White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/161876/

Version: Published Version

Article:
von Stumm, Sophie, Smith-Woolley, Emily, Cheesman, Rosa et al. (2021) School quality ratings are weak predictors of students' achievement and well-being. Journal of Child Psychology & Psychiatry. 13276. pp. 339-348. ISSN: 1469-7610

https://doi.org/10.1111/jcpp.13276

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry **:* (2020), pp **–** doi:10.1111/jcpp.1327

School quality ratings are weak predictors of students' achievement and well-being

Sophie von Stumm,[1] Emily Smith-Woolley,[2] Rosa Cheesman,[3] Jean-Baptiste Pingault,[3,4] Kathryn Asbury,[1] Philip S. Dale,[5] Rebecca Allen,[6] Yulia Kovas,[7,8] and Robert Plomin[3]

[1]Psychology in Education Research Centre, Department of Education, University of York, York, UK; [2]The Physiological Society, London, UK; [3]MRC Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [4]Department of Clinical, Educational & Health Psychology, Division of Psychology & Language Sciences, Faculty of Brain Sciences, University College London, London, UK; [5]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, USA; [6]School of Education, University of Brighton, Brighton, UK; [7]Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Tomsk, Russia; [8]Department of Psychology, Goldsmiths, University of London, London, UK

Background: In England, all state-funded schools are inspected by an independent government agency, the Office for Standards in Education, Children's Services and Skills (Ofsted). Inspections aim to hold schools accountable and to promote the improvement of education, with the results made available to the public. Ofsted reports intend to index school quality, but their influence on students' individual outcomes has not been previously studied. The aim of the current study was to explore the extent to which school quality, as indexed by Ofsted ratings, is associated with students' educational achievement, well-being and school engagement. **Methods:** We use an England population-based sample of 4,391 individuals, for whom school performance at age 11 and GCSE grades at age 16 were accessed from the National Pupil Database, and who completed measures of well-being and school engagement at age 16. **Results:** We found that Ofsted ratings of secondary school quality accounted for 4% of the variance in students' educational achievement at age 16, which was further reduced to 1% of the variance after we accounted for prior school performance at age 11 and family socioeconomic status. Furthermore, Ofsted ratings were weak predictors of school engagement and student well-being, with an average correlation of .03. **Conclusion:** Our findings suggest that differences in school quality, as indexed by Ofsted ratings, have little relation to students' individual outcomes. Accordingly, our results challenge the usefulness of Ofsted ratings as guides for parents and students when choosing secondary schools. **Keywords:** School quality; Ofsted; well-being; educational achievement; school engagement.

Introduction

In England, parents can choose where to send their children to secondary school. To help with this decision-making process, many turn to the reports by the Office for Standards in Education, Children's Services and Skills (Ofsted). Ofsted is an independent government agency whose purpose is to 'inspect and regulate services that care for children and young people' (Ofsted, 2020). The primary aim of these inspections is to drive improvement within schools and hold them to account. School inspections happen once every four years and comprise lesson observations, teacher meetings, paperwork checks and pupil interviews. Once an inspection has been conducted, a school is awarded an overall effectiveness rating that informs parents and the government of the quality of education that pupils attending the school receive. This score falls into one of four categories: 'Outstanding' (21% of schools received this rating in 2018), 'Good' (64%), 'Requires Improvement' (11%) or 'Inadequate' (4%; Institute for Government, 2019). In particular for those schools that are deemed to be 'Outstanding', this rating can act as a marketing tool, driving up interest from parents, students, potential teachers (Waterreus, 2003) and even house prices (Black, 1999; Gibbons & Machin, 2008; Leech & Campos, 2003). In contrast, schools that are judged to be underperforming suffer reputational damage and special measures are taken to improve the school, including the dismissal of senior managers and teaching staff and the replacement of the school governors by an appointed executive committee (Hutchinson, 2016; Roberts, 2019). These schools will also be placed under further, more frequent inspections. Although there is no doubt that Ofsted serves an important function by inspecting and rating schools' quality, it is less clear whether differences between schools in Ofsted ratings are associated with students' educational and social–emotional outcomes.

Ofsted inspections

All state-funded schools in England are inspected by Ofsted. In 2017/18, £44 million was spent on 6,079 school inspections, with an average of £7,200 per school inspection (National Audit Office, 2018). The frequency of visits and the length of inspection depend on the school's existing rating. For example,

Conflict of interest: No conflicts declared.

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

a school judged to be 'Good' at their last inspection will normally receive a one-day short inspection every four years (Ofsted, 2015). At the other end of the rating scale, a school whose overall effectiveness category is judged to be 'Inadequate' will receive more regular inspections and can even be closed down (Ofsted, 2015; Roberts, 2019).

After the inspection, schools receive a detailed report, which includes the overall effectiveness rating (Inadequate, Requires Improvement, Good or Outstanding). This rating is published by Ofsted for each school and publicly available on the Internet. In particular, these reports are deemed useful by parents when deciding where to send their children to secondary school. A survey of 1,000 parents in the United Kingdom found that Ofsted ratings were the third most important factor to parents when choosing a school, after location and suitability to the child's needs (Wespieser, Durbin, & Sims, 2015). A separate report of over 1,000 parents found that Ofsted ratings are the second most important information source for parents choosing schools, after word of mouth from other parents (Ofsted, 2017a).

