

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/121371/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Wilkins, David and Antonopoulou, Vivi 2019. Ofsted and Children's Services: What performance indicators and other factors are associated with better inspection results? (Corrigendum). *The British Journal of Social Work* 10.1093/bjsw/bcy134

Publishers page: <https://doi.org/10.1093/bjsw/bcy134>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Ofsted and Children's Services: What Performance Indicators and Other Factors Are Associated with Better Inspection Results?

David Wilkins and Vivi Antonopoulou

Abstract

'Failing' an inspection of The Office for Standards in Education, Children's Services and Skills (Ofsted) has severe consequences on a local authority. Senior managers may lose their jobs and the workforce as a whole can be destabilised. In extreme cases, central government can decide whether the authority is no longer capable of running children's services. On the other hand, receiving positive Ofsted judgements often brings with it a national reputation for excellence. This study reports the findings of an analysis of key performance indicators, expenditure and deprivation in relation to Ofsted inspections for eighty-seven local authorities in England undertaken between 2014 and 2016. Our aim was to examine the association between these factors and Ofsted judgements. Our findings suggest that for most of the factors we considered, there is no clear pattern of better or worse performance between local authorities with different Ofsted ratings. However, 'good and outstanding' authorities tend to outperform other authorities in relation to some procedural variables. By itself, the level of local authority deprivation was most clearly associated with the Ofsted rating, and expenditure was associated with the authority's level of deprivation but not their Ofsted judgement. Comparisons are made with the concept of 'value-added' performance in relation to schools.

Keywords: Child protection, children and families, organisation, social policy

Introduction

The Office for Standards in Education, Children's Services and Skills (Ofsted) is responsible for the inspection of children's services in England. The aim of these inspections is to 'raise standards and improve lives' (Ofsted, 2014, p. 5). Yet, in the past Ofsted have been criticised for focusing too much on compliance with policies and procedures and not enough on 'the experiences of children, young people and their families' (Munro, 2011, p. 7). In response to these criticisms, Ofsted revised their inspection framework to focus more on the quality of practice and outcomes for children and less on compliance. Under this new framework, Ofsted inspected children's services across three areas—Child Protection (CP), Looked After Children (LAC) and Leadership—and provided an overall judgement of 'inadequate', 'requires improvement', 'good' or 'outstanding'. Under this framework,

it has proved more difficult for local authorities to achieve a grade of ‘good’ or ‘outstanding’ than before. In Ofsted’s 2016 annual social care report, the most recently published within the time frame considered for this article (2014– 2016), of the eighty-seven authorities inspected, 2 per cent (n1/42) were graded outstanding, 24 per cent (n1/421) were graded good, 49 per cent (n1/443) were graded as requiring improvement and 24 per cent (n1/421) were found to be inadequate. This compares with 9 per cent outstanding, 69 per cent good and 22 per cent adequate in 2007/2008, during which year no authorities were rated inadequate (Impower, 2015).

This may suggest that ‘the bar has been raised’ for children’s services and yet it has also been argued that these judgements do not take sufficient account of local deprivation and spending levels (Rowlands, 2010; Bywaters et al., 2014a,b; Jones, 2015a). In 2015, the government announced that when children’s services are judged to be ‘inadequate’, they could be subject to mandatory intervention by the Department for Education (DfE) (HM Government, 2015). In some places, such as Doncaster and Slough, responsibility for the provision of children’s services has been removed from the local authority. Even in less extreme circumstances, it remains the case that the impact of a negative Ofsted judgement can still be problematic—‘confidence ... is lost, thresholds become very low ... the workforce implodes and becomes unstable ... there is a heavy dependence on agency workers ... workloads increase ... backlogs of assessment build up, cases are unallocated and corners are cut’ (Jones, 2015b).

