

RESEARCH ARTICLE

The Strengths and Difficulties Questionnaire and standardized academic tests: Reliability across respondent type and age

Maria Keilow^{1,2*}, Hans Henrik Sievertsen^{1,3}, Janni Niclasen⁴, Carsten Obel²

1 VIVE—The Danish Center for Social Science Research, Copenhagen, Denmark, **2** Department of Public Health, Aarhus University, Aarhus, Denmark, **3** Department of Economics, University of Bristol, Bristol, United Kingdom, **4** Frederikshøj skole og dagbehandling, Primulavej 2, DK-2720 Vanløse and Steno Diabetes Center Copenhagen, Capital Region of Denmark, Health Promotion, Diabetes Prevention Research, Gentofte, Denmark

* mkw@vive.dk



OPEN ACCESS

Citation: Keilow M, Sievertsen HH, Niclasen J, Obel C (2019) The Strengths and Difficulties Questionnaire and standardized academic tests: Reliability across respondent type and age. PLoS ONE 14(7): e0220193. <https://doi.org/10.1371/journal.pone.0220193>

Editor: Mette Goertz, University of Copenhagen, DENMARK

Received: February 22, 2019

Accepted: July 10, 2019

Published: July 25, 2019

Copyright: © 2019 Keilow et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Due to restrictions related to Danish law and protecting citizen privacy, the combined set of data as used in this study can only be made available through a trusted third party, Statistics Denmark. University-based Danish scientific organizations can be authorized to work with data within Statistics Denmark and such organization can provide access to individual scientists inside and outside of Denmark. Requests for data may be sent to Statistics Denmark or the Danish Data Protection Agency. The DNBC data are open access and permission to use these data may

Abstract

Exploiting nation-wide data from the Danish National Birth Cohort, we show that children's emotional and behavioral problems measured by the Strengths and Difficulties Questionnaire (SDQ) are closely related to their performance in standardized academic tests for reading and mathematics in sixth grade. The relationship is remarkably linear across the entire distribution for both the total difficulties score and subscale scores of the SDQ; higher scores on the SDQ (more problems) are related to worse performance in academic tests. We assess the similarity across respondent type; parent (child age 7 and 11), teacher (child age 11) and self-reported scores (child age 11), and find that teacher and parent reported scores have very similar slopes in the SDQ–test score relationship, while the child reported SDQ in relation to the academic test performance has a flatter slope.

Introduction

While individual behavior and mental health are interesting outcomes in themselves, the growing literature on the importance of these characteristics for educational achievement and labor market outcomes [1–4] has sparked the demand for validated instruments. One widely used potential candidate for measuring children's emotional and behavioral problems is the Strengths and Difficulties Questionnaire (SDQ), which is increasingly used by economists, sociologists [5–14], and by other professionals.

The SDQ has been extensively evaluated for its internal consistency, test-retest reliability, and inter-rater agreement [1,15]. However, less is known about how the SDQ relates to measures of academic test performance across respondent type and child age.

The objective of this paper is twofold: First, we assess how the SDQ is related to performance in standardized academic tests across three respondent types (parent, teacher, and child self-reported scores) and two age-levels of the child (age 7 and 11). Second, we exploit our sample size to nonparametrically identify the functional form of the relationship between

be granted by the DNBC Steering Committee. The procedure and conditions for approval is described here: <https://www.dnbc.dk/access-to-dnbc-data>.

Funding: Sievertsen acknowledges financial support from the Danish Council for Independent Research through grant 09-070295. This paper uses data from the Danish National Birth Cohort (DNBC). The DNBC was established with a significant grant from the Danish National Research Foundation. Additional support was obtained from the Danish Regional Committees, the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Health Foundation and other minor grants. The DNBC Biobank has been supported by the Novo Nordisk Foundation and the Lundbeck Foundation. Follow-up of mothers and children have been supported by the Danish Medical Research Council (SSVF 0646, 271-08-0839/06-066023, 0602-01042B, 0602-02738B), the Lundbeck Foundation (195/04, R100-A9193), The Innovation Fund Denmark 0603-00294B (09-067124), the Nordea Foundation (02-2013-2014), Aarhus Ideas (AU R9-A959-13-S804), University of Copenhagen Strategic Grant (IFSV 2012), and the Danish Council for Independent Research (DFF – 4183-00594 and DFF - 4183-00152). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the SDQ and the performance in standardized academic tests, both on the aggregate level (total difficulties score) and for each of the five subscales of the SDQ (hyperactivity/inattention, conduct problems emotional symptoms, peer relationship problems, and prosocial behavior). In other words, this paper assesses whether the SDQ can be used as a continuous outcome, and whether it matters, if scores from the parent, teacher or child self-reported versions of the SDQ questionnaire are used.

The study most similar to ours is one carried out by Kristoffersen et al. [12], who show that teacher and parent reported SDQ is closely associated with academic performance in ninth grade final exams. Kristoffersen et al. also find that teachers are more likely than parents to report extreme SDQ values (very low/high level of problems). Compared to Kristoffersen et al. [12], we add the self-reported version of the SDQ for children aged 11 in our analyses. Furthermore, we assess the functional form of the relationship at the aggregate as well as the subscale level. The latter is important, as studies using SDQ data typically focus on differences across the three pre-defined thresholds; normal, borderline and abnormal, although variation within these categories might also be informative. We therefore assess whether the relationship between the SDQ scores and the performance in standardized academic tests reflects these three groups, or whether the variation within the groups is informative.