Ofsted inspections and individual-level outcomes

Why do parents look to Ofsted reports of schools? Because they believe that Ofsted ratings index aspects of school quality that shape students' individual outcomes, including their educational achievement and also their well-being and happiness (Coldron & Boulton, 1991, 1996). But to what extent does the Ofsted rating of a school actually predict such *individual-level* outcomes? Although parents and students evidently want to know if going to a better Ofsted-rated school means higher examination results or greater student well-being, we could not find a single published study looking at the association between school-level Ofsted ratings and individual-level outcomes.

However, several studies have tested associations between individual student outcomes and school quality measured in other ways (Karvonen, Tokola, & Rimpelä, 2018), for example student-rated (Keith & Cool, 1992), parent-rated (Gibbons & Silva, 2011) and teacher-rated (Hoy, Hannum, & Tschannen-Moran, 1998) school quality, as well as more objective measures of school quality, such as pupil–teacher ratio, percentage of teachers with advanced degrees and pupil expenditure (Eide & Showalter, 1998). These studies reported small-to-moderate associations between school quality and pupil outcomes. For example, an analysis of the Trends in International Mathematics and Science Study (TIMSS) showed that class size, teacher education and teacher experience, which are objective markers of school quality, are inconsistently and weakly associated with students' test scores in maths and science across 40 countries (Hanushek & Luque, 2003; see also Hanushek, 1986, for US-focused analyses). With regard to social–emotional outcomes, one study of more than 10,000 pupils from the Longitudinal Survey of Young People in England found that school quality was only weakly associated with pupil happiness and well-being at school (Gibbons & Silva, 2011).

Overall, existing research converges on the conclusion that ratings of school quality tend to inform and dominate parents' perceptions of educational excellence, but they are not strongly associated with students' educational achievement or their enjoyment of the learning environment (Gibbons & Silva, 2011; Kutsyuruba, Klinger, & Hussain, 2015). Here, we explore for the first time whether Ofsted ratings, which intend to index school quality, fit this pattern or whether they capture aspects of school quality that meaningfully adds to pupil's individual outcomes.

The present research

Students are nonrandomly distributed across schools, because parents' choice of school for their children depends on a variety of factors, including personal preferences, resources and schools' reputation. Furthermore, in some cases, schools use students' individual characteristics, such as ability or achievement on school entry examinations, to select their student population. Thus, any observed associations between school quality and pupils' individual outcomes may be attributable to systematic differences between children who attend different schools (i.e. selection biases). To isolate any unique effects of school quality on student outcomes, it is important to account for students' covariates (Karvonen et al., 2018; Rivkin, Hanushek, & Kain, 2005). In the present study, we focus on the influence of Ofsted-rated quality of secondary school on students' educational achievement, well-being and their school engagement, after taking their family background and their prior educational achievement into account. Any remaining differences in achievement gains can be thought of as the school's influence on academic progress or its 'added value'.

We use an England representative sample of 4,391 teenagers for whom independent Ofsted quality ratings of their secondary school were available, as well as extensive information on individual outcomes at age 16 and their academic achievement prior to entering secondary school at age 11. Our primary goal was to investigate whether the overall Ofsted ratings were associated with a range of individual student outcomes, including educational achievement, well-being and school engagement while accounting for differences between students on entry into the school. We predicted significant but weak associations between Ofsted ratings and individual student outcomes, and we expected these associations to reduce substantially when students' prior achievement and family background were considered.

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

Methods

Sample

The sample for this study was drawn from the Twins Early Development Study (TEDS). TEDS is a large, population-based sample of twin pairs born in England and Wales between 1994 and 1996 and followed from birth to the present day (Rimfeld et al., 2019). Ethical approval for this study was received from King's College London Ethics Committee. In the present study, we included 4,391 unrelated individuals (one twin randomly from a pair, to preserve independence of data) who attended state school in England at age 16 years, and for whose schools Ofsted school quality ratings were available. In other parts of the United Kingdom, specifically in Wales, Northern Ireland and Scotland, state schools are also regularly inspected, as are private independent schools across the United Kingdom. However, respective inspection agencies and assessment frameworks, scope and criteria differ between countries.

Participants with severe medical or psychiatric problems or whose mothers experienced severe medical complications during pregnancy were excluded from the analysis. We also excluded those who attended nonmainstream schools, such as special schools for those with learning disabilities. The analysis sample included 2,403 females (55%) and 1,988 males (45%). This discrepancy in gender distribution resulted from boys' or men's greater attrition relative to girls' or women's: the 50% of the boys who were assessed at 18 months (relative to 50% girls) reduced to 49% boys in early childhood, and then to 48% in adolescence, and finally to 45% at age 18 years (Rimfeld et al, 2019), when they were asked to consent to sharing their school's Ofsted rating. Similar gender differences in attrition have been widely observed (Watson & Wooden, 2009). Written informed consent was given for all participants involved. This sample of 4,391 individuals is broadly representative of the United Kingdom's population for education and socioeconomic characteristics (see Table S1).

Measures

Ofsted-rated school quality

Headline quality rating. In the current study, there were 4,391 participants for whom we had the overall Ofsted ratings of the school that they attended at age 16 ('Overall effectiveness: How good is the school?'). Of these, 27% attended an 'Outstanding' school, 47% attended a 'Good' school, 22% attended a 'Requires Improvement' school, and 4% attended a school rated as 'Inadequate'. These statistics were roughly similar to the national percentages previously reported (Ofsted, 2017b). Ofsted reports, which include the overall quality rating, are publicly available on the Internet for all state-funded secondary schools: https://reports.beta.ofsted.gov.uk/. Test–retest reliability of Ofsted ratings is not available; however, in 2015/16, Ofsted carried out inspections on the same schools by different inspectors. Of the 24 schools inspected, inspectors agreed on the outcome in 22 cases (National Audit Office, 2018).

