanged between both PISA 2003 and 2009, and between PISA 2009 and 2012 among Swedish students. For the OECD average, perseverance is unchanged between 2003 and 2009, but increases somewhat between 2009 and 2012.

Student response patterns divided according to multiple-choice items and open-ended items

This supplementary analysis divides student response patterns according to whether the items are multiple-choice or open-ended items.[17] Our hypothesis is that students find the threshold for answering a multiple-choice item with predefined response options lower than the threshold for beginning to formulate an answer on their own.

In order for there to be support for the hypothesis that students do not have the same engagement, and therefore do not invest the same effort, in taking the knowledge test in recent years' surveys, we can expect students to answer open-ended items correctly in later surveys to a comparatively lower degree than multiple-choice items. Correspondingly, if comparatively more open-ended items than multiple-choice items are left unanswered by students in more recent studies, this may indicate that they are no longer investing effort to as great an extent as previously. This means that we are particularly looking at whether there are differences in student perseverance between the two item types between two surveys. The results of this analysis are presented in Tables 8–11 in the appendix.

Student response patterns for correctly answered items are not changed by item type

The general analysis of mathematics items reveals that item type does not appear to make any difference to the share of students answering items correctly. When we compare all test items in mathematics between PISA 2003 and PISA 2012, we compare the average response shares both with and without consideration of the items' level of difficulty. A comparison of student response patterns without considering the item's level of difficulty shows a significant difference in the change in the share of students answering items correctly depending on item type. When we then also consider the items' level of difficulty and the order in which the item comes within the block, this difference disappears. This indicates that the item type in itself does not make a difference to whether students answer the items correctly. Instead, it appears, above all, to be the order in which the item comes in the test booklet and its level of difficulty that make a difference.[18]

17 Note that when we divide the test items into two groups, the statistical uncertainty in the comparisons of the students' average response shares might become greater since we now have fewer test items in each group.

18 When we use a regression model to investigate the correlation between the items' level of difficulty and their item type for mathematics items in PISA 2003 and PISA 2012, the item type does not appear to affect how difficult the item is experienced to be in PISA 2003, while a correlation of this kind is found in PISA 2012. In the latter survey, students find open-ended items more difficult than multiple-choice items.

The decline we reported for the share of Swedish students answering mathematics items correctly, which lends support to the hypothesis of a decreasing level of knowledge, does not appear to change when student response patterns are divided according to the two item types. Although the difference between the years appears to be somewhat more stable for open-ended items, perseverance does not differ between the item types within the same year for any of the three PISA surveys investigated.

Reading items reveal in principle no differences at all in the average share of students answering the items correctly between multiple-choice items and open-ended items. The only difference that emerges is that the decline in perseverance, according to the share of students answering the items correctly, as noted between PISA 2003 and PISA 2009 – thus lending some support to the hypothesis of decreasing engagement – is primarily due to a decline among open-ended items.

Item type makes some difference to the share of students leaving items unanswered

When we compare student perseverance for the share of students leaving items unanswered, this is generally lower for open-ended items than for multiple-choice items, both in Sweden and in the OECD on average. However, no such pattern is seen among skipped mathematics items in PISA 2003 and PISA 2006. This indicates that item type appears to make some difference to the share of students skipping items. In contrast, it does not generally appear to be the case that student perseverance decreases more for open-ended items than for multiple-choice items between the years, which does not lend support to the hypothesis of decreasing engagement.

However, there are two things about student response patterns for skipped mathematics items that we may comment on. The first concerns the comparison between PISA 2003 and PISA 2012. No differences emerge between multiple-choice items and open-ended items in the general analysis between the share of students leaving mathematics items unanswered when these come first in the test booklet. In contrast, comparing only trend items, students appear to skip open-ended items to a somewhat greater extent than multiple-choice items when these come last in the test booklet. This could suggest decreased student engagement in PISA 2012 compared with PISA 2003. However, the differences in perseverance between these surveys are not significantly different from each other, which means that the pattern that emerges lends no clear support to the hypothesis of decreasing engagement.[19]

The second thing that should be commented on concerns the comparison between PISA 2006 and PISA 2012. The Swedish students' perseverance