Methods

Data

Data for this study consists of registry data and data collected by the Danish National Birth Cohort (DNBC). The study was approved by the National Agency for Education and Quality under the Danish Ministry of Education, the Danish National Birth Cohort Steering Committee, and the Danish Data Protection Agency, respectively. Data were analyzed anonymously on secure servers at Statistics Denmark in line with current data protection requirements. Regional ethics committees approved the collection of data for the DNBC prior to this study. Consent was collected from parents and teachers for data being used for research purposes in accordance with the approved procedures.

Measurements

The Strengths and Difficulties Questionnaire. The SDQ was developed by the English child psychiatrist Robert N. Goodman. The questionnaire assesses emotional and behavioral problems in children and adolescents [16–18]. It consists of 25 items and an impact part and has the advantage of being relatively short and uniform across respondent type (parent, teacher or child) [19]. In addition, the inclusion of positive items makes SDQ suitable across non-clinical samples. The established scoring procedure for the SDQ links each of the 25 items to one of five distinct subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and pro-social behavior. The sum scores for each of these scales range from zero to ten. The first four categories can be aggregated by adding the subscale scores to a total difficulties score (ranging from zero to 40 points). In the nonparametric part of the analysis, we exclude SDQ levels with less than 100 observations due to issues of imprecision. Consequently, the graphs do not necessarily show the entire range of the scale. All observations are, however, included in the linear regression analysis.

The Danish National Birth Cohort (DNBC). The DNBC is a longitudinal nationwide cohort and includes over 100,000 pregnancies (from more than 92,000 women) enrolled from 1996–2002, representing 31 percent of all pregnancies in Denmark during that period [20]. The DNBC cohorts thus include children, who were born from 1997–2003. The women were interviewed twice during pregnancy and when the child was 6 and 18 months, 7 and 11 years.

In this paper, we use data from the fifth wave (children age 7) and the sixth wave (children age 11) of the DNBC. Both waves contain parent reported SDQ scores. The sixth wave also includes child and teacher reported SDQ scores.

The Danish National Tests. To measure student academic achievement, we include data from the Danish National Tests taken in the years 2010–2014. These academic tests are self-scoring and adaptive tests that are provided from the Danish Ministry of Education systematically and nationwide throughout compulsory public schools with certain subjects tested at certain grade levels [21]. We use tests in mathematics and reading in sixth grade, when the children are approx. twelve to thirteen years old. The tests at this grade level were chosen a priori to the study because of their timing in relation to the SDQ measurements for the available cohorts in the study sample; i.e. children take these tests at age 13 and thus the tests are subsequent to the SDQ measurements at age 7 and 11. Moreover, at sixth grade, tests are available for both mathematics and reading. In all analyses, academic test scores are standardized to a mean of zero and a unit standard deviation.

Sample selection

We link data on SDQ scores from the DNBC [22] to data on results from the Danish National Tests [21] and a number of socioeconomic covariates. Of the 461,635 children born between 1997 and 2003 and observed in the register data, we can match 68,602 to at least one complete SDQ score in the DNBC and of those, we are able to match 42,863 children to sixth grade test scores in the Danish National Tests in either reading or mathematics. Last, 1095 children are excluded due to missing covariates. The final sample consists of 41,768 children, corresponding to nine percent of the relevant birth cohorts, i.e. the 1997–2003 cohorts.

There are three sources of non-random selection: First, selection bias due to non-participation may occur. Women were recruited to the DNBC through their general practitioner at routine health control visits during their pregnancy. Approx. two-thirds of those women, who were invited to participate in the DNBC, enrolled and completed at least one interview. Groups with lower socioeconomic resources in terms of occupation, education, and income are underrepresented in the DNBC compared to the general population [20,22,23]. Since participation in the DNBC required Danish language skills, families with non-western origin are also underrepresented. Teachers of children participating in DNBC were recruited for the age 11 follow-up and included approx. 19,000 teachers, corresponding to 65 percent of those, who were invited. Second, the Danish National Tests only cover children in public schools. Selection into private schools may be related to emotional or behavioral problems. In addition, children with severe mental health problems may be excused from taking the test. The relationship between the SDQ scores and the standardized academic test scores might therefore be different for the children not covered by our data. Third, we only include complete cases. This sample selection criterion is not very restrictive, as all covariates are observed for nearly all observations, i.e. for 97 percent.

Statistical analyses

We first describe sample characteristics in terms of variable means. We compare population means to the study sample means and compare the study subsample means among each other to assess external and internal validities, respectively.

Next, we present nonparametric analyses of the relationship between SDQ scores (total difficulties score and subscale scores) and standardized academic test scores (reading and mathematics). So far, researchers have very often relied on the binary or categorical scoring of the SDQ in categories of abnormal/borderline vs. normal scores [12]. To the best of our

knowledge, there is no evidence of whether variation in the SDQ scores away from these thresholds is correlated with school performance. To assess this, we compute mean academic test scores for each level of the total difficulties scale and present this relationship in graphs. We compare the relationship for all respondent types; parents (children age 7 and 11), teachers (children age 11) and children age 11.