Despite the significant influence that Ofsted has on the social care sector (Tilbury, 2004), compared with the education sector, we know relatively little about their inspections of children’s services. There have been numerous studies of Ofsted’s school inspections, including whether inspections help improve exam performance (Rosenthal, 2004)¹, whether they make a difference to General Certificate of Secondary Education (GCSE) results (Shaw et al., 2003)², how schools behave before and after an inspection (Ouston et al., 1997)³ and whether Ofsted’s judgements are reliable (Campbell and Husbands, 2000)⁴. In relation to children’s services, far less has been published. Notable exceptions include a Local Government Association report into whether Ofsted inspections help improve children’s services (Impower, 2015). This report found that when a department is judged to be ‘inadequate’ or ‘in need of improvement’, they tend to suffer ‘an accelerated decline’ (2015, p. 3).

¹ No—in the year of inspection, exam performance was slightly worse than in other years.

² For a small number of schools, inspection was associated with slight improvements. For the majority of schools, inspection did not improve examination achievement.

³ School inspections were found to contribute positively to the process of school improvement.

⁴ The methodology of inspection was found to be insufficiently reliable for the consequences which flow from it.

How do Ofsted inspect local authorities?

We do, however, know what methods Ofsted use when inspecting children's services, because their inspection handbook is a public document (Ofsted, 2017a). It describes a 'combination of case-tracking, case sampling, observations of practice and interviews' (p. 23). The latter may be conducted with social workers, children and young people, parents and carers, foster carers and adopters and senior staff within the authority. Inspectors may also spend time 'shadowing staff in their day-to-day work' and 'observing practice in multi-agency/single agency meetings' (p. 24). This suggests that Ofsted aim to take a holistic approach by gathering data from a variety of sources and then seeking to 'triangulate [the] evidence' (p. 32 and p. 35). This approach has been both commended and critiqued by Munro (2014) who said, 'the new framework is looking at the right aspects of work' (p. 1) but 'the complicated causal links between professional practice and outcomes make it difficult to make judgements about causality' (p. 2). In other words, although the inspection framework sets out what Ofsted inspectors do, it is not clear how they use this data to form a judgement about causal links between what they observe about the service and the quality of practice and resultant outcomes.

What factors are associated with different Ofsted judgements?

The question of what factors are most strongly associated with Ofsted's judgments is critical. To date, we are aware of only one other published study that explored this question in relation to Ofsted's new inspection framework. La Valle et al. (2016, p. 21) looked at eleven performance variables (mostly related to children in care) and found no significant associations between improved performance and better Ofsted judgements. Indeed, two of the best performing authorities—Lambeth and Sandwell—were judged by Ofsted to be 'inadequate' while one of the worst performing authorities—East Sussex—was judged to be 'good'. Similarly, in relation to Ofsted's previous inspection framework, Hood et al. (2016) examined all 152 local authorities in England and found three variables predicted an 'inadequate' judgement—fewer timely assessments, higher re-referral rates and higher proportions of agency workers.

There is a clear need for more research regarding the role of Ofsted in relation to children's services. Such research would help us understand more clearly how Ofsted form their judgements and how inspection activity might be improved. With this article, we aim to contribute to this debate by exploring which of the variables we considered, if any, were associated with Ofsted's judgements.

Aims and objectives

In this study, we sought to address the following questions—(i) What is the association between selected local authority performance variables, expenditure, deprivation and Ofsted judgements? and (ii) Can these associations explain differences in obtained Ofsted judgements?

Method

We explored the relationship between a selection of key variables and Ofsted judgements for eighty-seven English local authorities inspected between 2014 and 2016 (at the time of data collection, this represented every authority inspected under that framework). Each of the key variables relates to one of three areas—finance and spending, levels of deprivation and performance. All the data are publicly available.⁵

Measures of Ofsted’s judgements

Ofsted inspection data for each local authority were retrieved from Ofsted’s website. We started by considering the association between the overall Ofsted judgement and each of the sub-category judgements (CP, LAC and Leadership). As the correlations were so high, for the rest of our analysis we used only the overall Ofsted judgement for each authority (Table 1).

	Ofsted child protection judgement	Ofsted looked after children judgement	Ofsted leadership judgement	Ofsted overall judgement
Ofsted overall judgement	.904**	.822**	.929**	
Ofsted child protection judgement		.694**	.883**	.904
Ofsted looked after children judgement	.694**		.835	.822**
Ofsted leadership judgement	.883**	.835**		.929**

Table 1. Spearman’s rho correlation matrix for Ofsted overall and sub-category judgements.