19 Here, it does make a difference that open-ended items have a higher degree of difficulty than multiple-choice items in PISA 2012, while no correlation of this kind is found in PISA 2003. See preceding footnote. Furthermore, the comparison of student response patterns for trend items in these years suggests that the difference is due to the share of students skipping multiple-choice items being lower than before rather than to their skipping open-ended items more than before. This indicates that the difference between the response shares for open-ended items and multiple-choice items concerns the items' level of difficulty in 2012 rather than item type. Overall, the changes do not appear to lend support to a decrease in engagement between 2003 and 2012.

decreases between these two surveys with respect to the share of students leaving open-ended items unanswered, but not for multiple-choice items. This means that this analysis lends some support to a possible decrease in student engagement between these years. It should be noted, however, that this change is not supported in the share of correctly answered items, and it is impossible to determine exclusively whether this is due to decreased engagement or to whether students with lower ability also get more tired than previously and therefore omit difficult items to a greater degree.

Summary of student response patterns by item type

When we divide student response patterns according to the two item types, the results do not generally appear to change. Overall, the differences that emerge lend no further support to the hypothesis that engagement in taking the PISA knowledge test has decreased in the later studies.

The results we reported for the share of Swedish students answering items correctly do not appear to change when student response patterns are divided according to the two item types, neither for mathematics items nor for reading items.

For the share of students skipping items, item type appears to make some difference. In contrast, it does not generally appear to be the case that student perseverance decreases more for open-ended items than for multiple-choice items between the years. Therefore the analysis of student response patterns for skipped items broken down by item type also does not generally lend any further support to a decrease in engagement between the years.

It should be noted that this analysis shows that student perseverance decreases for the share of students leaving open-ended mathematics items unanswered between PISA 2006 and PISA 2012, while the equivalent does not apply to multiple-choice items in mathematics. However, it is difficult to determine exclusively whether these changes are due to decreased engagement or to whether students with lower ability also get more tired than previously and therefore omit difficult items to a greater degree.

Summary discussion

This study investigates the response patterns of Swedish students in the PISA knowledge test and whether these have changed between the different PISA surveys. The aim is to investigate whether the students' engagement in taking the PISA knowledge test has changed between the years and, by extension, whether the Swedish students' PISA results are reliable.

This concluding chapter presents the overall picture that emerges of student response patterns and a compilation of the support for the two hypotheses tested in the study. The first hypothesis says that the results decline in PISA is due to a decreasing level of knowledge in students, while the second hypothesis says that the results decline is due to decreased student engagement in taking the knowledge test.

Changes in student perseverance between the years

For all the surveys examined in this analysis, we see that student perseverance diminishes as the test goes on. This is true both for Sweden and for the OECD average. This means that the share of students answering items correctly is higher if the items come early compared with late in the test, and conversely that the share of students leaving items unanswered is lower when the items come early compared with late.

When we compare perseverance in mathematics items between PISA 2003 and 2012, the general analysis shows this to be unchanged both according to the share of students answering items correctly and according to the share of students leaving items unanswered. However, according to the analysis of trend items alone, perseverance expressed as the share of students leaving items unanswered appears to have increased in PISA 2012. Perseverance in PISA 2006 is greater than in PISA 2003, both with regard to the difference in the share of students who answered the items correctly and those who left them unanswered. In contrast, perseverance in PISA 2006 does not differ from 2012 in any respect. Overall, perseverance appears to be lowest in PISA 2003 and highest in PISA 2006, while perseverance in PISA 2012 falls somewhere in between.

This result differs somewhat from the pattern among the OECD countries on average. There, perseverance in PISA 2012 is greater than in PISA 2003. Besides this, perseverance based on mathematics items in PISA 2012 appears to be lower among Swedish students than for the OECD countries on average.

Student perseverance in reading is generally unchanged in Sweden, both expressed as the share of students answering items correctly and as the share of students leaving items unanswered. The picture is not unequivocal, however. Perseverance as measured by the share of students answering items correctly decreases between PISA 2003 and 2009. This decline indicates that student engagement may have decreased between these years.

This deviation, with decreasing perseverance in the share of students answering reading items correctly between 2003 and 2009, also emerges for the OECD average.

