SFI THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH

Beatrice Schindler Rangvid

03:2016 WORKING PAPER

STUDENT ENGAGEMENT IN INCLUSIVE CLASSROOMS

SFI – THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH

INKLUSIONSPANELET

STUDENT ENGAGEMENT IN INCLUSIVE CLASSROOMS

Beatrice Schindler Rangvid

THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH, COPENHAGEN, DENMARK

Working Paper 03:2016

STUDENT ENGAGEMENT IN INCLUSIVE CLASSROOMS*

Beatrice Schindler RANGVID
The Danish National Centre for Social Research (SFI)

October 2016

Abstract

There is general agreement that to thrive and learn at their best, students must be engaged. However, schools face a particular challenge to provide a suitable and engaging learning environment for SEN (special educational needs) students who are educated in general education classes. Using data from a large scale student panel survey, I document substantial differences in engagement between students with and without special needs in regular classes. I then show that concerning academic achievement, well-being and motivation, engagement appears to be at least as important a determinant for SEN-students as for other students. This highlights the need for better inclusion initiatives aimed at strengthening engagement of SEN-students in regular classrooms.

Keywords: education economics; instrumental variables; special needs education; mainstreaming; student engagement
JEL Classification: I20

* Contact information: bsr@sfi.dk; The Danish National Centre for Social Research (SFI), Herluf Trolles Gade 11, 1052 Copenhagen K, Denmark. This analysis is part of a project commissioned by the Danish Ministry for Children, Education and Gender Equality. The usual disclaimer applies. Thanks to Paul Bingley, Mette Gørtz and the participants of the Expert Panel Workshop held at SFI on June 8, 2016 for valuable comments and advice.

1. Introduction

There is general agreement that engagement in learning is important for success in school. School engagement refers to the students' level of connectedness, involvement, and commitment to school as well as to learn and achieve. Researchers have studied and measured student engagement in many different ways and - regardless of the definition used - student engagement is found to be a robust predictor of student achievement and behavior in school. Students who are engaged and connected to their schools demonstrate increased academic achievement (Goodenow, 1993; Willingham, Pollack & Lewis, 2002; Willingham, Pollack & Lewis, 2002; Roderick & Engle, 2001; Lee & Smith, 1995; Lee & Smith, 1993), higher attendance rates, lower drop-out rates (Croninger & Lee, 2001; Connell et al., 1995) and fewer anti-social behaviors (Wang & Eccles, 2012; Chen, 2005; Bryant et al., 2003; Caraway, Tucker, Reinke, & Hall, 2003).

Recent longitudinal studies suggest that both family and school characteristics have a significant influence on the changes of engagement over time (Wang & Eccles, 2012). Student engagement is a psychosocial process that results from the interaction between individual and contextual factors (Fredricks, Blumenfeld, & Paris, 2004). Specifically, the classroom environment (Guardino & Antia, 2012), and school characteristics help creating positive experiences of school (Moreira et al., 2015; Moreira, Oliveira, Dias, et al., 2014; NRCIM, 2004; Wang & Eccles, 2012; Wang & Eccles, 2013)

Existing studies have demonstrated the importance of student engagement on outcomes for mainstream students. However, empirical evidence for the specific group of SEN students is scarce. The present study contributes to filling the gap.

The general notion is that engagement is important for all students. However, students with Special Educational Needs (SEN) are at increased risk of experiencing difficulties (both related to behavior and learning) which tend to lower engagement with school.

Providing a good fit between individual and contextual characteristics is a challenge that schools face for all students. However, the challenge is even greater in the case of students with SEN, because the school- and classroom environment that may be suitable for the needs of the typical students may not be adequate for SEN students. To some extent engagement in learning may be an inherent trait. However, engagement can be promoted eg. by using teaching styles that enhance student engagement (Zepke & Leach, 2010).

Providing a suitable and engaging learning environment for SEN students is of heightened concern for school systems aiming at raising inclusion rates, i.e. educating increasing numbers of SEN students in general education classes. In recent years, a general consensus has developed to move from a broad definition of inclusion according to which education of SEN students often took place in separate classes or schools to a more narrow one, according to which SEN students should be educated in regular classrooms alongside their same-aged peers. Today, in many countries, large shares of SEN students are educated in regular classes. However, although attendance of general education is necessary for reaping the benefits of school, it is by no means sufficient. Instead, it is argued, SEN students must be fully engaged in school and classroom activities – both behaviourally and emotionally - to reap the full benefits of inclusive education.

This study uses data from Denmark to empirically investigate the impact of student engagement on academic achievement, well-being and motivation for SEN students who are educated in regular classes and their non-SEN peers. I use data from a student panel survey conducted as part of a project commissioned by the Ministry for Children, Education and Gender equality. The study repeatedly surveys about 9,000 students in grades 5 to 9 in participating schools. Engagement is based on this student reported data and is measured by four indices covering student participation (learning/social), student-teacher relations and academic acknowledgement. Moreover, data from administrative registers on student background and academic achievement are linked to the survey data at the student level. Since there is no official marker of students with special needs in the register data, I construct a definition of students with special needs/challenges using both survey and register-information. The resulting share of SEN students in this analysis is somewhat higher than other assessments of the share of SEN students.[1]

The primary methodological challenge of the analysis is potential measurement error due to self-reported data, because information on both engagement and on some of the outcomes comes from the student themselves. As we expect the potential bias on both input and output to be in the same direction, this would lead to an overestimation of the

[1] Due to the definition used in this analysis, 20% of all students have potentially special needs/challenges. Another assessment indicates a somewhat lower number: 14-15% (Nielsen & Skov, 2016).

impact of engagement on outcomes. To examine the extent of the bias, this study exploits peer-reported data, as instruments for self-reported data.

The results of this study demonstrate substantial differences in engagement between students with and without special needs in regular classes. I then show that engagement appears to be at least as important for SEN-students as for other students concerning outcomes like academic achievement, well-being and motivation. This highlights the need for inclusion initiatives aimed at strengthening engagement of SEN-students.

The structure of the paper is as follows: Section 2 provides a short presentation of the background for this study, and section 3 explains the engagement measure used. Section 4 presents the empirical strategy, and Section 5 introduces the data. Section 6 reports the results and Section 7 offers concluding remarks.

2. Background

Following years with steadily increasing numbers of student educated in segregated settings in Danish public schools, in the beginning of this decade a political consensus was reached to reverse this trend. Per pupil costs in segregated education tend to be considerably higher than in regular classrooms, while the evidence on the benefits of segregated education is not conclusive. Moreover, there is now a general consensus that full inclusion of SEN students in regular classrooms with their same-aged peers is the favored educational setting.