Last, we regress academic test scores on the SDQ scores using ordinary least squares. We restrict data to complete case analyses. Estimating these linear regressions, we present unconditional estimates and estimates conditioned on a rich set of controls that include parental education, income, and age at childbirth, child gender, birthweight and non-western origin. We also include school fixed effects in the conditional models to correct for unobservable variation across schools. These regression results are presented for reading and mathematics test scores, for SDQ total difficulties score as well as subscale scores and for each respondent type. While parametric tests rely on assumptions regarding both distribution and functional form, semiparametric tests rely on one or the other (e.g., our linear regression results assume a linear functional form). The graphs presented rely on neither and are therefore considered nonparametric. All analyses are performed using Stata version 14.

Results

Sample characteristics

Linking the DNBC data to administrative records from Statistics Denmark, [Table 1](#) presents variable means by subsamples for all covariates included in the analyses. Parent educational level is measured as years of schooling for the parent with the longest education. Parents' income is presented as total gross household income in 1,000 Euro (2010 price level), and equalized by dividing the total household gross income by the square root of the number of household members. Non-western origin indicates that either or both parents have a non-western origin.

[Table 1](#) shows that for the external validity of our results, children in our sample have higher birthweight, and their parents have higher incomes and have completed more years of schooling compared to the general population. In addition, families with non-western origin are underrepresented in the sample. Statistical t-tests show that these differences are significant. However, important for the internal validity of our study, the subsamples in columns (2)

Table 1. Variable means by subsamples.

	Subsamples				
	(1) Population	(2) Parent, Age 7	(3) Parent, Age 11	(4) Teacher, Age 11	(5) Child, Age 11
Female	0.49	0.50	0.52	0.52	0.54
Birthweight (gr.)	3,498	3,568	3,579	3,597	3,577
Non-western origin	0.11	0.01	0.01	0.01	0.01
Parents' years of schooling	14.59	15.58	15.72	15.92	15.75
Gross household income (1,000 Euro)	40.69	45.86	46.31	46.83	46.26
Mother's age at child birth	30.00	30.57	30.70	30.82	30.72
Father's age at child birth	32.67	32.79	32.87	33.01	32.89
Observations	461,635	33,584	28,919	11,819	26,458

Source: Own calculations on data from Statistics Denmark and from the DNBC. Parental variables are measured in the calendar year before the child was born. Column (1) shows variable means for all children born 1997–2003. Columns (2) to (5) indicate respondent type subsample.

<https://doi.org/10.1371/journal.pone.0220193.t001>

to (5) have remarkably similar means. While the selection into the sample is non-random, as suggested, given at least one SDQ score is observed, there is no selection across subsamples of respondent type.

SDQ and standardized academic tests

We now turn to the relationship between SDQ scores and children’s performance on the standardized academic tests. We focus on tests in reading and mathematics in sixth grade. Fig 1 presents the relationship between children’s SDQ total difficulties scores and their standardized academic test scores.

Fig 1 shows the plots for reading and mathematics test scores, respectively. In these graphs, we exclude total difficulties levels with fewer than 100 observations, as these estimates become very imprecise. (All observations are, however, included in the regressions presented in the following). The dashed line shows the cumulative share of children observed; from this line, we see that 80 percent of the observations have total difficulties scores of nine or less, and 98 percent of the observations have scores of 21 or less.

We first note that for both subjects (reading and mathematics) and all respondent types, the relationships are remarkably linear. Thus, the relationship between total difficulties and standardized academic test scores is very constant, independent of whether we are looking at variation at the lower end of the distribution (few problems) or further to the right in the distribution of the total difficulties score. We also note, however, that some differences appear across respondent types as the total difficulties score increases: The steepest (flattest) slopes are observed for the teacher (child) reported scores, but overall the slopes are very similar across respondent type and child age. These observations are similar for reading and mathematics test scores.