For Sweden, however, student response patterns for reading deviate somewhat from that which is noted on average for all students in the OECD when we compare PISA 2009 with PISA 2012. For the OECD on average, persever-

ance increases between these years, while perseverance in reading is at a relatively stable level in Sweden between these years.

The analysis that divides student response patterns according to item type reveals no difference in student perseverance between open-ended items and multiple-choice items for the share of students answering items correctly in any of the surveys investigated.

Perseverance for the share of students leaving items unanswered is generally lower for open-ended items than for multiple-choice items, both in Sweden and in the OECD on average. However, no such pattern is seen among skipped mathematics items in PISA 2003 and PISA 2006. In contrast, it does not generally appear to be the case that student perseverance decreases more for open-ended items than for multiple-choice items between the years.

Clear support for the hypothesis of a decreasing level of knowledge in PISA

Support for the hypothesis of a decreasing level of knowledge is seen in many of the comparisons made, particularly when we consider the change in the share of students answering items correctly. For Sweden, a clear decline emerges in the share of students answering mathematics items correctly both when these items come first and when they come last in the test booklet between both PISA 2003 and 2012 and between PISA 2006 and 2012. The changes in the share of correct answers are equal in size, regardless of the position of the items. This lends clear support to the hypothesis of a decreasing level of knowledge in mathematics. The analysis also gives indications of an increase in the share of students skipping items between PISA 2006 and 2012. Although the rise between these years when the items come early in the test booklet is not statistically significant, perseverance is unchanged between the surveys. This provides additional, albeit somewhat weaker, support for this hypothesis.

If we look at reading items, the Swedish students' response patterns mainly support the hypothesis of a decreasing level of knowledge in the later surveys. The comparison between PISA 2009 and 2012 reveals that there has been a steady decline in the share of students answering items correctly, albeit to a somewhat lesser extent than in the comparison of mathematics items. The same pattern also emerges when looking at the share of students leaving items unanswered between PISA 2003 and PISA 2009, and also to some extent between PISA 2009 and PISA 2012. The average share of students leaving items unanswered increases somewhat, regardless of whether the items come first or last in the test booklet, together with unchanged perseverance. These results generally persist when we divide student response patterns according to item type.

Some support for the hypothesis of decreased test-taking engagement between PISA 2003 and PISA 2009

In the comparison of the share of students answering reading items correctly between PISA 2003 and 2009, we see a response pattern that lends support to the hypothesis of decreasing student engagement. Between these years, the share of students answering items correctly is unchanged when the reading items

come first in the test booklet, but decreases when the items come last, which thus means that perseverance decreased from PISA 2003 to 2009. However, it should be noted that a similar, though somewhat smaller, change also emerges in the OECD on average.

In the analysis that divides student response patterns into multiple-choice items and open-ended items, the Swedish students' perseverance appears to decrease, expressed as the share of students leaving open-ended items unanswered between PISA 2006 and PISA 2012. The equivalent is not the case for multiple-choice items. This means that this analysis lends some support to a possible decrease in student engagement between these years. However, it should be noted that it is difficult to determine exclusively whether these changes are due to decreased engagement or to whether students with lower ability also get more tired than previously and therefore to a greater degree omit items experienced as being more difficult.

Concluding remarks

The Swedish students' response patterns mainly suggest that it is the level of knowledge that has decreased during the 2000s. This conclusion is based on the fact that the patterns supporting this hypothesis are unequivocal; there is a clear decline in the share of students answering the items correctly in both mathematics and reading, regardless of whether the items come early or late in the test booklet, between most of the PISA surveys compared. Changes in response patterns among the share of students leaving items unanswered also support this hypothesis. These changes, however, are usually somewhat smaller and somewhat more difficult to assess.

However, it should not be overlooked that there are also certain patterns in the responses that support the hypothesis of decreasing test-taking engagement. This is particularly the case in the comparison of response patterns for reading items between PISA 2003 and 2009, based on the decline in the share of students answering items correctly. A circumstance that weakens this is that the share of students leaving reading items unanswered instead lends support to the hypothesis of a decreasing level of knowledge, with a small but steady increase regardless of whether the item comes early or late in the test booklet and with unchanged perseverance between the years.