A special education reform was implemented in Denmark in 2012 with the specific aim to increase inclusion rates from 94.4% to 96% by 2015.[2] The reform included a legislative change in the Folkeskole Act, which narrowed the definition of special needs education to include only extensive support (more than 12 lessons/week). This meant that from 2013 only students with extra support exceeding 12 lessons per week were included in the category of SEN-students.[3] Moreover, an agreement between the municipalities and the national government outlined the objectives for increased inclusion and the financial responsibility for special needs education was decentralized from the municipal level to the schools. This created strong financial incentives for the

[2] The inclusion rate is the overall share of students educated in inclusive settings (SEN and non-SEN students).

[3] Students still get extra support up to 11 lessons per week, but they are not "visible" as students receiving support in the administrative registers.

schools to provide for SEN students in regular classes, because having students attend special classes or special schools is substantially more costly.

A range of initiatives to support the transition towards a more inclusive education have been implemented as part of the reform. A core iniative is school reform including improving teachers' and staff's skills and professional development and counselling, particularly with respect to inclusion of SEN students. Initiatives to make regular classrooms more inclusive include a strengthened focus on individualised teaching in regular classrooms, temporary subdivision of class, additional lessons, two teachers in class, teachers' assistants, and individual support to help the SEN students overcome practical obstacles related to school attendance. Moreover, information and attitude campaigns for parents and students were conducted and follow-up work was implemented, for example the monitoring of the transition towards more inclusion. This study is part of the follow-up research on the reform commissioned by the Ministry of Children, Education and Gender Equality.

A major challenge for schools is that having SEN students placed in regular class rooms is not sufficient to ensure a learning environment for them where they prosper and thrive. Students need to be engaged in their learning environment to reap the benefits of inclusive education. This study documents student engagement of SEN-students educated in regular classes and then examines whether engagement is positively related to academic achievement, well-being and motivation.

3. Measuring engagement

School engagement refers to the students' level of connectedness, involvement, and commitment to school as well as to learn and achieve. Engagement is a construct with several dimensions that includes affective connections within the academic environment (e.g., positive adult-student and peer relationships) and active student behavior (e.g., participation and effort). The internal nature of some dimensions of engagement means that it is difficult to define and measure. Engagement is about a connection with learning, it is not simply about good classroom behaviour or attendance. Students who are not participating in discussions or completing their work are not engaged. The construct of student engagement has been studied and measured by researchers in many different ways, but there is still no consensus about the number and the type of dimensions. However, regardless of the definition higher levels of engagement seem to be linked with improved performance in school.

Four indices have been constructed by factor analysis methods from the range of questions in the student surveys of the Danish National Panel Study of Inclusion:

- **Participation in learning activities** includes the extent to which a student frequently participates in classroom discussions; puts up his/her hand, when the teacher asks a question (and he/she knows the answer); actively participates when working with classroom peers; and dares to say in class when there is a task/exercise he cannot do.
- **Participation in social activities** includes different ways to socialize with school peers. Students may be together with children from their class during recess or they may be together in their leisure time outside school. Finally, they may participate in school-related social arrangements (e.g. parties with the class, outings with parents or other arrangements with the entire class invited).
- **Student-teacher relations**. Experiences of student-teacher relations include five items that examine the extent to which 1) students like their teachers, 2) students are treated fairly by their teachers, 3) teachers do something about it when a child is bullied, 4) teachers do something for the well-being of all students in the class, and 5) teachers succeed in making their class interesting.
- **Academic acknowledgement** includes five items examining how often 1) peers ask the student for help in class, 2) teachers compliment the student, 3) peers compliment the student when he is doing well in school, 4) when the student thinks that he has done well, his teachers agree, 5) when the student thinks that he has done well, his peers agree.

For use in the regressions, the indices are standardized to a mean of 0 and standard deviation of 1. Higher index-values signify more frequent participation, better student-teacher relations etc. Student engagement variables enter the regressions in lagged form to ensure that the dependent variables are pre-determined.

4. Empirical strategy

The empirical analysis begins by documenting differences in school engagement between SEN-students and other students. The second step of the analysis examines the effect of engagement on a range of student outcomes. Estimating the effect of engagement on student outcomes is a challenge due to observed and unobserved differences between students that might affect both the levels of school engagement and outcomes. To alleviate concerns about selection bias, I take several steps: first, I exploit the panel structure of the survey to obtain pre-determined measures of the dependent variables for use in the analyses. Since the data for the analysis comes from a student panel survey with five data-collection waves, I use the measures of engagement from the previous data collection to avoid regressing engagement on contemporaneous measures of student outcomes.

Second, I also include the lagged outcome as an explanatory variable in the regression, thus effectively using a value-added approach to estimate the effect of engagement. In a value-added model, the coefficients on engagement measure the effect of engagement on *changes*, not *levels*, in the outcome variables. This helps identification, since engagement does not need to be exogenous with respect to the *level* of eg. student motivation, but only with respect to *changes.* This is a much weaker assumption.

Third, even *changes* in the outcome may be correlated with student and family background. I therefore control for such differences by including variables describing the socio-economic status of the student, eg. gender, ethnicity, parental education, income and labour market status.

Moreover, to ensure that time-invariant factors at the school level do not bias the results, I include school fixed effect. Last, to rule out influences from aggregate time shocks, I add year fixed effects.[4]

Formally, the basic model can be written as:

$$Y_{ist} = \alpha + \gamma E_{is,t-1} + \varphi Y_{t-1} + \beta X_{ist} + \sigma_s + \nu_t + \varepsilon_{ist}$$

where Y_{ist} is the outcome of student i in school s at time t, $E_{is,t-1}$ are measures of engagement, Y_{t-1} is the lagged outcome, X_{ist} is a vector of variables describing the

[4] The time fixed effects are effectively "data-collection fixed effects" with a set of dummy variables to indicate which survey wave the observation comes from.

student and family background, σ_s are time-invariant school effects and ν_t capture fixed differences across survey waves.

Thus, this study rules out many potential biases and covers all the basics that other related papers cover. Still, the findings of this study must be interpreted with some caution, because one still cannot be sure that all selection bias is controlled for. If there are underlying unmeasured factors that affect both engagement and changes in outcomes, these methods cannot entirely rule out that there is remaining bias.

The analysis considers both academic and non-academic outcomes. The non-academic outcomes – well-being and motivation & effort – may be regarded as mediating factors linking student engagement to academic outcomes.