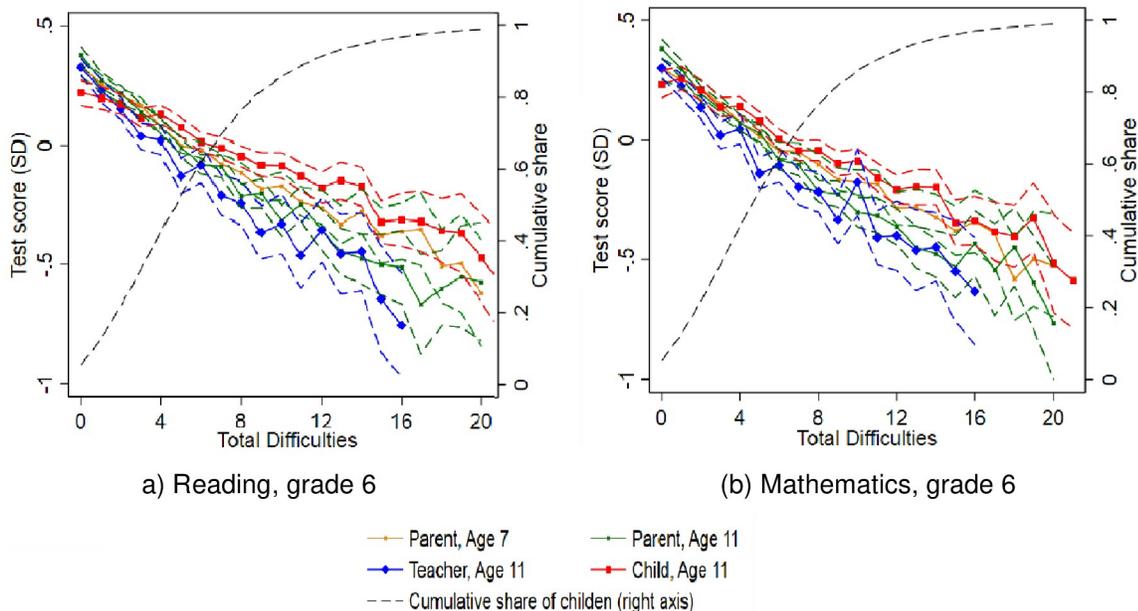


Fig 1. Relationship between SDQ total difficulties and academic test scores in reading and mathematics in sixth grade. Source: Own calculations on data from Statistics Denmark and from the DNBC. Mean academic test scores for each level of the total difficulties scale. Dashed lines indicate pointwise 95 percent confidence bands. The graphs only contain SDQ levels with at least 100 observations, corresponding to at least 98 percent of the sample. The cumulative share of children observed (shown on the right axis) refers to the parent-reported SDQ scores at age 7.

<https://doi.org/10.1371/journal.pone.0220193.g001>

Fig 2 and Fig 3 show the nonparametric relationships for each of the five SDQ subscales for reading and mathematics test scores, respectively.

From Fig 2 and Fig 3 we see that especially the scales for hyperactivity/inattention and conduct problems show very tight and linear relationships between levels of difficulties and

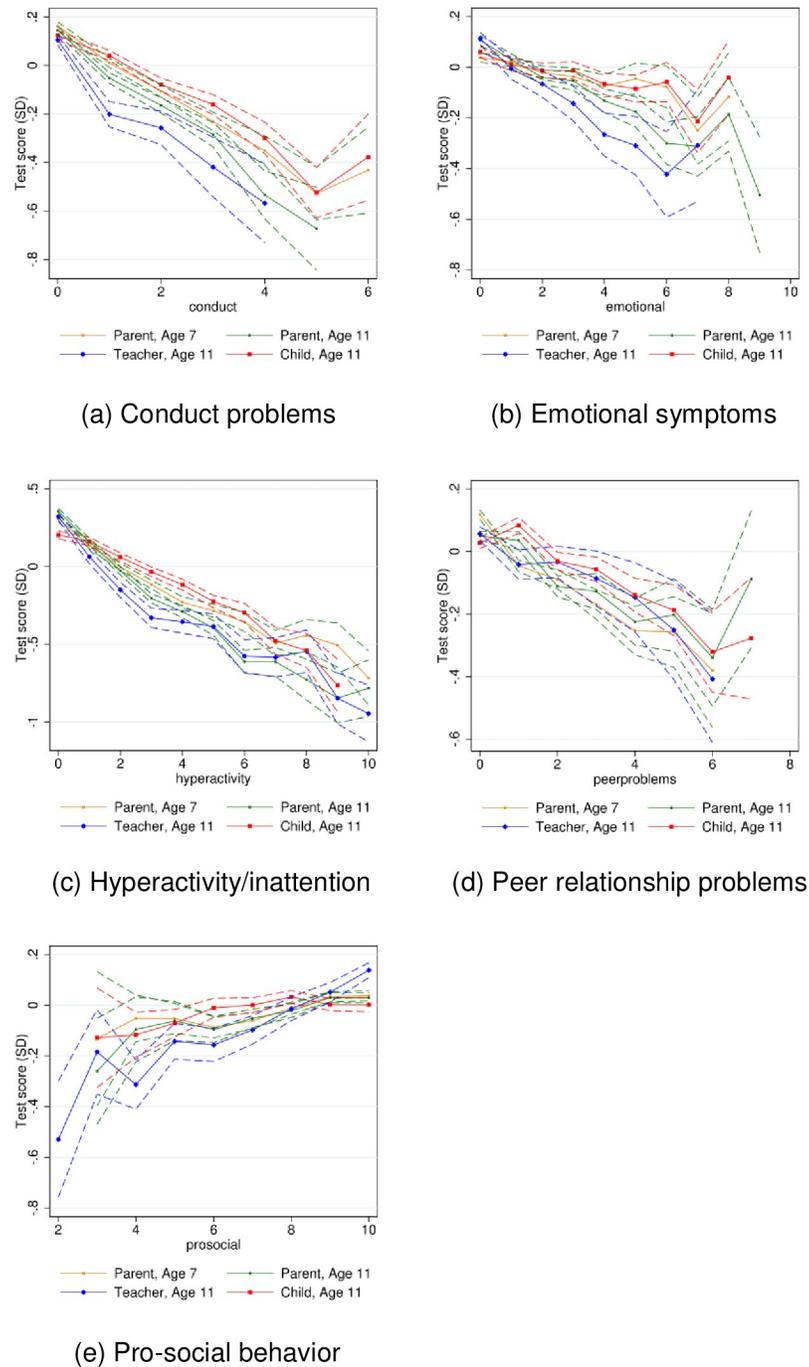


Fig 2. Relationship between SDQ subscales and reading test scores in sixth grade. *Source:* Own calculations on data from Statistics Denmark and from the DNBC. Mean reading test scores for each level of the SDQ subscale. Dashed lines indicate pointwise 95 percent confidence bands. The graphs only contain SDQ levels with at least 100 observations.