The study lends no support to a decrease in perseverance between the surveys, suggesting that student engagement has hardly changed, and thus also not been able to contribute to declining results to any greater extent. In this analysis, changes in student response patterns cannot be directly linked to results in the PISA test. It is therefore not possible to estimate how great a part this represents of the results decline expressed in terms of the number or percentage of points. The clearest support for decreased perseverance shown by the analysis is found in correctly answered reading items. This result may mean that there is a particular drop in engagement for reading items, or that reading items are perceived as being more difficult than mathematics items when they come at the end of a test booklet. However, this is something the analysis cannot answer.

The Swedish students' perseverance generally appears to be less than for students in the OECD on average, which in turn suggests that Swedish students could generally be less engaged than the OECD average. However, since the

analysis of student response patterns mainly suggests that it is the students' level of knowledge that has decreased, this result is of course to be understood in the light of the fact that the students' level of knowledge and their perseverance can reinforce each other in such a way that the students' perseverance might be less when the students' level of knowledge is lower.

Overall, there is some support for decreased test-taking engagement in Swedish students. These results are not unequivocal, however. The overall assessment of the results is that this potential decrease in engagement would not be able to explain any major part of the results decline demonstrated by Swedish students over the past decade. However, as mentioned earlier, it is not possible on the basis of this analysis to estimate how great a part this represents of the results decline expressed in terms of the number or percentage of points.

This in-depth analysis does not claim to provide an exhaustive answer to the question of whether it is a decline in the engagement or the knowledge level of students that has led to a deterioration in PISA results. However, it does represent one piece in the puzzle of student motivation and effort. Further knowledge could be contributed by more studies of this and related areas, such as comparisons with similar countries or the addition of items in science, which is the main subject in PISA 2015. The analysis gives no indication that the results decline in PISA is not a fair picture of the students' knowledge in the areas PISA measures.

References

Cosgrove, J., Cartwright, F., (2014) Changes in achievement on PISA: the case of Ireland and implications for international assessment practice. *Large-scale Assessments in Education*, 2:2

OECD (2014) PISA 2012 *Technical Report*. OECD Publishing, Paris. National Agency for Education (2006) *Med fokus på matematik och naturvetenskap. En analys av skillnader och likheter mellan internationella jämförande studier och nationella*

kursplaner. Skolverkets aktuella analyser. Stockholm: National Agency for Education.

National Agency for Education (2013a) *PISA 2012: 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap.* Report 398. Stockholm: National Agency for Education

National Agency for Education (2013b) *PISA 2012: 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap. Resultaten i koncentrat.* Report 398. Stockholm: National Agency for Education

National Agency for Education (2013c). *Kommunernas resurstilldelning till grundskolor.* Report 391. Stockholm: National Agency for Education

National Agency for Education (2015). *Med fokus på matematik – Analys av samstämmighet mellan svenska styrdokument och den internationella studien PISA.* Stockholm: National Agency for Education

Appendix

Table 4. Estimates of student response patterns in mathematics, between positions between years.

Compilation of changes in student response patterns for mathematics items when the block is in position 1 (P1) and in position 4 (P4) in the test booklet for the 2003, 2006 and 2012 studies. Figures in red indicate that the changes support the hypothesis of decreasing knowledge. No changes in the table support the hypothesis of decreasing engagement.

		Change P1 between years		Change P4 between years	
Sweden		**2003/2012**	**2006/2012**	**2003/2012**	**2006/2012**
Share correct	General analysis, without control	**-7.5**		-5.9	
	General analysis, control†	**-6.9**		**-5.3**	
	Analysis of trend items	**-4.7**	**-2,6**	**-2.8**	**-5.0**
Share unanswered	General analysis, without control	2.5		1.4	
	General analysis, control†	2.2		1.1	
	Analysis of trend items	**1.5**	0,6	**-2.9**	**2.9**
OECD		**2003/2012**	**2006/2012**	**2003/2012**	**2006/2012**
Share correct	General analysis, without control	-2.9		0.6	
	General analysis, control†	**-2.3**		1.1	
	Analysis of trend items	-1.0	0,5	**2.7**	**1.2**
Share unanswered	General analysis, without control	-0.2		**-5.7**	
	General analysis, control†	-0.4		**-5.9**	
	Analysis of trend items	**-0.7**	**-1.3**	**-8.0**	**-2.9**

If the differences in share for each position between two years are significantly different from each other, these are marked in bold type.