Individual items. Depending on the length of the Ofsted inspection and the risk criteria addressed in their visit, we also had data available on up to 26 individual inspection items, such as 'The extent to which pupils contribute to the school and wider community' and 'The schools capacity for sustained improvement'. The intercorrelations among the 26 individual Ofsted items revealed moderate-to-high associations, with an average correlation of $r = .59$ (see Figure S1). See Table S2 for the individual items, along with their sample sizes, means and standard deviations.

To guide our decision on the most appropriate measure of Ofsted-rated school quality to use, we conducted principal components analysis (PCA) on the 26 individual items (Table S3). The scree plot (Figure S2) and item loadings (Table S4) supported one general 'school quality' principal component, explaining 59% of the variance. The extracted unrotated component correlated highly with all 26 individual items (Figure S2; average $r = .77$), as well as with the Ofsted overall quality rating (Figure S2; $r = .93$). This suggests that the Ofsted overall quality rating captures what is common among the individual items. This result justified our use of the overall quality rating in subsequent analyses in order to maximise the sample size (N of overall quality rating = 4,391; N of Ofsted extracted component, which requires complete data for all items = 1,114).

Outcomes at age 16

Educational achievement. At the end of compulsory education, students in the United Kingdom sit the 'General Certificate for Secondary Education' (GCSE) examinations. Almost all students take the three core subjects: English, mathematics and science. In addition, students take a range of other subjects such as geography, history and art. All subjects were graded from 4 (G, the lowest grade) to 11 (A*, the best possible grade), in line with the GCSE grading system that was in place when the twins were 16 years old (i.e. 2010 to 2012). In the current sample, GCSE results were obtained in three ways: from questionnaires sent via mail; from telephone interviews with twins and their parents; and with data from the National Pupil Database (NPD; https://www.gov.uk/government/collections/national-pupil-database). The NPD is a pupil-level database that matches pupil and school characteristic data to pupil-level attainment in England. GCSE scores from NPD and TEDS correlate at .99; therefore, we used NPD ratings when TEDS data were missing. There were 4,379 students who had GCSE data and Ofsted data.

In the present study, we focused on the three core subjects: English, mathematics and science, which are taken by all students. Because English, mathematics and science grades correlated highly ($r = 0.70$–0.82), we created a GCSE composite requiring at least two grades to be present.

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

Student-reported school engagement. At age 16, students answered seven questionnaires about their experience of school engagement, including teacher–student relations, control over and relevance of schoolwork, peer support for learning, family support for learning, homework behaviour, homework feedback, attitudes to school and peer victimisation. Details of these questionnaires can be found in Appendix S1.

Well-being. At age 16, students also answered six questionnaires relating to their well-being. These questionnaires assessed the following: academic self-concept, future aspirations and goals, life satisfaction in relation to school, subjective happiness, grit and ambition. Details of these questionnaires can be found in the Supplementary Materials.

Student covariates

To estimate the relationship between school quality and pupil outcomes more rigorously, we considered individual characteristics of students as covariates. We selected two covariates that previous studies have shown to be influential on student achievement: family socioeconomic status and prior achievement (Hemmings, Grootenboer, & Kay, 2011; Sirin, 2005; von Stumm, 2017).

Socioeconomic status. A measure of family socioeconomic status was created by calculating the mean of five measures: maternal and paternal education (measured on a scale from 1 to 8, where 1 = no education and 8 = postgraduate qualifications), maternal and paternal occupation (indexed by the Standard Occupational Classification on a scale from 1 to 9, where 1 = elementary administration and service occupations and 9 = managers, directors and senior officials), and maternal age at birth of first child. These measures were collected at first contact, when the study members were on average 18 months old. All measures were standardised to have a mean of 0 and a SD of 1, and at least three measures were required to calculate the arithmetic mean.

Prior achievement. Children's academic performance at the end of primary school, which in the United Kingdom is a different institution than their secondary school, was assessed with a standardised examination at age 11. The examination spans English, mathematics and science tests. We used the 'fine point score' of each of these tests from the NPD (for details on the scoring method, see Department of Education, 2017).

Analysis

Associations between Ofsted ratings and individual outcomes. We calculated Spearman's rank correlation to explore the relationship between the Ofsted overall quality rating and educational achievement, well-being and school engagement measures. In addition to investigating individual differences in outcomes, we also estimated the average differences among students attending schools of different quality using ANOVA with polynomial trend analysis and planned contrasts. Trend analysis tests the relationship between the group means (Inadequate/Requires Improvement/Good/Outstanding) comparing linear, quadratic and cubic trends. A linear trend would suggest a proportionate change in the value of the outcome across ordered categories, for example GCSE scores increasing proportionately across each Ofsted categories (Inadequate/Requires Improvement/Good/Outstanding). By contrast, quadratic and cubic trends suggest that the relationship between outcome measures (educational achievement, well-being and school engagement) and Ofsted-rated school quality changes across the ordered categories of Ofsted school quality.

To test the influence of Ofsted-rated school quality on individual achievement, independent of student characteristics (family socioeconomic status and prior achievement), we conducted regressions and observed the unique variance explained by Ofsted-rated school quality. We also looked at the unstandardised beta coefficients to estimate the average GCSE difference between different Ofsted-rated schools. Finally, we ran ANCOVA to investigate the adjusted means of the Ofsted-rated school quality categories.