<https://doi.org/10.1371/journal.pone.0220193.g002>

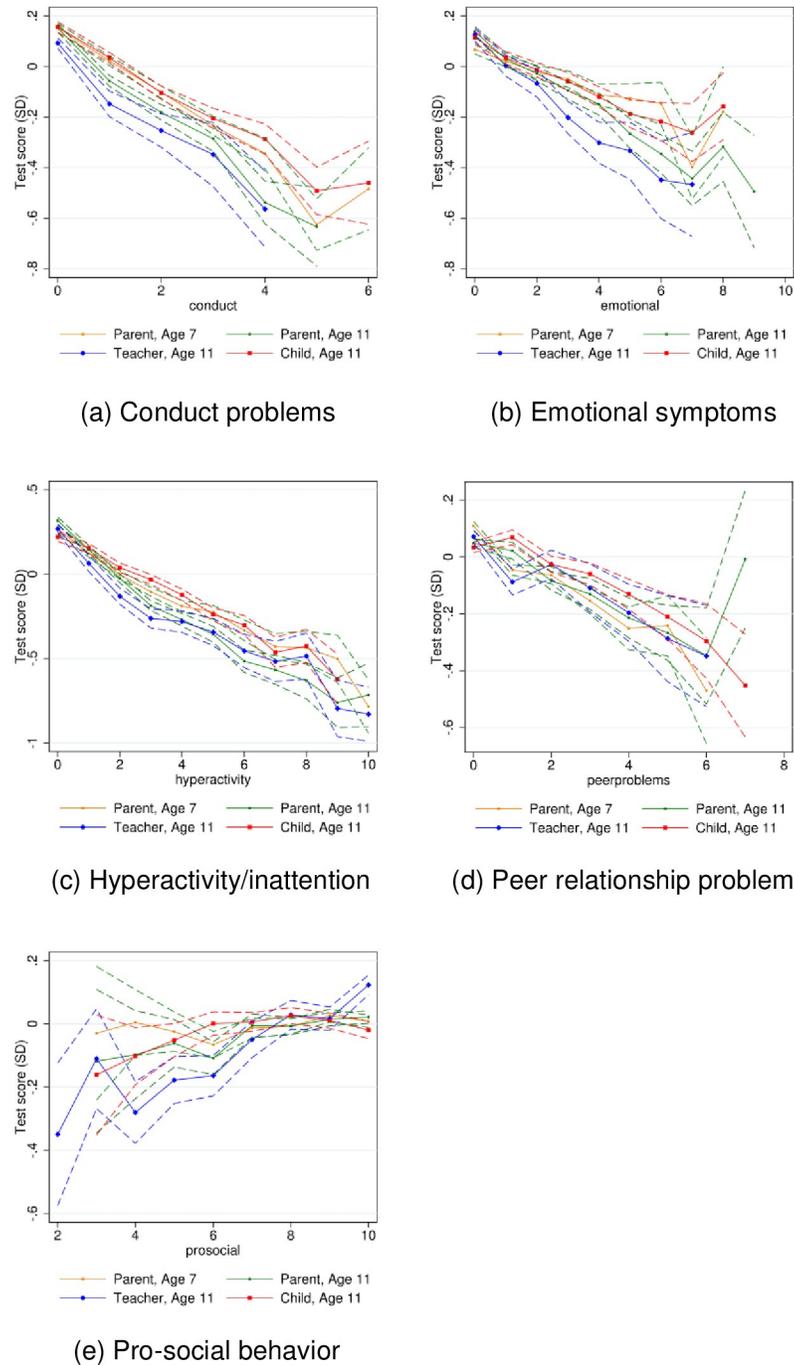


Fig 3. Relationship between SDQ subscales and mathematics test scores in sixth grade. *Source:* Own calculations on data from Statistics Denmark and from the DNBC. Mean mathematics test scores for each level of the SDQ subscale. Dashed lines indicate pointwise 95 percent confidence bands. The graphs only contain SDQ levels with at least 100 observations.

<https://doi.org/10.1371/journal.pone.0220193.g003>

academic test performance, across all respondent types. Again, although the slopes are very similar across respondent type and child age, the flattest slopes are observed for the child reported SDQ.

Table 2. Regression results for SDQ total difficulties. Dependent variable: Academic test score in reading (A) and mathematics (B) in sixth grade.

	(1) Parent, Age 7	(2) Parent, Age 11	(3) Teacher, Age 11	(4) Child, Age 11
<i>A. Reading</i>				
Total difficulties, unconditional	-0.041** (0.001)	-0.050** (0.001)	-0.050** (0.002)	-0.030** (0.001)
Total difficulties, conditional	-0.027** (0.001)	-0.038** (0.001)	-0.042** (0.002)	-0.021** (0.001)
Observations	33,167	28,563	11,598	26,142
<i>B. Mathematics</i>				
Total difficulties, unconditional	-0.043** (0.001)	-0.053** (0.001)	-0.051** (0.002)	-0.036** (0.001)
Total difficulties, conditional	-0.032** (0.001)	-0.042** (0.001)	-0.046** (0.002)	-0.027** (0.001)
Observations	33,079	28,483	11,550	26,087

Source: Own calculations on data from Statistics Denmark and from the DNBC.