Instrumenting self-reported engagement

After the basic model has been established, the primary methodological challenge is the self-reported nature of the data on student engagement. Assuming that the measurement error of the independent variable student engagement is a classical error in variables, the measurement error in the dependent variable may lead to a downward bias in the estimated coefficients and may thus lead to an underestimate of the true effects of student engagement. For example, a study by Ashenfelter & Krueger (1994) found significant evidence of measurement error in self-reported schooling levels. To investigate if measurement error due to self-reported nature of the data is a major concern in this study, I use a method similar to that suggested in Ashenfelter & Krueger (1994). In the second wave of the student panel survey, we asked students not only to answer questions regarding their own engagement, but also asked them to report on two of their peers' engagement. To keep the survey at a reasonable length, two key questions regarding student engagement were chosen.[5] Practically, this was done as follows. For each class an electronic list with students' names was provided. The program code ensured for the first student on the list, that this student was asked about the engagement of the second and third student on the list, the second student was asked about the third and fourth student etc. and the last student in the list – to close the circle – was asked about the first and second student. The reason to collect data on not only one, but two peers, was to reduced the risk of missing peer-reported data, since the

[5] Note, that each students was asked six additional questions: two questions about two peers, plus a question on well-being of two peers.

probability of having a valid answer from at least one peers is greater when asking two. Also, with two answers from two reports, there are two instruments for each instrumented variable and thus, an overidentification tests can be run for the instruments.[6]

5. Data and definitions

I use survey data linked to administrative microdata to examine the effect of student engagement on a range of outcomes. The survey data has been collected within the project "Danish Panel Data Study of Inclusion", that has been commissioned by the Ministry for Children, Education and Gender Equality. The project is part of the follow-up research for the policy initiative to provide more inclusive education. This large-scale data collection monitors the transition towards more inclusive schools from the students' point of view. Student surveys have been administered five times over a two year period from spring 2014 to spring 2016. The surveys collect information on a range of topics including participation, student-teacher relations, acknowledgement, well-being at school, motivation and academic self-confidence and progress. Two grade cohorts of students in almost 200 schools are surveyed: students are followed from grade-levels 5 and 7 in the schoolyear 2013/14 throughout the schoolyear 2015/16 when these students are in levels 7 and 9.

In addition to data retrieved from the surveys, this study exploits administrative microdata containing extensive and reliable information on test scores and students' family background, as well as school and grade identifiers. The administrative data is linked to the survey data at the student-level by unique identification numbers.

Sample

Basically, all 9,350 students from schools and classes that participate in the Danish National Student Panel for Inclusion are included in the sample. Yet, the actual size of the estimation sample for the single regressions depends on the number of missing key variables. While scores from the national tests in reading and math are available for

[6] Note, that validity of the exclusion restriction is not an issue in this application, because the IVmethod is used to control for measurement error, not for identifying causal effects.

almost all students, variables deriving from the surveys are more often missing. For example, the response rate is highest in the first wave of the survey (about 88%), but was lower in later waves (eg. 57% in the last wave). An analysis of the representativeness of the sample of responses showed that the sample of students who had answered the surveys were approximately representative of the population of students at the relevant grade levels. For the estimations, I keep only records with non-missing data on the key variables (engagement and outcomes).

Variables

Outcomes The range of outcome variables considered include both indexes created from the survey questions and test scores from national standardized tests that are available from the administrative registers. The following outcomes are used:

- Reading scores
- Math scores
- Academic self-confidence and progress
- Well-being at school
- Motivation & effort

Reading and math test scores Beginning in 2010, reading tests were administered each spring to students enrolled in grade-levels 2, 4, 6, and 8 in public schools, creating a two year gap between assessments. Students take math tests in grade-levels 3 and 6.

These mandatory tests are high-profile tests. They are IT-based and adaptive, meaning that tests are taken online at computers with the test system choosing questions based on the student's level of proficiency as displayed during the test and automatically calculating test results. The tests simultaneously evaluate the skill levels within three profile areas of reading (language comprehension, decoding, and reading comprehension) and math (numbers & algebra, geometry, applied math). For the analyses in this paper, test scores have been standardized for each test, grade and year to have a mean of 0 and a standard deviation of 1.[7]

[7] Specifically, first the scores are standardized for each profile area for each grade-year combination. Then, scores are averaged across the three profile areas before I standardize the average for each grade-year combination. The resulting final measure of the reading score thus has a standard deviation of one and mean zero.

The remaining three outcome variables derive from questions in the student surveys and they have been constructed by factor analysis methods within the framework of the Danish National Panel Study of Inclusion.

- **Motivation and effort** includes the extent to which a student says he is interested in learning, actively participates when working together with classroom peers, and completes the assignments given to him by his teacher.
- **Academic self-confidence & progress**. Experiences of academic self-confidence & progress include four items that examine the extent to which 1) students think they do well in school, 2) students think they can do the tasks or assignments given by their teachers, 3) students feel they make progress, and 4) students think they read fluently (fast & accurate).
- **Well-being at school** includes six dimensions of enjoying being at school: students may like attending school; they may like their peers in class; treat each other well in class; they are allowed to participate and contribute when working with others in groups; not being disturbed by noise and not considering to switch schools.

The indices are all standardized to a mean of 0 and standard deviation of 1. Higher index-values signify higher motivation, better well-being, etc.

Controls The administrative data holds a range of individual student and family information like gender, immigration background, family type[8], maternal and paternal education level, income and labour market status. This information is included in the regressions as controls for socio-economic student background. Table A1 in the appendix shows descriptive statistics of the variables used in the regressions.

Definition of SEN students

The primary focus of this analysis are students with special needs who are educated in regular classes. As there is no formal definition of SEN-students that can be used to distinguish between SEN and non-SEN students in the data, I use the following criteria for being included in the sample of SEN-students:

- Student receives continuously extra support throughout the survey years

[8] I.e. whether the child lives with both parents or not.

- Student returned recently from segregated special needs education (special class or special school)
- Student has an SDQ-score (Strength and Difficulties Questionnaire) beyond the ”normal range”[9]
- Student has been diagnosed with a psychiatric disorder

Students meeting one or more of these criteria are included in the SEN-sample. According to this definition, 20% of the students from schools and classes participating in the Inclusion Panel Study are included in the sample of SEN-students. Table 1 provides some descriptives for the SEN-sample compared to the non-SEN sample, which suggest that SEN students significantly underperform with respects to all five outcomes. However, the percentage native Danes is equal in both subsamples and only slightly more SEN students are boys. However, substantially more SEN students do not live with both parents and come from less advantaged families (lower education and lower labour market status).