***p is significant at the 0.01 level (2-tailed).*

⁵ Spending, finance and key performance data from the DfE; deprivation data from the Department for Communities and Local Government and inspection data from Ofsted.

We then placed each authority into a category—low, medium or high—depending on their overall judgement. Local authorities with an ‘inadequate’ judgement were placed into the low category (n 1/4 22). Local authorities ‘in need of improvement’ were placed into the medium category (n1/444). As there were relatively few high-performing authorities at the time of our study, those judged ‘good’ or ‘outstanding’ were placed together into the high category (n 1/4 21).

Measures of spending

Obtaining an accurate estimate of local authority spending is difficult, not least because local authorities report spending in different ways (Bywaters et al., 2015, 2017). For this study, we used local authority budget returns⁶ to obtain a total spending figure for children’s services in the financial year of their most recent Ofsted inspection. These figures were adjusted for inflation to 2015/2016 levels using the retail price index. We divided this figure by the local child population at the time of the Ofsted inspection (using Ofsted’s figures), resulting in a ‘spending per child’ estimate for each authority.

There are clear limitations to this approach, not least because not all authorities report spending data in the same way, but authorities target their spending more heavily on some children than others. In addition, although it is relatively easy to know how many ‘children in need’ and ‘children in care’ there are in each authority, the breadth of local authority spending on children and families is far wider, including social work assessments (at least some of which will conclude the child was not ‘in need’), early years support, children’s centres and Sure Start, youth services and family support services. Obtaining reliable usage figures for each of these services in each authority is practically very difficult.

Thus, we acknowledge the limitations in our approach and accept that a nominal ‘spending per child’ figure may not represent actual spending patterns (although it is worth noting that similar figures have been used by other researchers in the field; Bywaters et al., 2017).

Measures of deprivation

Using the English Indices of Deprivation 2015, we ranked each local authority from least deprived (Woking) to most deprived (Manchester). Although ranking at the level of local authority obscures within- authority variation, as with spending data, this approach provides at least an indication of

⁶ Under section 251 of the Apprenticeship, Skills, Children and Learning Act 2009, local authorities are required to submit to central government statements about their planned and actual expenditure on education and children’s social care. These statements are available via the Department for Education website.

the relative position of each authority. We removed authorities not included in our sample and re-ranked the remaining authorities from one to eighty-seven.

Key variables in relation to children protection and children looked after

Finally, we obtained key variables in relation to CP and LAC services for each local authority for the year immediately prior to their most recent Ofsted inspection (Table 2).

Children protection – key performance indicators	<ol style="list-style-type: none"> 1. Referral rate per 10,000 2. Rate of repeat referrals (%) 3. Assessment rate per 10,000 4. Assessments overdue (%) 5. Children in need per 10,000 6. Initial child protection conferences overdue (%) 7. Child protection review conferences overdue (%) 8. Statutory child protection visits overdue (%) 9. Child in need and child protection plans lasting longer than 2 years (%) 10. Rate of repeat child protection plans (%)
Looked after children – key performance indicators	<ol style="list-style-type: none"> 1. Looked after children per 10,000 2. Looked after children without 5 or more ‘good’ GCSEs (%) 3. Looked after children placed more than 20 miles from home (%) 4. Looked after children ‘persistently absent’ from education (%) 5. Looked after children with a recorded episode of ‘missing from care’ (%)

Table 2. Details of key performance data obtained for each local authority.

These data were obtained from the DfE website. Some of these indicators relate to procedural outputs, things ‘produced’ by the local authority—e.g. the proportion of assessments completed on time or out of time. Some of the indicators relate to the quality of practice or perhaps even child and family outcomes— e.g. the rate of repeat referrals and the proportion of children in care who did not achieve ‘good’ GCSE results. We chose these indicators based on our own judgement about their likely significance, and because many of them are routinely discussed and cited in the wider literature (Bilson and Martin, 2017). This means there are inevitably other indicators that we did not include but might have—e.g. the number of special guardianship orders). Expenditure and deprivation, we considered to be structural variables, related not to procedure or practice but to the wider context and environment within which local authorities operate. Table 3 shows the percentage of authorities in each of our Ofsted categories (‘inadequate’, ‘in need of improvement’ and ‘good and outstanding’) compared to high-, medium- and low levels of deprivation.