* p < 0.1

** p < 0.05.

Standard errors are clustered on the school level. Point estimates from separate regressions. In Panel A (B), the dependent variable is the sixth grade reading (mathematics) test score. Academic test scores are standardized to a mean of zero and a unit standard deviation. The first row in each panel shows point estimates from regressions that only include a constant and the total difficulties score. The second row shows point estimates from regressions where we also control for school fixed effects, parental education, parental income, parental age at child birth, child gender, child birthweight and non-western origin. Columns (1) to (4) indicate respondent types for the included SDQ score.

<https://doi.org/10.1371/journal.pone.0220193.t002>

As the nonparametric analyses clearly indicate a linear relationship between the SDQ and the academic test scores, including the SDQ subscale scores, a linear model is warranted.

Table 2 presents result from estimating a linear model regressing academic test scores on the SDQ total difficulties scores. We present unconditional estimates and estimates conditioned on a rich set of controls using ordinary least squares. (All control variables included in the analyses are presented in Table 1).

Panel A, Table 2 shows point estimates for reading for unconditional regressions (without controls) and for conditional regressions, where we control for school fixed effects, parental education, parental gross income, parental age at child birth, child birthweight, gender, and non-western origin. One additional unit on the total difficulties scale (indicating more difficulties) is associated with three to five percent of a standard deviation lower academic test score, unconditionally, and two to four percent of a standard deviation lower test score, conditionally.

For mathematics (Panel B, Table 2), the coefficients are slightly larger; three to five percent of a standard deviation lower test score, conditionally. Teacher and parent reported SDQ have remarkably similar coefficients for same-aged children (age 11). In line with the presented graphical evidence, the estimates for child reported SDQ scores have the lowest coefficients of the three respondent types.

Table 3 shows regression results by the SDQ subscales conditioned on the control variables.

In line with the graphical evidence (Fig 2 and Fig 3), the relationship between SDQ scores and academic test performance presented in the regression estimates (Table 3) is strongest for conduct problems and hyperactivity/attention, whereas the relationship for prosocial behavior seems the weakest.

Table 3. Regression results for each SDQ subscale. Dependent variable: Academic test score in reading (A) and mathematics (B) in sixth grade.

	(1) Parent, Age 7	(2) Parent, Age 11	(3) Teacher, Age 11	(4) Child, Age 11
<i>A. Reading</i>				
Conduct problems	-0.079** (0.004)	-0.102** (0.005)	-0.120** (0.008)	-0.068** (0.004)
Emotional symptoms	-0.002 (0.003)	-0.034** (0.003)	-0.063** (0.005)	-0.018** (0.003)
Hyperactivity/inattention	-0.075** (0.003)	-0.111** (0.003)	-0.111** (0.004)	-0.064** (0.003)
Peer relationship problems	-0.039** (0.005)	-0.024** (0.004)	-0.027** (0.006)	-0.020** (0.004)
Pro-social behavior	0.013** (0.003)	0.013** (0.004)	0.037** (0.004)	-0.004 (0.003)
<i>B. Mathematics</i>				
Conduct problems	-0.090** (0.004)	-0.116** (0.005)	-0.132** (0.009)	-0.086** (0.004)
Emotional symptoms	-0.022** (0.003)	-0.050** (0.003)	-0.073** (0.005)	-0.035** (0.003)
Hyperactivity/inattention	-0.077** (0.003)	-0.110** (0.003)	-0.116** (0.004)	-0.071** (0.003)
Peer relationship problems	-0.047** (0.005)	-0.030** (0.005)	-0.039** (0.006)	-0.025** (0.004)
Pro-social behavior	0.013** (0.003)	0.019** (0.004)	0.047** (0.005)	0.002 (0.004)

Source: Own calculations on data from Statistics Denmark and from the DNBC.

* $p < 0.1$

** $p < 0.05$.

Standard errors are clustered on the school level. Point estimates for separate regressions. In Panel A (B), the dependent variable is the sixth grade reading (mathematics) test score. Academic test scores are standardized to a mean of zero and a unit standard deviation. All regressions include school fixed effects and controls for parental education, parental income, parental age at child birth, child gender, birthweight and non-western origin. Columns (1) to (4) indicate respondent types for the included SDQ score.

<https://doi.org/10.1371/journal.pone.0220193.t003>

The results for the parent reported SDQ scores indicate that the association with academic test performance is stronger at age 11 than at age 7. For example, whereas one additional unit higher score for hyperactivity/inattention is associated with eight percent of a standard deviation lower test scores at age 7, it is associated with eleven percent of a standard deviation lower test scores at age 11. This tendency is visible for all subscales except peer relationship problems.

Discussion

Using data from Statistics Denmark and a largescale nation-wide birth cohort, we show that emotional and behavioral problems as measured by the SDQ are closely related to student performance in standardized academic tests of reading and mathematics in sixth grade. The relationship is remarkably linear across the entire distribution and for both the total difficulties scores and individual subscales.