† With control for sequence (items 1–12), the item's degree of difficulty and block.

Table 5. Estimates of the students' diminishing perseverance in mathematics items.

Compilation of changes in student response patterns for mathematics items between position 1 (P1) and position 4 (P4) within the same year, for the 2003, 2006 and 2012 studies. No changes in the table support the hypothesis of decreasing engagement.

		Change between P1 and P4 within the same year		
Sweden		**2003**	**2006**	**2012**
Share correct	General analysis, without control	**-11.0**		**-9.4**
	General analysis, control†	**-11.0**		**-9.4**
	Analysis of trend items	**-12.0[b]**	**-7.7[c]**	**-10.1**
Share unanswered	General analysis, without control	**14.4**		**13.3**
	General analysis. control†	**14.4**		**13.3**
	Analysis of trend items	**16.7[a, b]**	**10.0[c]**	**12.3[c]**
OECD		**2003**	**2006**	**2012**
Share correct	General analysis, without control	**-9.5**		-6.0
	General analysis. control†	**-9.5[a]**		**-6.0[c]**
	Analysis of trend items	**-10.2[a, b]**	**-7.1[c]**	**-6.4[c]**
Share unanswered	General analysis, without control	**12.8[a]**		**7.2[c]**
	General analysis, control†	**12.8[a]**		**7.2[c]**
	Analysis of trend items	**14.1[a, b]**	**8.4[c]**	**6.8[c]**

If the differences in share between position 1 and position 4 within a year (perseverance) are significantly different from each other, these are marked in bold type.

† With control for sequence (items 1–12), the item's degree of difficulty and block.
a. Different from 2012.
b. Different from 2006.
c. Different from 2003.

Table 6. Estimates of student response patterns in reading, between positions between years.

Compilation of changes in student response patterns for reading items when the block is in position 1 (P1) and in position 4 (P4) in the test booklet for the 2003, 2009 and 2012 studies. Figures in red indicate that the changes support the hypothesis of decreasing knowledge. Figures in blue indicate that the changes support the hypothesis of decreasing engagement.

Analysis of trend items	Change P1 between years		Change P4 between years	
Sweden	**2003/2009[a]**	**2009[b]/2012**	**2003/2009[a]**	**2009[b]/2012**
Share correct	-0.8	**-2.7**	**-6.6**	**-1.6**
Share unanswered	**1.2**	**1.5**	**3.2**	0.6
OECD	**2003/2009[a]**	**2009[b]/2012**	**2003/2009[a]**	**2009[b]/2012**
Share correct	0.4	-0.3	**-2.6**	**3.0**
Share unanswered	**-0.5**	0.1	0.3	**-3.0**

If the differences in share for each position between two years are significantly different from each other, these are marked in bold type.

a. Compare with 2003 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests.

b. Compare with 2012 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests.

Table 7. Estimates of the students' diminishing perseverance in reading items.

Compilation of changes in student response patterns for reading items between position 1 (P1) and position 4 (P4) within the same year, for the 2003, 2009 and 2012 studies. Figures in blue indicate that the changes support the hypothesis of decreasing engagement.

Analysis of trend items	Change between P1 and P4 within the same year			
Sweden	**2003**	**2009[a]**	**2009[b]**	**2012**
Share correct	**-11.3[e]**	**-17.1[f]**	**-14.5**	**-13.4**
Share unanswered	**11.1**	**13.1**	**14.7**	**13.8**
OECD	**2003**	**2009[a]**	**2009[b]**	**2012**
Share correct	**-10.0[e]**	**-13.0[f]**	**-13.1[d]**	**-9.8[e]**
Share unanswered	**9.9**	**10.6**	**11.2[d]**	**8.2[e]**

If the differences in share between position 1 and position 4 within a year (perseverance) are significantly different from each other, these are marked in bold type.

a. Compare with 2003 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests.

b. Compare with 2012 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests.

d. Perseverance different from 2012.

e. Perseverance different from 2009.

f. Perseverance different from 2003.

Table 8. Estimates of student response patterns in mathematics, between positions between years, by item type.