		LAC with				
	Ofsted rating	Repeat Referrals (%)	Assessments Overdue (%)	missing episode (%)	Levels of deprivation	Spend per child
Ofsted rating		-.064	-.224*	.519**	-.245*	-.146
Sig. (2-tailed)		.558	.043	.000	.022	.179
N		87	82	45	87	87
Repeat Referrals (%)	-.064		.089	-.070	-.321**	-.045
Sig. (2-tailed)	.558		.427	.649	.002	.678
N	87		82	45	87	87
Assessments Overdue (%)	-.224*	.089		-.195	.069	.021
Sig. (2-tailed)	.043	.427		.200	.540	.848
N	82	82		45	82	82
LAC with missing episode (%)	.519**	-.070	-.195		-.154	.048
Sig. (2-tailed)	.000	.649	.200		.313	.753
N	45	45	45		45	45
Levels of Deprivation	-.245*	-.321**	.069	-.154		.655**
Sig. (2-tailed)	.022	.002	.540	.313		.000
N	87	87	82	45		87
Spend per child	-.146	-.045	.021	.048	.655**	
Sig. (2-tailed)	.179	.678	.848	.753	.000	
N	87	87	82	45	87	

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3. Statistically significant correlations between Ofsted scores categories and key factors.

Associations between Ofsted judgements and key variables

Spearman's rho non-parametric correlational analyses were performed to establish the pattern of relationships between Ofsted judgements and the key variables we included. Significant correlations were found between Ofsted judgements and levels of deprivation (Spearman's $\rho = -0.245$, $p = 0.022$), the proportion of assessments overdue (Spearman's $\rho = -0.224$, $p = 0.043$), and the proportion of children in care with 'missing episodes' (Spearman's $\rho = 0.519$, $p < 0.001$). Table 4 depicts the statistically significant associations.

Ofsted categories	Repeat referrals (%)			CIN plans > 2 years (%)			CP plans > 2 years (%)			Repeat CP plans (%)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Inadequate	22	23.2	7.54	22	32.0	8.02	17	2.9	1.84	21	16.0	4.88
In need of improvement	44	21.7	7.05	44	30.7	6.67	28	2.7	2.60	44	15.0	4.73
Good and outstanding	21	21.2	5.14	21	33.0	6.69	17	2.1	2.36	21	17.7	5.56
Total	87	22.0	6.73	87	31.6	7.01	62	2.6	2.34	86	15.9	5.04

Table 4. Mean scores for Ofsted scores categories and CP 'practice' variables.

Comparison of means between Ofsted categories

We calculated the overall mean between our three Ofsted categories for a selection of key variables. First, we calculated the mean per Ofsted category for a group of 'practice' variables—repeat referrals, Child in Need (CIN) and CP plans lasting longer than two years, repeat CP plans (where the same child is subject of a CP plan more than once), children in care persistently absent from school, children in care with a missing episode and children in care without five 'good' GCSEs. In England, CIN and CP plans are used to coordinate support for children and their families. The key difference between them is that CIN plans are used when the child is 'in need' of additional support and are legally voluntary (they cannot be implemented without parental consent), whereas CP plans are used when the child is at risk of significant harm and can be implemented (if not entirely effectively) without parental consent. For these variables, a lower mean indicates better performance (Tables 5 and 6). To take one example, for the group of 'inadequate' authorities we considered, they had a mean of 23.2 per cent repeat referrals. This compares with figures of 21.7 per cent in 'in need of improvement' authorities and 21.2 per cent in 'good and outstanding' authorities. (The figures for standard deviation give an indication of the range of performance in each of these groups.)

Secondly, we calculated the mean per category for a group of procedural variables — Initial Child Protection Conferences and Review Child Protection Conferences overdue, CP visits overdue and Assessments overdue. For these variables, a lower mean indicates better performance (Table 7).