While teacher and parent reported scores show very similar results, the relationship between the child's self-reported SDQ and the academic test performance has a flatter slope. Normative data from various samples show that children report more extreme SDQ scores compared to

both teacher and parent assessments for total difficulties as well as subscales [24,25]. If children report more emotional and behavioral difficulties (and more pro-social strengths), this difference in itself would parallel shift this graph to the right compared to the graph for teacher and parent respondents, though it would not necessarily translate into the flatter slope that we observe. However, this difference between children's self-reported scores and parent or teacher scores is not stable across the distribution, rather the difference increases with higher self-reported scores. For example, our data shows that for those children who report a total difficulties score on the SDQ higher than the median 76 percent of parents report a lower score than their child does, indicating that they assess fewer problems than the child does. In fact, parents in this group on average report a four point lower total difficulties scores than the child self-report. For children who report scores lower than or equal to the median the difference is much smaller; here, parents on average report a 0.2 points higher score than the child and only 40 percent of parents provide lower scores than their child. It appears that for higher self-reported scores (more problems), parents are often more positive than their child and the difference in the assessment of problems quite large, whereas the assessments are much more similar at the lower end of the distribution. We believe that these differences in assessment at the higher end of the distribution explain the differences in the slopes that we observe.

Although the SDQ has been extensively examined for its internal consistency, test-retest reliability, and inter-rater agreement, less is known about how the questionnaire relates to academic performance outcomes such as standardized tests in school. Studies examining this relationship across respondent type and child age are even scarcer. This study adds to previous research by identifying the functional form of this relationship for three types of respondents and two age levels. We show that the variation in the SDQ scores away from the abnormal/borderline vs. normal thresholds (i.e. using SDQ as a continuous variable) is correlated with children's performance in standardized academic tests. The results are remarkably similar for all types of respondents and age levels. To the best of our knowledge, this characteristic of the relationship between the SDQ and standardized academic tests has not been addressed in the literature.

Strengths and limitations

A major strength of this study is the large sample size, which enable the nonparametric analyses of the functional form of the relationship between SDQ and academic test scores. In addition, using data from three respondent types and two age levels enables the comparison across these subsamples, which adds to the level of detail presented in the study.

Limitations of this study include the fact that the study sample consist of a non-random subsample of the population of children born from 1997–2003. The results are therefore not informative about the relationship between emotional and behavioral problems and academic test scores for children excluded from the DNBC due to non-response. Similarly, for children who have not completed the Danish National Tests (including children who do not attend a public school or children exempt from taking the test due to severe mental health problems or learning disabilities), we can only speculate about the relationship between their SDQ scores and academic performance.

Implications

The SDQ is a reliable, valid and widely used instrument for identifying emotional and behavioral problems in primary school children in research and among professionals. Assessing the functional form on both aggregate and subscale levels of the SDQ in terms of academic test scores, this study provides relevant information on the reliability and use of the questionnaire.

Due to limited knowledge about the informative value of the lower end of the SDQ distribution (fewer problems), researchers have so far have been reluctant to use the raw subscale scores. The results of this study imply that the use of SDQ scores as continuous variables is justified and that the lower end of the SDQ scale is informative, even at a subscale level.

Acknowledgments

Sievertsen acknowledges financial support from the Danish Council for Independent Research through grant 09–070295. This paper uses data from the Danish National Birth Cohort (DNBC). The DNBC was established with a significant grant from the Danish National Research Foundation. Additional support was obtained from the Danish Regional Committees, the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Health Foundation and other minor grants. The DNBC Biobank has been supported by the Novo Nordisk Foundation and the Lundbeck Foundation. Follow-up of mothers and children have been supported by the Danish Medical Research Council (SSVF 0646, 271-08-0839/06-066023, O602-01042B, 0602-02738B), the Lundbeck Foundation (195/04, R100-A9193), The Innovation Fund Denmark 0603-00294B (09–067124), the Nordea Foundation (02-2013-2014), Aarhus Ideas (AU R9-A959-13-S804), University of Copenhagen Strategic Grant (IFSV 2012), and the Danish Council for Independent Research (DFF– 4183–00594 and DFF—4183–00152). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: Maria Keilow, Hans Henrik Sievertsen, Janni Niclasen, Carsten Obel.

Formal analysis: Maria Keilow, Hans Henrik Sievertsen, Janni Niclasen.

Methodology: Maria Keilow, Hans Henrik Sievertsen, Janni Niclasen, Carsten Obel.

Project administration: Maria Keilow.

Software: Hans Henrik Sievertsen.

Supervision: Maria Keilow, Hans Henrik Sievertsen, Janni Niclasen, Carsten Obel.

Validation: Maria Keilow, Janni Niclasen.

Visualization: Maria Keilow, Hans Henrik Sievertsen.

Writing – original draft: Hans Henrik Sievertsen, Janni Niclasen.

Writing – review & editing: Maria Keilow, Carsten Obel.