Compilation of changes in student response patterns for mathematics items by item type when the block is in position 1 (P1) and in position 4 (P4) in the test booklet for the 2003, 2006 and 2012 studies. Figures in red indicate that the changes support the hypothesis of decreasing knowledge. Figures in blue indicate that the changes support the hypothesis of decreasing engagement.

		Difference in change between multiple-choice items and open-ended items							
		Change P1 between years				Change P4 between years			
		2003/2012		2006/2012		2003/2012		2006/2012	
Sweden		**Multiple-choice items**	**Open-ended items**	**Multiple-choice items**	**Open-ended items**	**Multiple-choice items**	**Open-ended items**	**Multiple-choice items**	**Open-ended items**
Share correct	General analysis, without control	2.1*	**-13.9***			4.6*	**-12.9***		
	General analysis, control†	**-6.5**	**-7.1**			**-4.0**	**-6.1**		
	Analysis of trend items	**-3.6**	**-5.7**	-0.9*	**-4.1***	-0.2*	**-5.3***	**-3.2**	**-6.8**
Share unanswered	General analysis, without control	0.8	**4.4**			-3.9*	**5.8***		
	General analysis, control†	3.2	2.4			-1.4	**3.7**		
	Analysis of trend items	**1.0**	**1.9**	0.6	0.6	**-5.2***	-0.6*	-0.1*	**5.7***
OECD		2003/2012		2006/2012		2003/2012		2006/2012	
Share correct	General analysis, without control	5.6*	**-8.6***			9.2*	-5.3*		
	General analysis, control†	**-2.5**	**-2.1**			1.0	1.3		
	Analysis of trend items	-0.9	-1.1	0.4	0.6	**3.2**	**2.3**	**1.8**	0.5
Share unanswered	General analysis, without control	-0.5	0.9			**-7.1**	**-3.7**		
	General analysis, control†	1.4	-0.8			**-5.2**	**-5.4**		
	Analysis of trend items	**-0.3**	-1.2	**-0.3***	**-2.3***	**-7.8**	**-8.3**	**-2.8**	**-3.0**

If the differences in share between multiple-choice items and open-ended items for each position between two years are significantly different from each other, these are marked with *.

If the differences in share for each position between two years are significantly different from each other, these are marked in bold type.

† With control for sequence (items 1–12), the item's degree of difficulty and block.

Table 9. Estimates of the students' diminishing perseverance in mathematics items, by item type.

Compilation of changes in student response patterns for mathematics items by item type between position 1 (P1) and position 4 (P4) within the same year, for the 2003, 2006 and 2012 studies. Figures in blue indicate that the changes support the hypothesis of decreasing engagement.

		Difference in change between multiple-choice items and open-ended items					
		Change between P1 and P4 within the same year					
		2003		2006		2012	
Sverige		Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items
Share correct	General analysis, without control	**-11.5**	**-10.8**			-9.0	**-9.7**
	General analysis, control†	**-11.5**	**-10.8**			**-9.0**	**-9.7**
	Analysis of trend items	**-13.1**[b]	**-10.9**	**-7.4**[c]	**-7.9**	**-9.6**	**-10.6**
Share unanswered	General analysis, without control	**14.0**	**14.6**			**9.3***	**15.9***
	General analysis, control†	**14.0**	**14.6**			**9.3***	**15.9***
	Analysis of trend items	**15.1**[a, b]	**18.1**[b]	**9.5**[c]	**10.4**[a, c]	**8.9*[c]**	**15.6*[b]**
OECD		**2003**		**2006**		**2012**	
Share correct	General analysis, without control	-8.8	**-9.9**			-5.2	-6.6
	General analysis, control†	**-8.8**[a]	**-9.9**[a]			**-5.2**[c]	**-6.6**[c]
	Analysis of trend items	**-10.0**[a]	**-10.3**[a, b]	**-7.4**	**-6.7**[c]	**-5.9**[c]	**-6.9**[c]
Share unanswered	General analysis, without control	**10.7**[a]	**13.8**			4.1[c]	**9.2**
	General analysis, control†	**10.7**[a]	**13.8**[a]			**4.1*[c]**	**9.2*[c]**
	Analysis of trend items	**11.7*[a, b]**	**16.3*[a, b]**	**6.7*[a, c]**	**9.9*[c]**	**4.2*[b, c]**	**9,2*[c]**

If the differences in share between multiple-choice items and open-ended items within a year (perseverance) are significantly different from each other, these are marked with *.