Table 1: Descriptives for the SEN- and non-SEN sample

	SEN	Non-SEN
Reading scores	-0.38	0.10
Math scores	-0.46	0.11
Academic self-confidence & progress	-0.51	0.14
Motivation & effort	-0.54	0.15
Well-being at school	-0.43	0.12
%male	53%	50%
%lives with mom & dad	70%	80%
%native Dane	94%	94%
%mother low-educated (unskilled)	27%	18%
% mother high-educated (long college)	5%	8%
% mother income transfer	18%	12%
%mother high-wage	9%	14%

[9] The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioural screening questionnaire about 3-16 year olds. The self-report version that is used in the student survey is suitable for young people aged around 11-16. For the analyses, I use three subscales of the SDQ: emotional symptoms, conduct problems and hyperactivity/inattention. The scores in each scale range between 0 and 10. Students with scores below 6 are generally considered to be within the ‘normal’ range. According to the definition adopted for use in this study, a student has considerable difficulties if he/she scores outside the normal range in one or more of the three subscales.

6. Results

The aim of this analysis is to investigate whether engagement matters for the outcomes of students with special needs who are educated in regular classes. The first step in this analysis is to document the level of engagement of SEN-students compared to students without special needs. In the second part of the analysis, I investigate whether higher levels of engagement improve SEN-students' outcomes.

Descriptive analysis

As a first step, I present descriptive statistics of engagement levels of SEN students and other students. Figure 1 shows the gap between SEN-students and other students for the four indexes that together describe student engagement. In the table, both raw differences and regression-corrected differences are shown. Regression-adjusted differences control for observed differences in socio-economic background, school fixed effects and time fixed effects to disentangle the correlation between engagement and SEN-status net of such differences.

Figure 1: Engagement indexes: raw and regression-adjusted gaps

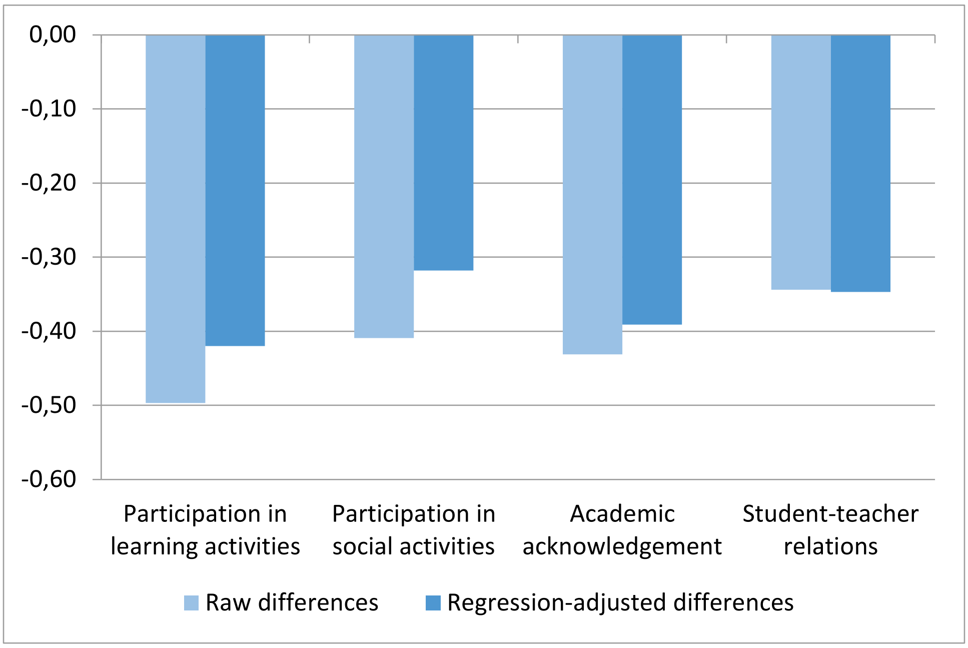


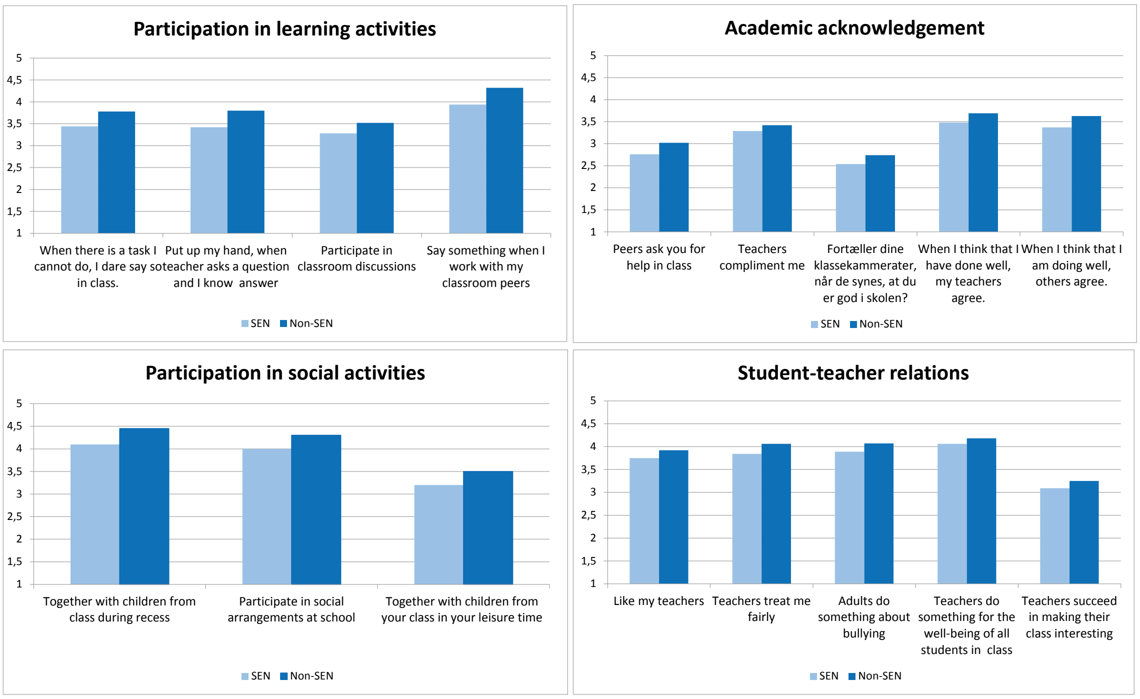
Figure 1 shows that for all four indexes of engagement, SEN-students score substantially lower than other students. The largest raw difference is for the index concerning participation in learning activities, where SEN-students score half a standard deviation lower than other students. As to the indexes for academic acknowledgement and participation in social activities, SEN-students score about 0.4 SD lower, and the gap for the index of student-teacher relations is 0.35 SD. The regression-adjusted differences somewhat reduce the gap for the two participation indexes by about 0.10 SD, but explain only little of the gap in academic acknowledgement. Regression-adjustment leaves the gap for student-teacher relations unchanged.