References

1. Achenbach TM, Becker A, Döpfner M, Heiervang E, Roessner V, Steinhausen HC, et al. Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *J Child Psychol Psychiatry Allied Discip.* 2008; 49: 251–275. <https://doi.org/10.1111/j.1469-7610.2007.01867.x> PMID: 18333930
2. Heckman J, Stixrud J, Urzua S. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *J Labor Econ.* 2006; 24: 411–482.
3. Heckman, James J; Rubinstein Y. The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *Am Econ Rev.* 2001; 91: 145–149.
4. Anderson LR. Adolescent mental health and behavioural problems, and intergenerational social mobility: A decomposition of health selection effects. *Soc Sci Med. Elsevier;* 2018; 197: 153–160. <https://doi.org/10.1016/j.socscimed.2017.11.026> PMID: 29245082

5. Gupta ND, Simonsen M. Non-cognitive child outcomes and universal high quality child care. *J Public Econ.* 2010; 94: 30–43.
6. Gupta ND, Simonsen M. The effects of type of non-parental child care on pre-teen skills and risky behavior. *Econ Lett. Elsevier B.V.*; 2012; 116: 622–625. <https://doi.org/10.1016/j.econlet.2012.06.020>
7. Gupta ND, Simonsen M. Academic performance and type of early childhood care. *Econ Educ Rev.* 2016; 53: 217–229. <https://doi.org/10.1016/j.econedurev.2016.03.013>
8. Haskins AR. Paternal incarceration and child-reported behavioral functioning at age 9. *Soc Sci Res.* 2015; 52: 18–33. <https://doi.org/10.1016/j.ssresearch.2015.01.001> PMID: 26004445
9. Kristoffersen JHG, Nielsen HS, Simonsen JM, Visby Kraegpøth M, Skyt Nielsen H, Simonsen M. Disruptive school peers and student outcomes. *Econ Educ Rev. Elsevier Ltd.*; 2015; 45: 1–13. <https://doi.org/10.1016/j.econedurev.2015.01.004>
10. Chor E, Andresen ME, Kalil A. The impact of universal prekindergarten on family behavior and child outcomes. *Econ Educ Rev. Elsevier Ltd.*; 2016; 55: 168–181. <https://doi.org/10.1016/j.econedurev.2016.10.002>
11. López Turley RN, Gamoran A, McCarty AT, Fish R. Reducing children's behavior problems through social capital: A causal assessment. *Soc Sci Res. Elsevier Ltd.*; 2016; 61: 206–217. <https://doi.org/10.1016/j.ssresearch.2016.06.015> PMID: 27886729
12. Kristoffersen JHG, Obel C, Smith N. Gender differences in behavioral problems and school outcomes. *J Econ Behav Organ. Elsevier B.V.*; 2015; 115: 75–93. <https://doi.org/10.1016/j.jebo.2014.10.006>
13. Hernández-Alava M, Popli G. Children's Development and Parental Input: Evidence From the UK Millennium Cohort Study. *Demography. Demography*; 2017; 1–27. <https://doi.org/10.1007/s13524-016-0546-y>
14. Dee TS, Sievertsen HH. The gift of time? School starting age and mental health. *Health Econ.* 2018; 27: 781–802. <https://doi.org/10.1002/hec.3638> PMID: 29424005
15. Stone LL, Otten R, Engels RCME, Vermulst AA, Janssens JMAM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-Year-olds: A review. *Clin Child Fam Psychol Rev.* 2010; 13: 254–274. <https://doi.org/10.1007/s10567-010-0071-2> PMID: 20589428
16. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry.* 2001; 40: 1337–1345. <https://doi.org/10.1097/00004583-200111000-00015> PMID: 11699809
17. Niclasen J, Teasdale TW, Andersen A-MN, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish Strength and Difficulties Questionnaire: the SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS One.* 2012; 7: e32025. <https://doi.org/10.1371/journal.pone.0032025> PMID: 22384129
18. Niclasen J, Skovgaard AM, Andersen AMN, Sørhovd MJ, Obel C. A confirmatory approach to examining the factor structure of the strengths and Difficulties Questionnaire (SDQ): A large scale cohort study. *J Abnorm Child Psychol.* 2013; 41: 355–365. <https://doi.org/10.1007/s10802-012-9683-y> PMID: 23008011
19. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry.* 1997; 38: 581–586. PMID: 9255702
20. Jacobsen TN, Nohr EA, Frydenberg M. Selection by socioeconomic factors into the Danish National Birth Cohort. *Eur J Epidemiol. Springer Netherlands*; 2010; 25: 349–355. <https://doi.org/10.1007/s10654-010-9448-2> PMID: 20349116
21. Beuchert LV, Nandrup AB. The Danish National Tests—A Practical Guide. *Econ Work Pap.* 2014; 25.
22. Olsen J, Melbye M, Olsen SF, Sørensen TIA, Aaby P, Andersen A-MNM, et al. The Danish National Birth Cohort—its background, structure and aim. *Scand J Public Health.* 2001; 29: 300–307. <https://doi.org/10.1177/14034948010290040201> PMID: 11775787
23. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology. Lippincott Williams & Wilkins*; 2006; 17: 413–418. <https://doi.org/10.1097/01.ede.0000220549.14177.60> PMID: 16755269
24. Mellor D, Mellor D. Normative data for the strengths and difficulties questionnaire in Australia. *Aust Psychol.* 2005; 40: 215–222. <https://doi.org/10.1080/00050060500243475>
25. Meltzer H, Gatward R, Goodman R, Ford T. The mental health of children and adolescents in Great Britain. 1999.