If the differences in share between position 1 and position 4 within a year (perseverance) are significantly different from each other, these are marked in bold type.

a. Perseverance different from 2012.

b. Perseverance different from 2006.

c. Perseverance different from 2003.

† With control for sequence (items 1–12), the item's degree of difficulty and block.

Table 10. Estimates of student response patterns in reading, between positions between years, by item type.

Compilation of changes in student response patterns for reading items by item type when the block is in position 1 (P1) and in position 4 (P4) in the test booklet for the 2003, 2006 and 2012 studies. Figures in red indicate that the changes support the hypothesis of decreasing knowledge. Figures in blue indicate that the changes support the hypothesis of decreasing engagement.

	Difference in change between multiple-choice items and open-ended items							
Analysis of trend items	Change P1 between years				Change P4 between years			
	2003/2009[a]		2009[b]/2012		2003/2009[a]		2009[b]/2012	
Sweden	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items
Share correct	-0.7	-0.9	**-2.1**	**-3.2**	**-6.2**	**-6.9**	-1.4	**-1.8**
Share unanswered	**0.9**	**1.4**	**0.7***	**2.1***	-1.4*	**6.1***	0.3	0.8
OECD	2003/2009[a]		2009[b]/2012		2003/2009[a]		2009[b]/2012	
Share correct	0.5	0.3	-0.2	-0.3	**-3.1**	**-2.3**	**3.5**	**2.6**
Share unanswered	**-0.4**	-0.5	0.1	0.1	**-1.2***	**1.2***	**-2.7**	**-3.2**

If the differences in share between multiple-choice items and open-ended items for each position between two years are significantly different from each other, these are marked with *.

If the differences in share for each position between two years are significantly different from each other, these are marked in bold type.

a. Compare with 2003 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests.

b. Compare with 2012 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests.

Table 11. Estimates of the students' diminishing perseverance in reading items, by item type.

Compilation of changes in student response patterns for reading items by item type between position 1 (P1) and position 4 (P4) within the same year, for the 2003, 2009 and 2012 studies. Figures in blue indicate that the changes support the hypothesis of decreasing engagement.

Analysis of trend items	Difference in change between multiple-choice items and open-ended items							
	Change between P1 and P4 within the same year							
	2003		2009[a]		2009[b]		2012	
Sverige	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items
Share correct	**-10.1**	**-12.1**[d]	**-15.8**	**-17.9**[e]	**-14.0**	**-14.9**	**-13.2**	**-13.5**
Share un-answered	**8.5***	**12.7***	**6.9***	**17.0***	**12.2***	**16.8***	**11.7***	**15.5***
OECD	2003		2009[a]		2009[b]		2012	
Share correct	**-9.2**[d]	**-10.5**	**-13.3**[e]	**-12.8**	**-13.3**[c]	**-12.9**	**-9.6**[d]	**-10.0**
Share un-answered	**6.7***	**11.8***	**6.5***	**13.2***	**8.9***[c]	**13.1***[c]	**6.2***[d]	**9.8***[d]

If the differences in share between multiple-choice items and open-ended items within a year (perseverance) are significantly different from each other, these are marked with *.

If the differences in share between position 1 and position 4 within a year (perseverance) are significantly different from each other, these are marked in bold type.

a. Compare with 2003 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests, and these two years are thus never compared.