Results

Associations between Ofsted ratings and educational achievement

The Ofsted overall quality rating correlated .21 with students' GCSE scores, accounting for 4.4% of the variance. Figure 1 depicts the flow of pupils from the four quality categories to GCSE grades. The figure shows that fewer students in Outstanding schools achieved lower grades as compared to students in schools rated 'Requires Improvement' or 'Inadequate'. Despite the mean differences, what is striking is the variability of GCSE grades obtained by students attending schools of different quality. Each school quality category contains students who achieved a wide mix of grades at GCSE.

Turning to our analyses of means, a linear trend best described the relationship between the Ofsted school quality categories and students' educational achievement ($F = 201.96$, $p = 7.68 \times 10^{-45}$; Table S5). The difference between Inadequate and Requires Improvement schools was a third (.33) of a grade ($t = 3.06$, $p < .05$), which was similar to the difference between Requires Improvement and Good (0.30 of a grade;

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

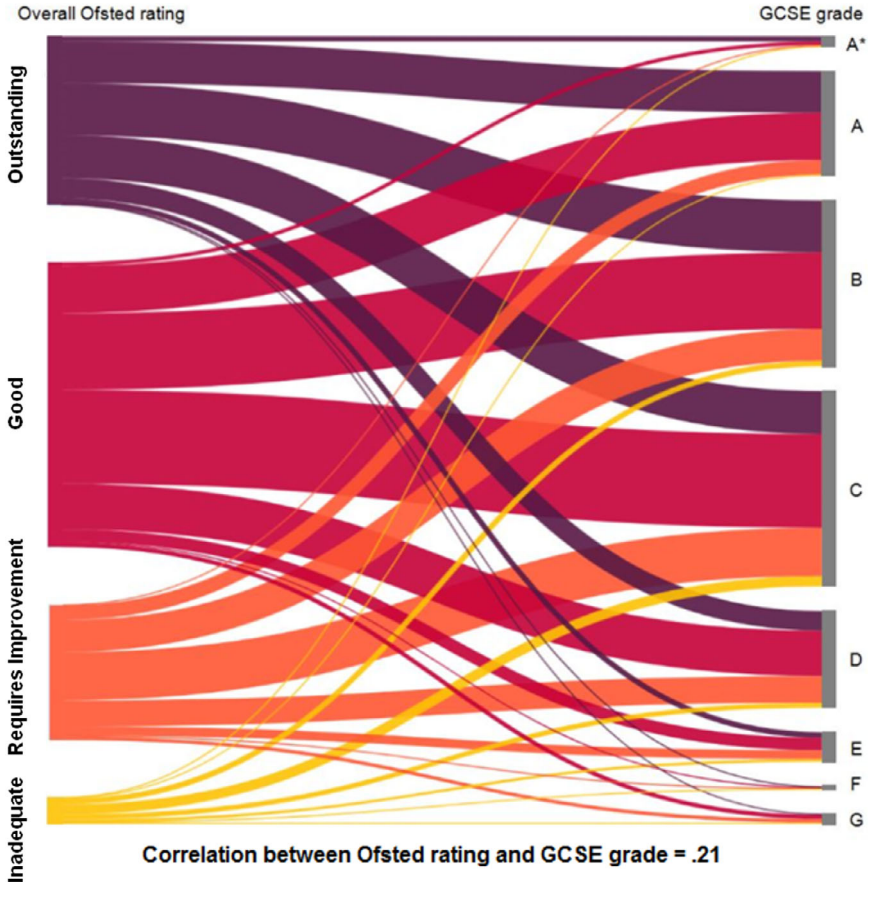

Figure 1 Flow of Ofsted ratings to GCSE grades

$t = 6.35, p < .001$), and to the difference between Good and Outstanding (0.34 of a grade; $t - 7.78$, $p < .001$). The biggest GCSE difference was therefore between those attending Inadequate schools and those attending Outstanding schools, with almost a grade difference (0.94 of a grade; $t = 9.93$, $p < .001$). Students attending Inadequate schools scored on average a GCSE grade of C ($M = 8.17$, $SD = 1.23$), whereas those in Outstanding schools had a mean GCSE grade of B ($M = 9.11$, $SD = 1.20$).

Once we controlled for student covariates, the variance in GCSE predicted by the Ofsted overall quality rating fell from 4.4% to <1% (Table S6). Furthermore, the unstandardised beta associated with the Ofsted overall quality rating ($B = .13$) indicated that the average GCSE difference between the categories (Inadequate/Requires Improvement/Good/Outstanding) was now approximately one tenth of a grade, which was confirmed by the ANCOVA with pairwise comparisons (Table S7). At the extremes, between Inadequate and Outstanding schools, the grade difference was 0.4 ($p = 2.91 \times 10^{-9}$). The GCSE difference between attending an Ofsted-rated 'Good' school (the most common Ofsted category) and an Outstanding school was approximately 0.1 of a GCSE grade ($p = .001$), once student covariates are taken into account. Figure 2 shows the raw and adjusted GCSE means for each Ofsted school quality category.