Ofsted Categories	CLA persistently absent from education (%)			CLA with missing episode (%)			CLA without 5 'good' GCSEs (%)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Inadequate	17	6.7	6.14	14	4.8	2.89	13	79.4	6.33
In need of improvement	29	5.3	1.99	18	6.2	3.19	18	77.5	5.19
Good and outstanding	16	4.8	1.95	13	9.0	1.78	13	83.3	3.64
Total	62	5.5	3.63	45	6.6	3.17	44	79.8	5.62

Table 5. Mean scores for Ofsted scores categories and CLA 'practice' variables.

	ICPCs overdue (%)			Review conferences overdue (%)			CP visits overdue (%)			Assessments overdue (%)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Inadequate	21	30.1	17.41	22	5.1	5.81	20	39.1	24.85	21	24.4	15.93
In need of improvement	43	30.5	21.04	44	4.8	5.28	40	44.4	29.68	41	24.2	16.42
Good and outstanding	18	24.5	15.97	21	2.8	3.67	20	34.4	31.92	20	15.0	9.69
Total	82	29.1	19.08	87	4.4	5.12	80	40.6	29.09	82	22.0	15.29

Table 6. Mean scores for Ofsted scores categories and CP 'procedural' variables.

Ofsted Ranking	Deprivation levels	N	Mean for spending per child	SD
Inadequate	low deprivation	6	734.78	92.82
	medium	8	850.62	103.75
	high deprivation	8	981.79	206.73
Total average			£855.73	134.43
In need of improvement	low deprivation	10	639.43	102.46
	medium	15	921.99	157.67
	high deprivation	19	973.65	204.12
Total average			£845.02	154.75
Good and outstanding	low deprivation	13	658.79	90.82
	medium	6	964.88	305.94
	high deprivation	2	1102.81	94.83
Total average			£908.82	163.86

Table 7. Mean spending per child for each OFSTED category and deprivation group.

Analysis of variance (ANOVA) tests were carried out to examine these mean scores and a Bonferroni correction was applied when multiple comparisons were made. Significant differences were found

for two of the variables—the proportion of children in care without five ‘good’ GCSEs ($F(2, 41) = 4.86, p = 0.013, g^2 = 0.198$), and the proportion of children with ‘missing episodes’ from care ($F(2, 42) = 8.12, p = 0.001, g^2 = 0.279$). Post hoc comparisons using Bonferroni correction indicated that the mean score for children in care without five GCSEs for the ‘in need of improvement’ category ($M = 477.5, SD = 5.19$) was significantly different than for the good/outstanding category ($M = 483.3, SD = 3.63$)—average performance in the ‘in need of improvement’ category was better than the ‘good/outstanding’ category. However, the ‘inadequate’ category ($M = 479.42, SD = 6.33$) did not significantly differ from the ‘in need of improvement’ or ‘good/outstanding’ categories ($p > 0.05$). Similarly, post hoc comparisons using Bonferroni correction indicated that the mean score for missing episodes for the ‘inadequate’ category ($M = 44.78, SD = 42.88$) and the ‘in need of improvement’ category ($M = 46.22, SD = 43.19$), were significantly different from the ‘good/outstanding’ category ($M = 49, SD = 1.78$)—authorities in the two former categories had fewer children with recorded ‘missing from care’ episodes than those in the ‘good/outstanding’ category.

Ofsted categories and spending

We calculated the overall mean ‘spend per child’ for our three different Ofsted categories. The figure for all eighty-seven authorities was £854.60. ANOVA between each of our Ofsted categories did not produce any statistically significant differences on expenditure ($p > 0.05$). When levels of deprivation are taken into account, an ANOVA yielded significant differences between local authorities in terms of spend per child, for the ‘inadequate category’ ($F(2, 21) = 4.83, p = 0.02$); for the ‘in need of improvement’ category ($F(2, 43) = 13.26, p = 0.000$); and for the ‘good and outstanding’ category, ($F(2, 20) = 9.42, p = 0.002$). Post hoc Tukey Honestly Significant Differences (HSD) tests showed that spend per child was significantly different between the low- and high- deprivation groups ($p < 0.05$); whereas the medium-deprivation group was not significantly different from the other two groups (Table 8).