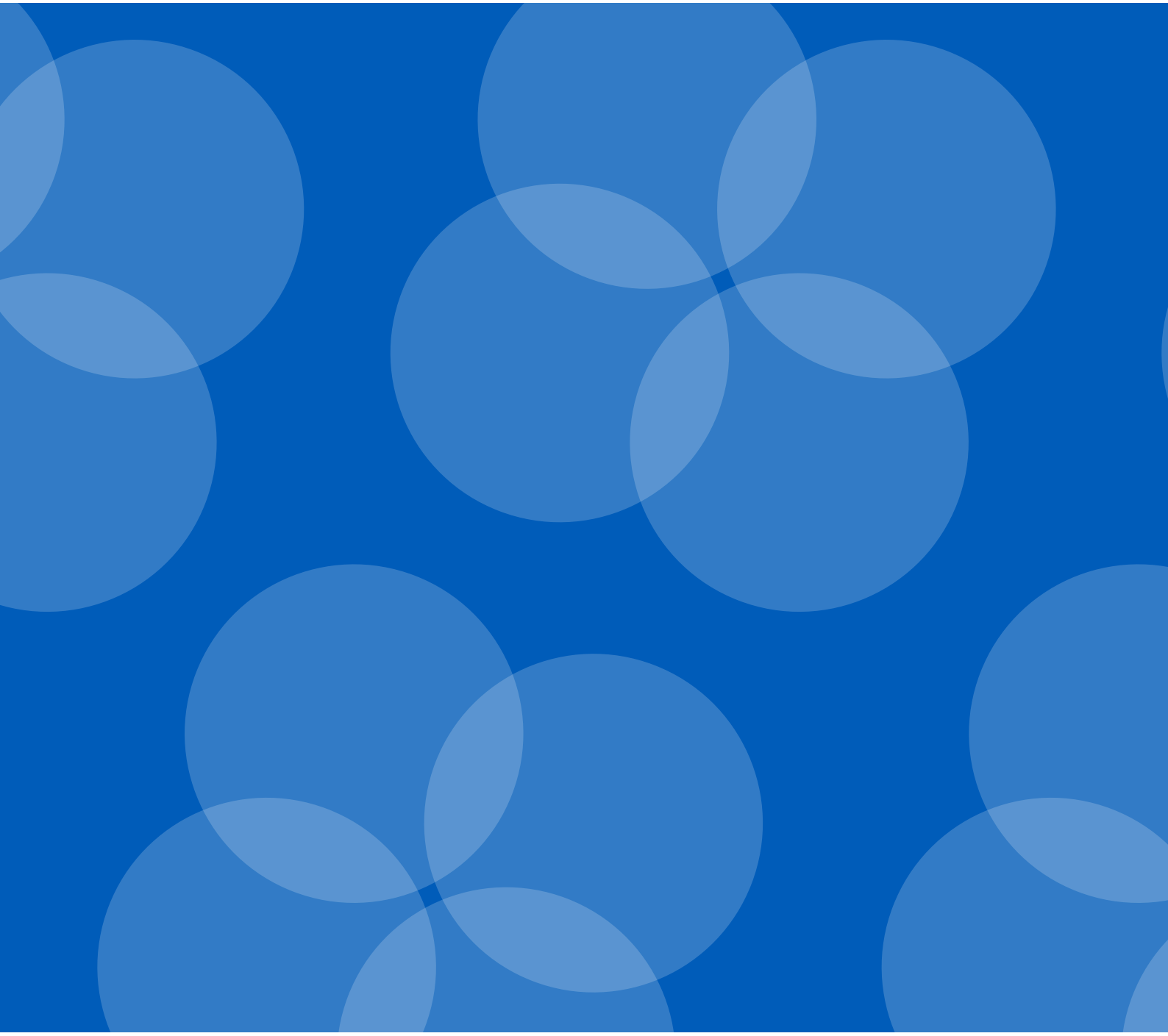
b. Compare with 2012 – only the trend items included. Different test items are included for 2009 depending on the comparison year. None of the reading items from the 2003 tests are included in the 2012 tests, and these two years are thus never compared.

c. Perseverance different from 2012.

d. Perseverance different from 2009.

e. Perseverance different from 2003.

The PISA 2012 results demonstrated a sharp decline in all three PISA competencies: reading, mathematics and science. One of the potential partial explanations discussed was the context in which the PISA survey is conducted. In light of this deterioration in results, the National Agency for Education has, among other things, taken the initiative for the two studies presented here. The studies complement each other in the sense that the analysis of one is based on student responses to questions about effort and motivation in taking the PISA test, while the analysis of the other is based on the students' actual response patterns in the PISA tests. Even though the results from both analyses are to be interpreted with caution, they reinforce each other. Independently of each other, the analyses demonstrate that the sharp decline in results indicated by PISA cannot be explained by a lack of motivation or effort in students.



Skolverket

www.skolverket.se