Associations between Ofsted ratings and students' well-being and school engagement

Spearman's correlations between the Ofsted overall quality rating and the 14 student-reported measures of well-being and school engagement ranged from −.04 (ambition) to .07 (homework behaviour), with an average correlation of .03 (see Figure S3). After correction for multiple testing, only the correlation between Ofsted ratings and homework behaviour remained significant. A series of additional ANOVAs supported these results (see Table S8). Figure 3 depicts the means and 95% confidence intervals for well-being and school engagement for students in schools rated as Inadequate, Requires Improvement, Good and Outstanding. It shows that students attending 'Inadequate'-rated schools reported similar levels of happiness, attitudes to school, homework, student–teacher relations and ambition as those attending 'Outstanding'-rated schools.

Discussion

The purpose of this study was to explore the relationship between school quality as rated by Ofsted and individual-level outcomes for pupils. We found that the Ofsted overall quality rating 'Overall effectiveness: how good is the school' accounted for 4.4% of the differences in educational achievement

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

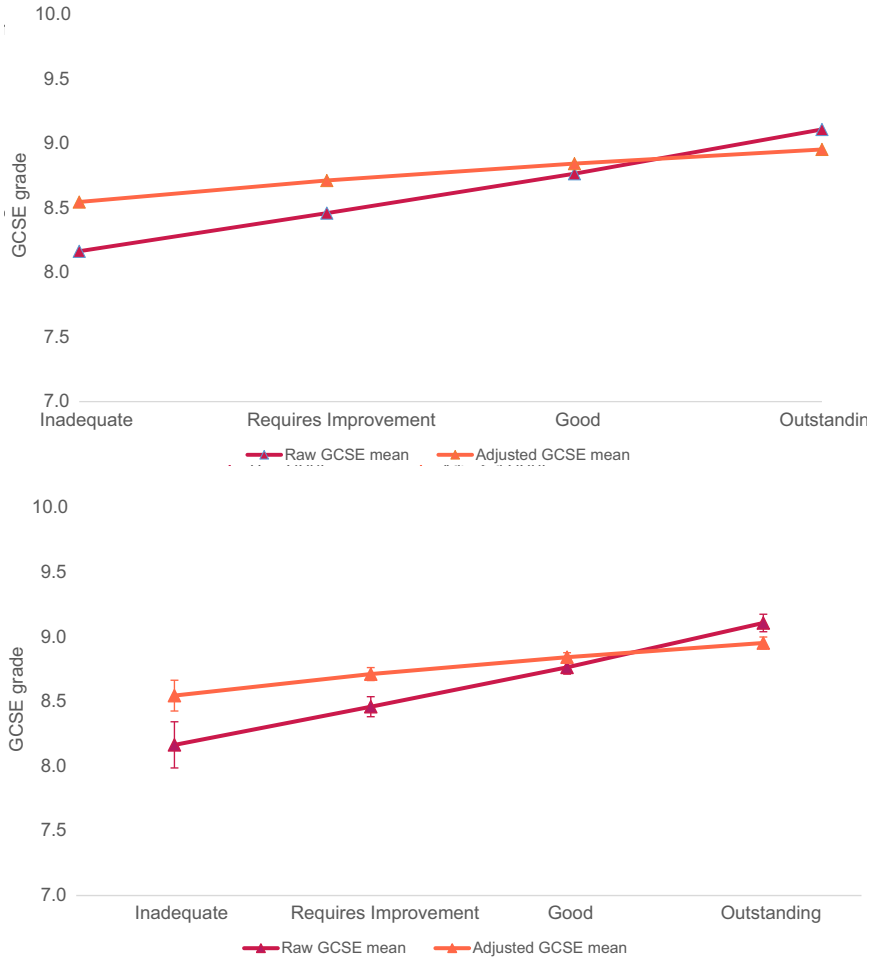

Figure 2 Raw and adjusted GCSE means and 95% confidence intervals. GCSE was graded from 4 (lowest grade: G) to 11 (highest grade: A*)

at age 16. However, most of this association could be attributed to family socioeconomic status and prior achievement in primary school. Once the covariates were included, Ofsted ratings of school quality predicted <1% of the observed differences in GCSE examination grades. This finding suggests that even the small benefits of school quality for students' individual outcomes can be largely attributed to schools' selection of student intake, not to their added value. We also found that Ofsted-rated school quality was a weak predictor of student well-being and school engagement. Overall, our findings suggest that individual student outcomes are largely independent of schools' Ofsted-rated quality. Our findings align with earlier reports that pupils' individual outcomes show little relation to markers of school quality (Gibbons & Silva, 2011; Hanushek, 1986; Hanushek & Luque, 2003).

Ofsted states that their ratings 'allow parents to make informed decisions about where to educate their children' (Ofsted strategy 2017-22, p. 3). Indeed, one of the Ofsted's priorities is to make their reports 'better focused on the issues that parents care about when choosing or seeking assurances about a school' (p9). However, we find that the factors that parents care about most – educational achievement and students' well-being – are negligibly predicted by Ofsted ratings. Pupils' average GCSE difference between schools of varying quality was just a tenth of a GCSE grade. Put another way, attending a 'Good' school over a 'Requires Improvement' school is associated with a GCSE boost of just 0.1 of a grade on average.

By accounting statistically for student covariates, such as prior achievement, in the prediction of GCSEs we generate a proxy of academic progress. Academic progress (referred to as 'Progress 8' by the Department for Education) is calculated as achievement at age 16 independent of previous achievement at 11, and is thought to index value added by schools. In other words, academic progress is students' change (i.e. gains and losses) in school performance between the ages of 11 and 16 years. In the present study, we find that Ofsted-rated quality of a school has little relation to the progress students make during secondary school.