Overall, there are huge engagement gaps between SEN-students and other students, which are not explained by differences in student background, time-invariant differences at the school level or aggregate time shocks.

Additional calculations show that the left tails of the distributions for SEN-students are significantly fatter than for other students, meaning that much larger percentages of SEN-students have poor engagement. For all but one of the four index of engagement, 35% of SEN-students have an engagement level that is lower than the 20% percentile of non-SEN students. Only for student-teacher relations, the situation is somewhat better with only 30% of SEN-students scoring below the 20% percentile of non-SEN students.

Next, I examine whether gaps between SEN and non-SEN students are particularly large for specific items within each index. Figure 2 shows average values of all items that enter the four indexes of engagement for SEN and non-SEN students. Values are shown on the original scale from 1 to 5, where 5 indicates the highest/best level. The overall impression is that SEN students report lower values than other students for all items. The difference to non-SEN students does not vary much across items. Thus, the difference in the four index of engagement for SEN students and other students we see in Figure 1 is not due to SEN-students reporting lower engagement for a selective number of items, but the impression is one of overall lower engagement levels for all items explored in the survey.

Figure 2: Items in engagement indexes. Average raw scores by SEN-status.



Measurement error

As explained in the section on the empirical method, a primary concern with self-reported data is measurement error. Therefore, before continuing to the empirical analysis of the impact of student engagement on outcomes, I investigate whether bias due to measurement error is a major concern in this study.

The analysis uses the peer-reported data on two key items of student engagement[10] that have been collected in the second survey wave. These peer-reported variables are used as instruments for students' self-reported engagement.

Table 2 presents results. The table shows the estimated coefficients for the variables of interest, i.e. the two key items of the participation index, in regressions on each of the five outcome variables. For each outcome, results from two different models are presented: first results from the simple OLS model using students' self-reported

[10] The two selected questions are the frequency of (1) *participation in classroom discussions* and (2) *being together with children from your class during recess*.

participation measures, and second, results from a model using peer-reported data on the two key items of participation as instruments for self-reported data. In addition to the variables of interest, all models include variables describing students' socio-economic background and the lagged outcome. Moreover, the F-statistics describing the strength of the instruments and overidentification tests are reported and they show that the instruments seem to be very strong and that they pass the test of overidentifying restrictions.

Table 2: OLS and IV-models of the effect of participation on outcomes.

		Participation, learning activities				Participation, social activities				
		Coef	se	Strength (F-test)	Overidentifying restrictions (Sargan χ^2)[i]	Coef	se	Strength (F-test)	Overidentifying restrictions (Sargan χ^2)[i]	n obs
Reading scores	OLS	0.060***	(0.011)			0.007	(0.009)			5398
	IV	0.186***	(0.023)	652	0.845	0.051	(0.029)	327	1.249	5398
Math scores	OLS	0.112***	(0.019)			-0.023	(0.015)			2879
	IV	0.280***	(0.047)	260	0.008	0.050	(0.058)	142	0.170	2879
Academic self-confidence &	OLS	0.073***	(0.016)			0.004	(0.017)			3598
progress	IV	0.146***	(0.036)	432	0.073	0.040	(0.043)	220	0.898	3598
Motivation &	OLS	0.088***	(0.018)			0.042*	(0.017)			3600
effort	IV	0.209***	(0.040)	432	0.001	0.172***	(0.048)	221	2.577	3600
Well-being (at	OLS	0.024	(0.015)			0.075***	(0.021)			3610
school)	IV	0.008	(0.033)	435	0.683	0.154**	(0.054)	222	0.284	3610

[i] *Results from Basmann's chi2 overidentification test are very similar.*
Note: Both models use the IV model specification and sample.

Strength of instruments To be valid as instruments, a variable must be sufficiently correlated with the endogenous regressors, but uncorrelated with the error term. Researchers have devoted much attention to the issue of weak instruments (i.e. instruments that are only weakly correlated with the endogenous regressors). In such cases, the usual estimators are biased toward the OLS estimator and inference can be severely misleading (e.g. Angrist & Pischke, 2009).

Thus, to be valid as instruments, peer-reported student engagement must be good predictors of self-reported engagement. This is tested in the first stage of the instrumental variable analysis, where the variables to be instrumented (here, the two indicators of self-reported engagement) are regressed on the instruments (peer-reported engagement from two peers for each of the two key engagement items). The F-statistics of joint significance of the instruments are reported in Table 2. The F-statistics are all far above the conventional level of 10, ruling out any concerns about a weak instrument problem.

Test of overidentifying restrictions Furthermore, the instruments must also be uncorrelated with the structural error term. If there are more instruments than variables to be instrumented, whether the instruments are uncorrelated with the error term can be assessed by a test (Davidson & MacKinnon, 1993). The overidentification tests suggest we have no obvious reason to distrust the validity of the sets of instruments employed.

IV versus OLS After having asserted the strength and validity of the instruments, we turn to compare the results from the simple OLS model and the IV model. The overall impression is that, qualitatively, the results using the simple OLS model is similar to the IV results. Estimates that are significant using the IV model are significant as well when using the OLS model with signs in the same direction. Likewise, estimates that are imprecisely estimated are so no matter whether the OLS or the IV model is used. Thus, this exercise suggests that we would get qualitatively similar results using the simple OLS specification and the IV correction for measurement error. Measurement error does not have a major impact on the qualitative results and thus on the conclusion of the analysis.

However, that being said, when comparing the size of the (significant) point estimates, it is clear that the IV estimates are all larger than the OLS estimates. These results provide evidence that measurement error is producing a downward bias in conventional estimates. Results of self-reported engagement should thus be regarded as a lower bound. This should be kept in mind when interpreting the results from the main analysis in the next subsection. The main analysis is a full-scale analysis using data from all five survey waves. Remember that the sample available for the IV analysis was restricted, because peer data was only collected in wave 2.

Main results

The examination of the role of measurement error suggested that the conventional OLS model produces qualitatively similar results as the IV-model controlling for measurement error (although the OLS results should be regarded as a lower bound estimate). Therefore, in the next part of the analysis, I do a full-scale analysis of the effect of student engagement on outcomes exploiting the full set of student engagement indicators and data from all five waves of survey data using the conventional OLS model.