This finding is important for two reasons. First, in a survey of parent views (Ofsted, 2017a), 32% of parents with children aged up to 18 years said that they would want to find out about children's

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

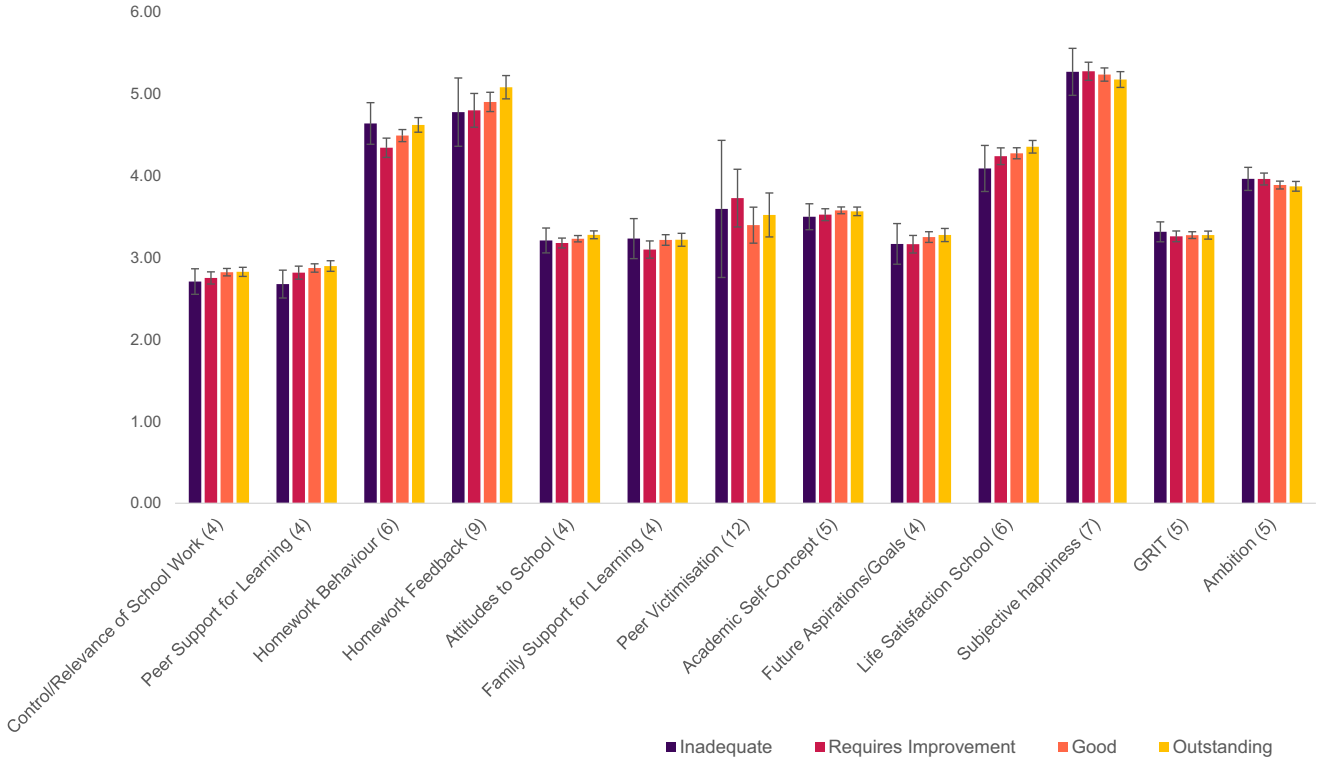

Figure 3 Means and 95% confidence intervals for well-being and school experience measures for students attending schools rated as: Inadequate, Requires Improvement, Good and Outstanding by Ofsted. *Note*. The maximum scores for each of the scales are in brackets

progress in maths at a school when deciding on which school to send their child to. However, if this is weakly predicted at secondary school level by Ofsted ratings, then parents may want to prioritise other factors when choosing a secondary school, for example the physical distance between the family home and the school. Second, it highlights that the examination differences between students attending different Ofsted-rated quality schools are largely accounted for by the school's student population intake: schools with higher Ofsted ratings admit better-performing students (see Hutchinson, 2016). This is in line with previous research suggesting that when schools are responsible for their own admissions, they are more likely to select more able pupils (Rimfeld, et al., 2019; Rivkin et al., 2005; Smith-Woolley et al., 2018; West, 2006).

Although achievement outcomes are important to parents, they are not the only reason why parents opt to send their children to one school over another (Coldron & Boulton, 1991, 1996). The factors most often cited in the literature on parental choice in education are student happiness, well-being and pupil behaviour. In the present study, we find that the correlations between Ofsted ratings and measures of student well-being and school engagement were very small (average $r = .03$) and on average nonsignificant. This suggests school quality, as rated by Ofsted, has little influence on individual-level well-being factors. Put another way, students attending schools with the worst Ofsted ratings report similar levels of happiness, bullying, future aspirations, life satisfaction in relation to school and ambition as those students attending schools with the highest Ofsted ratings. These results are in line with previous research that showed that parent-rated school quality is not strongly associated with pupil happiness and well-being at the school (Gibbons & Silva, 2011).