Table 3 presents the main results.[11] As before, only the estimates of main interest are displayed in the results table, but all regressions include variables describing students' socio-economic background, the lagged outcome and school fixed effects. Regressions that use outcomes deriving from the student panel surveys furthermore include a set of dummies indicating which survey wave the records come from (thus accounting for time fixed-effects).[12]

The results show mostly significant effects of student engagement indicators. In particular, all four dimensions of engagement seem to have a positive and significant effect on student motivation & effort and on student well-being. Moreover, the two academic dimensions of engagement (participation in learning activities and academic acknowledgement) have positive effects on the three academic outcomes as well (reading scores, math scores, academic self-confidence & progress), while the estimates of the two non-academic dimensions of engagement are insignificant (or even negative). The effect sizes are modest, varying between 0.03 SD and 0.09 SD. However, keeping the results from the measurement error analysis in mind, these are probably lower bound estimates.

These results indicate that engagement is important to student outcomes in school. Students who participate in learning and social activites, who experience academic acknowledgement and good relations with their teachers are more likely to score higher on the outcome scales.

With Hattie's rule of thumb that a learning gain of 0.4 SD corresponds to the learning gain of roughly one year of schooling, we can translate the estimated effect sizes into months of learning (Hattie, 2009). For example, the estimated effect on math scores of a one SD increase in participation in learning activities is 0.089 SD (Table 3). This corresponds to a learning gain of roughly 3 months. However, the IV results in Table 2 suggest that there is a scaling factor of 2.5 (=0.280/0.112) due to the self-reported nature

[11] A full results table for one of the regressions (motivation&effort) is provided in the appendix (Table A2).

[12] For the two outcomes from registerdata (reading and math test scores), only one record per student is available.

of the data. Thus, the true effect size might be as large as 7,5 months. For reading, the effect is 0.053 corresponding to 1,5 months. The scaling factor for reading is 3.1 (=0.186/0.060) and thus the effect may be as large as 0.16 corresponding to 5 months of learning gain.

Table 3 Effect of engagement on student outcomes. Full sample results.

	Reading scores	Math scores	Academic self-confidence and progress	Motivation & effort	Well-being (at school)
Participation, learning activities	0.053***	0.089***	0.075***	0.075***	0.029***
	(0.009)	(0.018)	(0.009)	(0.010)	(0.008)
Academic acknowledgement	0.039***	0.063***	0.037***	0.051***	0.036***
	(0.010)	(0.017)	(0.008)	(0.008)	(0.008)
Participation, social activities	0.001	-0.044**	0.014	0.030***	0.051***
	(0.009)	(0.014)	(0.007)	(0.008)	(0.009)
Student-teacher relations	0.006	0.025	0.012	0.034***	0.080***
	(0.009)	(0.020)	(0.007)	(0.008)	(0.009)
N	6,775	3,514	17,338	17,343	17,405

Robustness

In this subsection, I examine the sensitivity of the main results to changes in the model specification. First, in the main specification, I use the same specification for the measurement of the treatment period both for the test score regressions and the regressions with outcomes deriving from the surveys, i.e. I include the engagement measure with a one period lag. For example, in regressions with the post-score retrieved from the 3rd survey wave, student engagement measures from the 2nd wave are included. However, the timing of the reading and math test scores, where post-scores are measured roughly at the same time as the 3rd survey (Spring 2014), allows us to use student engagement measures not only from the 2nd survey, but also from the 1st survey (which is measured in Spring 2013, i.e. after the pre-scores that is from Spring 2012). Thus, instead of using only the engagement measure from the second survey, in this robustness check, I measure student engagement by the simple average of students' responses in the first and the second wave. This measures engagement over a longer time period – over two waves instead of only one.

However, the results are similar. Table 4 repeats the main results for reading and math scores (columns 1 and 3) and adds the results for the regressions using engagement data from waves 1 and 2 (columns 2 and 4).[13] Thus, the main conclusions are robust to this change in the specification.

Table 4: Sensitivity checks

	(1)	(2)	(3)	(4)	(5)	(6)
	Reading scores		Math scores		Well-being	
	Main results, repeated	Wave 1 and 2 engagement measures	Main results, repeated	Wave 1 and 2 engagement measures	Self-reported measure	Peer reported measure
Participation in learning activities	0.053^{***}	0.060^{***}	0.089^{***}	0.110^{***}	0.052***	0.061***
	(0.009)	(0.010)	(0.018)	(0.019)	(0.014)	(0.012)
Participation in social activities	0.001	-0.002	-0.044^{**}	-0.046^{**}	0.030*	0.045***
	(0.009)	(0.010)	(0.014)	(0.016)	(0.014)	(0.013)
Academic acknowledgement	0.039^{***}	0.038^{***}	0.063^{***}	0.056^{**}	0.040**	0.055***
	(0.010)	(0.011)	(0.017)	(0.018)	(0.014)	(0.012)
Student-teacher relations	0.006	-0.007	0.025	0.024	0.087***	0.034*
	(0.009)	(0.011)	(0.020)	(0.024)	(0.017)	(0.015)
N	6,775	6,775	3,514	3,514	10,848	10,848

Standard errors in parentheses $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Next, I examine the sensitivity of the conclusions when using peer-reported outcome data instead of self-reported data. Together with the data on peer-reported engagement collected during the 2nd survey wave, we also collected peer-reported data on one key element of the well-being indicator.[14] As an additional robustness check, I re-estimated the regression on well-being reported in Table 4 using peer-reported well-being. Because data was collected from two peers, I stack the data, so there are two records per student, one for each peer-reported outcome. Table 4 shows results using peer-reported data for student well-being (column 6). For comparison purposes, I also show results using the same sample, but estimating with self-reported well-being (column 5). The

[13] The samples for the robustness analyses are restricted to be the same sample as for the main specification.
[14] The following item was selected to represent well-being: *Do you like school?*

results show that no matter whether peer- or self-reported outcomes are used, the estimated coefficients on the engagement variables are all positive and significant. While three of the four estimates are larger with peer-reported outcome, the last estimate is only half the size than with self-reported data. However, overall, the main conclusions do not change.

SEN students

Heterogenous effects The main results suggest that engagement is important for the average student. However, in this study, we are explicitely interested in what engagement means for SEN students who are educated in regular classes. Therefore, the main model is reestimated including an indicator for whether the student has special educational needs and its interaction with the engagement variables. In this modified model, the engagement main effect is the effect for students without special needs, while the interaction term indicates, whether the effects is different for SEN-students.

Table 5: Heterogeneous effects of engagement for SEN-students.

		Reading scores	Math scores	Academic self-confidence and progress	Motivation & effort	Well-being (at school)
Participation, learning activities	Main effect	0.056***	0.100***	0.072***	0.078***	0.010
		(0.009)	(0.020)	(0.008)	(0.010)	(0.009)
	Interaction: SEN-students	-0.036	-0.079**	-0.009	-0.018	0.039*
		(0.020)	(0.029)	(0.021)	(0.024)	(0.018)
Academic acknowledgement	Main effect	0.041***	0.061**	0.034***	0.041***	0.035***
		(0.011)	(0.021)	(0.008)	(0.008)	(0.008)
	Interaction: SEN-students	-0.022	-0.002	0.007	0.022	-0.012
		(0.021)	(0.031)	(0.020)	(0.021)	(0.019)
Participation, social activities	Main effect	-0.014	-0.072***	0.008	0.027**	0.048***
		(0.010)	(0.017)	(0.008)	(0.008)	(0.009)
	Interaction: SEN-students	0.052*	0.090**	0.000	-0.011	0.009
		(0.023)	(0.033)	(0.018)	(0.018)	(0.020)
Student-teacher relations	Main effect	-0.001	0.016	0.004	0.024**	0.065***
		(0.010)	(0.021)	(0.008)	(0.009)	(0.010)
	Interaction: SEN-students	0.007	-0.003	0.001	0.010	0.045**
		(0.019)	(0.033)	(0.017)	(0.020)	(0.016)

Table 5 shows results for the engagement main effect and the interaction with the SEN-indicator. The estimate of main interest is the interaction effect indicating whether the engagement effect is for SEN-students than for other students. As can be seen from the table, most interaction estimates are insignificant indicating that the effect of engagement on outcomes is the same for SEN-students and others. However, some interaction effects are significant. All but one of these has a positive sign suggesting that if engagement has a differential impact on SEN-students and others, engagement tends to be more important for SEN-students. The overall impression is that engagement is at least as important for SEN students as for other students.

As before for the full sample results, these estimates can be used to calculate effect sizes for SEN students using the results in Table 5. For example, for SEN students the estimated effect on reading scores of a one SD increase in participation in learning activities is 0.021 SD (=0.100-0.079). Together with the scaling factor for SEN students for reading scores, this translates into roughly 6 months of learning gains – using Hattie's rule of thumb (Hattie, 2009).[15] For math, the estimate is 0.020 SD (=0.056-0.036) and using the scaling factor for math, this translates into approximately 4 months of learning gain.

Table 6: Medium term effects of engagement.

	Motivation & effort	Academic self-confidence and progress	Well-being (at school)
Participation in learning activities	0.231^{***}	0.191^{***}	0.065^{**}
	(0.023)	(0.022)	(0.022)
Participation in social activities	0.116^{***}	0.202^{***}	0.065^{*}
	(0.023)	(0.026)	(0.028)
Academic acknowledgement	0.032	-0.015	0.186^{***}
	(0.025)	(0.023)	(0.034)
Student-teacher relations	0.182^{***}	0.088^{**}	0.245^{***}
	(0.028)	(0.026)	(0.027)
N	3,402	3,402	3,424

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Medium term effects In the main estimation, outcomes that derive from the surveys are regressed on student engagement measured in the previous survey wave - about 6

[15] For SEN students, the OLS and IV estimates used to calculate the scaling factors are 0.21 and 0.02 for reading scores and 0.38 and 0.06 for math.

months earlier. The estimated effects thus measure short run effects of engagement on outcomes. However, one could imagine that medium term effects might be larger. Therefore, I present results for the three survey outcomes when the effect is measured after two years, corresponding to the overall time-span of the panel survey (2014-2016). The outcome is measured in the last survey (survey 5) and the lagged outcome is taken from the first survey (survey 1). Student engagement is measured as the average student engagement level during the first four survey waves (surveys 1 – 4). Table 6 presents the medium term results. The overall impression is that while roughly the same estimates are significant and with the same sign for both the short run (Table 3) and the medium term results, the estimates in the medium term regression are larger. This suggests that the effect from engagement accumulates over time: effects from consistently high engagement over a longer time period leads to even higher levels of motivation, achievement and well-being.

When estimating the medium run regression with interaction terms for SEN-students, I find that the medium-term effect for SEN-students only differs from the effect for non-SEN students in two instances (results not shown, but available upon request). The medium-term effect of participation in learning activities on well-being is significantly larger for SEN-students, as is the effect of academic acknowledgement on academic self-confidence and progress. Overall, the picture is similar to that for short-run effects: the effects for SEN students and other students are mostly homogenous and in some few instances, effects are larger for SEN-students.

7. Concluding remarks

The general notion is that student engagement is an important dimension for successful inclusion of students with special needs in regular classrooms. There is a general understanding that student engagement is important for both academic and non-academic outcomes and for regular students and students with special educational needs alike. This study rigorously tests this proposition using data from large scale student panel-survey linked to administrative data. The present study contributes to the sparse literature on the importance of student engagement on outcomes for SEN-students.

Identification of the model relies on a value-added approach combined with controlling for a large range of student background characteristics and school fixed effects to take account of underlying factors that affect both engagement and outcome-changes.

Overall, the results suggest that student engagement seems to be important for student outcomes. Smaller short-run effects accumulate to larger effects in the medium term. Student engagement is important for outcomes of students with special needs and other students *alike*. The few significant differences found indicate that engagement may be even more important for SEN-students than for others. An additional analysis into the potential role of measurement error due to self-reported data suggests that the results of the main model should be seen as lower bound estimates. Using the IV- versus OLS-results from the restricted sample analysis suggests that effects sizes may be scaled by a factor two to three.

The successful inclusion of SEN-students in regular classes is of much political interest, because educating students in segregated settings as special classes or special schools is costly and the evidence of the benefits is not conclusive. The findings yield valuable information for the design of inclusion policies. The results provide empirical evidence for the general notion that to be properly included in school, it is not sufficient to be placed in the class together with regular peers, but to learn at their best they must be engaged. Participating in learning and social activities, experiencing academic acknowledgement and having good student-teacher relations are all factors that motivate students and facilitate learning and well-being at school. Thus, student engagement initiatives should be a key feature in any inclusion policies.

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Bryant AL, Schulenberg JE, O'Malley PM, Bachman JG, & Johnston LD (2003): How Academic Achievement, Attitudes, and Behaviors Relate to the Course of Substance Use During Adolescence: A 6-Year, Multiwave National Longitudinal Study. *Journal of Research on Adolescence.* 13:361-397.

Caraway K, Tucker CM, Reinke WM, & Hall C (2003): Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools.* 40:417-427.

Chen JJL (2005): Relation of academic support from parents, teachers, and peers to Hong Kong adolescents' academic achievement: The mediating role of academic engagement. *Genetic, Social, and General Psychology Monographs.* 131:77-127.

Connell JP, Halpern-Felsher B, Clifford E, Crichlow W, Usinger P. (1995): Hanging in there: behavioral, psychological, and contextual factors affecting whether African-American adolescents stay in school. *J Adolesc Res.* 10(1):41-63.