There are several limitations to our study. First, the present study focused on Ofsted reports of secondary schools only, and we did not consider the impact of school quality at younger ages, when it may be more important. Indeed, a review of primary school quality on educational achievement across 29 countries concluded that the quality of primary schools and that of teachers contributed meaningfully to student achievement, especially in low-income countries (Heyneman & Loxley, 1983). In the present study, we go some way to account for differences between pupils when they enter secondary school by controlling for their prior achievement and family socioeconomic status. However, for testing potential cumulative effects of school quality across primary and secondary education, comprehensive longitudinal research is needed that elucidates the academic trajectories of students as they move through schools of varying quality.

Another limitation of the present study is the lack of objective measures of student well-being and

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

school engagement. Instead, we analysed data from 14 self-report measures. It is possible that students would be happier at different schools; yet, because they only have experience of attending their own school, they lack a comparative perspective. One way to explore this possibility would be to look at students who have attended multiple schools of varying quality and compare their well-being and satisfaction levels at each school. However, these students are often moved for a reason, such as family separation, military deployment, exclusion or bullying, and may not be representative of the wider student population. Indeed, students who switch schools are, on average, from lower income families and have greater behaviour problems and social interaction difficulties (Gasper, DeLuca, & Estacion, 2012; Sorin & Iloste, 2006). Furthermore, Ofsted ratings were only available for the secondary schools that students in our sample attended at age 16, but any secondary school changes that they may have experienced earlier were not recorded.

A final limitation to note is that the current sample was drawn from a twin study. Although we only used one twin from a pair for the current study, being a twin might influence the results. However, our sample appears to be largely representative of the general population for achievement (Table S5) and previous research has shown twins to be broadly representative of the general population for health (Andrew et al., 2001), personality (Johnson, Krueger, Bouchard, & McGue, 2002), psychiatric problems (Kendler, Martin, Heath, & Eaves, 1995), emotional/behavioural problems (Moilanen et al., 1999) and educational achievement (Rimfeld et al., 2019).

Conclusion

In the current study, we find that Ofsted-rated school quality is a weak predictor of secondary school outcomes at age 16, including educational achievement, well-being and school engagement, once schools' student selection criteria have been taken into account. These findings call into question the usefulness of Ofsted ratings as a guide for parents who are looking to make an informed choice for their children's secondary school. Furthermore, our study contests the notion that Ofsted inspections, which are perceived as exhausting, stressful and demoralising by teachers and other school staff (Hopkins et al., 2016; de Wolf & Janssens, 2007), capture differences in school quality that matter for students' individual outcomes.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Supplementary measures.

Table S1. Representativeness of the current sample.

Table S2. Sample sizes, means and standard deviations (SD) for Ofsted ratings.

Table S3. Principal component analysis of Ofsted items.

Table S4. Ofsted individual item loadings.

Table S5. Analysis of variance with polynomial trend analysis and planned contrasts of GCSE scores between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate.

Table S6. Results from multiple regression analysis predicting examination results at age 16 (GCSEs) from student covariates and Ofsted Headline Quality Rating.

Table S7. Analysis of covariance of GCSE scores between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate, accounting for covariates of prior achievement and socioeconomic status.

Table S8. Analysis of variance with polynomial trend analysis of school engagement and wellbeing measures between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate.

Figure S1. Correlation matrix of Ofsted individual items.

Figure S2. Scree plot showing eigenvalues for each principal component after performing PCA on individual Ofsted items.

Figure S3. Spearman correlation coefficients (rs) for the relationship between measures of student wellbeing and engagement and the Ofsted headline quality rating.

Acknowledgements

S.v.S. is supported by a Jacobs Foundation Early Career Fellowship (2017–2019). R.C. is supported by an ESRC studentship. R.P. is supported by a Medical Research Council Professorship award (G19/2). The authors gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a programme grant to R.P. from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938) and the European Commission (602768; 295366). Data for this study came from the Twins Early Development Study (TEDS) and the National Pupil Database. Researchers can apply for access to both (TEDS: https://www.teds.ac.uk/researchers/teds-data-access-policy; NPD: https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts). The authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Sophie von Stumm, Psychology in Education Research Centre, Department of Education, University of York, York, YO10 5DD, UK; Email: sophie.vonstumm@york.ac.uk

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

Key points

- In England, a government agency inspects all state-funded schools to objectively assess differences in the quality of the education that children receive. The inspection reports are widely used by parents to decide which school they will send their children to.
- Our findings show that differences in school quality, as indexed by government inspections, have little influence on students' educational achievement, well-being and school engagement.
- Parents who are looking to make an informed choice for their children's secondary school may be ill-advised to draw conclusions about individual student outcomes based on government school inspection reports.

References

Andrew, T., Hart, D.J., Snieder, H., de Lange, M., Spector, T.D., & MacGregor, A.J. (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Research and Human Genetics*, *4*, 464–477.

Black, S.E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*, 577–599.

Coldron, J., & Boulton, P. (1991). 'Happiness' as a criterion of parents' choice of school. *Journal of Education Policy*, *6*, 169–178.

Coldron, J., & Boulton, P. (1996). What do parents mean when they talk about 'discipline' in relation to their children's schools? *British Journal of sociology of education*, *17*, 53–64.

de Wolf, I.F., & Janssens, F.J.G. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education*, *33*, 379–396.

Department for Education. (2017). National pupil database, key stage 4, Tier 4, 2002–2016: Secure access, 4th ed. UK Data Service. SN: 7627, doi: 10.5255/UKDA-SN-7627-4.

Eide, E., & Showalter, M.H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters*, *58*, 345–350.

Gasper, J., DeLuca, S., & Estacion, A. (2012). Switching schools: Revisiting the relationship between school mobility and high school dropout. *American Educational Research Journal*, *49*, 487–519.