Croninger RG, Lee VE. (2001): Social capital and dropping out of high schools: benefits to at- risk students of teachers' support and guidance. *Teachers College Record.*103(4):548-581.

Davidson, R., & J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Fredricks, J.A., Blumenfeld, P.C. & Paris, A.H. (2004): School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109.

Goodenow, C. (1993): Classroom belonging among early adolescent students: Relationships to motivation and achievement. *J Early Adolesc.* 13(1):21-43.

Guardino, C. & Antia, S.D. (2012): Modifying the classroom environment to increase engagement and decrease disruption with students who are deaf or hard of hearing. *Journal of Deaf Studies and Deaf Education, 17*, 518-533.

Hattie, J. (2009): *Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement*. London/ New York: Routledge

Lee, V. E., Smith, J. B. (1993): Effects of school restructuring on the achievement and engagement of middle-grade students. *Soc Educ*. 66:164-187.

Lee, V. E., Smith, J. B. (1995): Effects of high school restructuring on gains in achievement and engagement for early secondary school students. *Soc Educ.* 68:241-270.

Nielsen, C.P. & P.R. Skov (2016) *Støtte til elever med særlige behov: En kvantitativ beskrivelse med udgangspunkt i Inklusionspanelet*. SFI-notat. København: SFI – Det Nationale Forskningscenter for Velfærd.

Roderick, M. & Engle, M. (2001): The grasshopper and the ant: Motivational responses of low- achieving students to high-stakes testing. *Educ Eval Policy Analysis.*23(3):197-227.

Wang MT & Eccles JS (2012): Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child development.* 83:877-895.

Wang, M. T. & Eccles, J. S. (2013): School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28*, 12-23.

Willingham, W.W; Pollack, J. M. Lewis, C. (2002): Grades and test scores: accounting for observed differences. *J Educ Meas.* 39(1):1-37.

Zepke, N. & L. Leach (2010): Improving student engagement: Ten proposals for action. *Active Learning in Higher Education*. 11(3) 167–177.

Appendix

Table A1: Descriptive statistics

Variable	Obs	Mean	SD	Min	Max
Outcomes					
Motivation & effort	23514	0	1	-5,76	1,36
Academic self-confidence & progress	23506	0	1	-4,79	1,90
Well-being at school	23591	0	1	-5,58	1,90
Reading scores	11360	0	1	-7,40	3,86
Math scores	11414	0	1	-8,25	5,57
Variables of interest (engagement indicators)					
Participation in learning activities	7312	0	1	-3,88	1,64
Participation in social activities	7256	0	1	-4,69	1,40
Academic acknowledgement	7319	0	1	-3,88	3,05
Student-teacher relations	7343	0	1	-4,54	1,93
Controls					
Male	11717	0,50	0,50	0	1
Lives with both parents	11715	0,78	0,42	0	1
1st generation immigrant	11717	0,01	0,08	0	1
2nd generation immigrant	11717	0,06	0,23	0	1
Psychiatric disorder	12707	0,05	0,22	0	1
Mother's age at birth	11643	29,94	4,77	16	47
Father's age at birth	11430	32,58	5,66	13	70
Mother's highest education (reference: lower secondary school)					
Vocational education and training	11483	0,37	0,48	0	1
High-school diploma	11483	0,08	0,27	0	1
Short tertiary education	11483	0,04	0,21	0	1
Bachelor	11483	0,24	0,43	0	1
University	11483	0,08	0,27	0	1
Father's highest education (reference: lower secondary school)					
Vocational education and training	11218	0,45	0,50	0	1
High-school diploma	11218	0,05	0,22	0	1
Short tertiary education	11218	0,07	0,26	0	1
Bachelor	11218	0,13	0,34	0	1
University	11218	0,09	0,29	0	1
Disposable income, mother (100,000 DKK)	11643	0,13	0,05	0	2,04
Disposable income, father (100,000 DKK)	11373	0,15	0,12	0	7,51
Mother's labour market status (reference: low-wage job)					
Self-employed	11643	0,03	0,17	0	1
High-wage job	11643	0,13	0,33	0	1
Medium-wage job	11643	0,21	0,41	0	1
Other wage levels	11643	0,12	0,32	0	1
Permanent income transfers	11643	0,13	0,34	0	1
Other employment categories	11643	0,08	0,28	0	1
Father's labour market status (reference: low-wage job)					
Self-employed	11430	0,09	0,29	0	1
High-wage job	11430	0,17	0,37	0	1
Medium-wage job	11430	0,13	0,34	0	1
Other wage levels	11430	0,20	0,40	0	1
Permanent income transfers	11430	0,07	0,26	0	1
Other employment categories	11430	0,03	0,18	0	1

Table A2: Full results, main regression for motivation & effort

Variable	Coef	se
Variables of interest (engagement indicators)		
Participation in learning activities	0.075***	(0.010)
Participation in social activities	0.030***	(0.008)
Academic acknowledgement	0.051***	(0.008)
Student-teacher relations	0.034***	(0.008)
Controls		
Lagged outcome (motivation & effort)	0.499***	(0.012)
Male	-0.153***	(0.015)
Lives with both parents	0.050**	(0.017)
1st generation immigrant	0.022	(0.088)
2nd generation immigrant	0.096**	(0.035)
Psychiatric disorder	-0.121*	(0.049)
Mother's age at birth	-0.000	(0.002)
Father's age at birth	-0.000	(0.002)
Mother's highest education (reference: lower secondary school)		
Vocational education and training	0.078***	(0.018)
High-school diploma	0.101***	(0.026)
Short tertiary education	0.105**	(0.033)
Bachelor	0.100***	(0.023)
University	0.101***	(0.030)
Father's highest education (reference: lower secondary school)		
Vocational education and training	0.048*	(0.019)
High-school diploma	0.101**	(0.031)
Short tertiary education	0.101***	(0.029)
Bachelor	0.089***	(0.025)
University	0.122***	(0.030)
Disposable income, mother (100,000 DKK)	0.031	(0.047)
Disposable income, father (100,000 DKK)	0.131	(0.214)
Mother's labour market status (reference: low-wage job)		
Self-employed	-0.020	(0.030)
High-wage job	0.019	(0.025)
Medium-wage job	0.049*	(0.020)
Other wage levels	0.017	(0.020)
Permanent income transfers	-0.070**	(0.023)
Other employment categories	-0.018	(0.028)
Father's labour market status (reference: low-wage job)		
Self-employed	0.032	(0.024)
High-wage job	0.024	(0.022)
Medium-wage job	0.029	(0.021)
Other wage levels	0.005	(0.018)
Permanent income transfers	-0.006	(0.034)
Other employment categories	0.107**	(0.034)
N	17343	