Gibbons, S., & Machin, S. (2008). Valuing school quality, better transport, and lower crime: Evidence from house prices. *Oxford Review of Economic Policy*, *24*, 99–119.

Gibbons, S., & Silva, O. (2011). School quality, child wellbeing and parents' satisfaction. *Economics of Education Review*, *30*, 312–331.

Hanushek, E.A. (1986). The economics of schooling production and efficiency in public schools. *Journal of Economic Literature*, *24*, 1141–1177.

Hanushek, E.A., & Luque, J.A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, *22*, 481–502.

Hemmings, B., Grootenboer, P., & Kay, R. (2011). Predicting mathematics achievement: The influence of prior achievement and attitudes. *International Journal of Science and Mathematics Education*, *9*, 691–705.

Heyneman, S.P., & Loxley, W.A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *American Journal of Sociology*, *88*, 1162–1194.

Hopkins, E., Hendry, H., Garrod, F., McClare, S., Pettit, D., Smith, L. … & Temple, J. (2016). Teachers' views of the impact of school evaluation and external inspection processes. *Improving Schools*, *19*, 52–61.

Hoy, W.K., Hannum, J., & Tschannen-Moran, M. (1998). Organizational climate and student achievement: A parsimonious and longitudinal view. *Journal of School Leadership*, *8*, 336–359.

Hutchinson, J. (2016). School inspection in England: Is there room to improve? Report. Education Policy Institute.

Institute for Government. (2019). Available from: https://www.instituteforgovernment.org.uk/explainers/schools-10-key-facts [last accessed 07 October 2019].

Johnson, W., Krueger, R.F., Bouchard, T.J., & McGue, M. (2002). The personalities of twins: Just ordinary folks. *Twin Research and Human Genetics*, *5*, 125–131.

Karvonen, S., Tokola, K., & Rimpelä, A. (2018). Wellbeing and academic achievement-differences between schools from 2002 to 2010 in the Helsinki Metropolitan Area. *Journal of School Health*, *88*, 821–829.

Keith, T.Z., & Cool, V.A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly*, *7*, 207–226.

Kendler, K.S., Martin, N.G., Heath, A.C., & Eaves, L.J. (1995). Self-report psychiatric symptoms in twins and their nontwin relatives: Are twins different? *American Journal of Medical Genetics*, *60*, 588–591.

Kutsyuruba, B., Klinger, D.A., & Hussain, A. (2015). Relationships among school climate, school safety, and student achievement and well-being: A review of the literature. *Review of Education*, *3*, 103–135.

Leech, D., & Campos, E. (2003). Is comprehensive education really free?: A case-study of the effects of secondary school admissions policies on house prices in one local area. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *166*, 135–154.

Moilanen, I., Linna, S.-L., Ebeling, H., Kumpulainen, K., Tamminen, T., Piha, J. … & Almqvist, F. (1999). Are twins' behavioural/emotional problems different from singletons'? *European Child and Adolescent Psychiatry*, *8*, S62–S67.

National Audit Office. (2018). *Ofsted's inspection of schools*. London UK: National Audit Office.

Ofsted. (2015). *School inspection handbook: Handbook for inspecting schools in England under section 5 of the education act 2005*. Manchester: Ofsted.

Ofsted. (2017a). *Annual parents survey 2017: Parents' awareness and perceptions of Ofsted*. London: Office for Standards in Education, Children's Services and Skills.

Ofsted. (2017b). *The annual report of her majesty's chief inspector of education, children's services and skills 2016/17*. London, UK: Office of Standards in Education.

Ofsted (2020). https://www.gov.uk/government/organisations/ofsted/about. [last accessed 28 March 2020].

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

Ofsted strategy (2017–22). The education inspection framework. May 2019, No 190015. Available from https://www.gov.uk/government/publications/ofsted-strategy-2017-to-2022. [last accessed 28 March 2020].

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., … & Plomin, R. (2019). Twins Early Development Study: A genetically sensitive investigation into behavioural and cognitive development from infancy to emerging adulthood. *Twin Research and Human Genetics*. https://doi.org/10.31234/osf.io/xqh52.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417–458.

Roberts, N. (2019). School inspections in England: Ofsted. House of Commons Library, Briefing Paper, Number 07091, 5 August 2019.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453.

Smith-Woolley, E., Pingault, J.-B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S. … & Plomin, R. (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *NPJ Science of Learning*, *3*, 3.

Sorin, R., & Iloste, R. (2006). Moving schools: Antecedents, impact on students and interventions. *Australian Journal of Education*, *50*, 227–241.

von Stumm, S. (2017). Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. *Intelligence*, *60*, 57–62.

Waterreus, J.M. (2003). *Lessons in teacher pay: Studies on incentives and the labor market for teachers*. Amsterdam, the Netherlands: University of Amsterdam.

Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In: R.M. Groves, G. Kalton, J.N.K. Rao, N. Schwarz, C. Skinner, & P. Lynn, eds. *Methodology of longitudinal surveys*. https://doi.org/10.1002/9780470743874.ch10.

Wespieser, K., Durbin, B., & Sims, S. (2015). *School choice: The parent view*. Slough, UK: National Foundation for Educational Research.

West, A. (2006). School choice, equity and social justice: the case for more control. *British Journal of Educational Studies*, *54*, 15–33.

Accepted for publication: 6 May 2020

© 2020 The Authors. Journal of Child Psychology and Psychiatry published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.