	B	SE (B)	e^B	CI(95%)	P-value
Deprivation levels	-.031	.014	.970	(.994,.996)	.026*
Assessments Overdue	-.055	.026	.947	(.899,.997)	.037*
LAC with missing episode	.583	.190	1.792	(1.235, 2.598)	.002**
Total Spend per child	-.002	.002	.998	(.995, 1.002)	.208

Reference category was “in need for improvement” Ofsted category.

*. Correlation is significant at the 0.05 level (2

**. Correlation is significant at the 0.01 level (2-tailed).

Table 8. Bivariate models of Ofsted judgement

Associations with Ofsted category ratings

Logistic regression analyses were conducted to test the association between all the variables we considered and membership of our three Ofsted categories. The level of deprivation (B 1/4 -0.031, Exp b 1/4 0.970, p 1/4 0.026), proportion of assessments overdue (B 1/4 -0.055, Exp b 1/4 0.947, p 1/4 0.037) and proportion of children with ‘missing episodes’ from care (B 1/4 0.583, Exp b 1/4 1.792, p 1/4 0.002) were found to be significant predictors of membership in the ‘good and outstanding’ category (see Table 9).

	$\chi^2(df)$	B	SE (B)	e ^B	CI(95%)	P-value
Full model	16.38 (4)	-	-	-	-	.003**
Predictors						
Deprivation levels		-.030	.015	.971	(.944, .999)	.043*
Assessments Overdue	-	-.059	.028	.943	(.893,.995)	.033*

*. Correlation is significant at the 0.05 level (2

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9. Multivariate model predicting Ofsted category judgement

Spend per child was not found to be associated with Ofsted category judgements ($p > 0.05$). Thus, the odds of a Local Authority belonging to the good/outstanding category ‘decrease’ by 3 per cent for every deprivation score change of 1, in other words, it becomes 3 per cent less likely to be in the good/outstanding category for every increase in deprivation ranking, and 5.3 per cent less likely for every change in proportion of assessments overdue. Conversely, it becomes 7.9 per cent more likely to be in the good/outstanding category for every change in proportion of reporting children with missing episodes of care.

Multivariate regression models were created predicting the Ofsted judgement categories⁷. Ofsted categories were included as the dependent variable in logistic regression models through a stepwise process. Only the significant predictors from the univariate regression analysis were included in the final model⁸. The resulting model (χ^2 16.38, df 4, p 0.003) contained two predictors, Deprivation ranking (B -0.030, Exp b 0.971, p 0.043) and assessments overdue (B -0.059, Exp b 0.943, p 0.033), and explained approximately 20.7 per cent of the variance in the Ofsted judgement categories (Cox and Snell r^2 0.181, Nagelkerke r^2 0.207) (Table 10). Thus, it would be 2.9 per cent less likely for a Local Authority to belong to the good/outstanding group for every deprivation score change and similarly, 5.7 per cent less likely for every change in the proportion of assessments overdue.

	χ^2 (df)	B	SE (B)	Exp b	CI(95%)	P-value
Full model	16.38 (4)	-	-	-	-	.003**
Predictors						
Deprivation levels		-0.030	.015	.971	(.944, .999)	.043*
Assessments Overdue	-	-0.059	.028	.943	(.893,.995)	.033*

*. Correlation is significant at the 0.05 level (2

tail).
 **. Correlation is significant at the 0.01 level (2-tailed).

Table 10. Multivariate model predicting Ofsted category judgement

Strengths and limitations

The key limitations of our study are (i) the lack of reliable local authority spending data, (ii) the use of authority-level deprivation data and (iii) missing data from some local authorities.

⁸ The variable children in care with a missing episode from care was not included in the final model because it was a variable with data provided for by only half of the sample (i.e. forty-four Local Authorities) and therefore, it would limit the parameter estimates in the model.

⁷ Ordinal regression analyses were first performed on the data but because the assumptions were violated, multinomial logistic regressions were then performed.

⁸ The variable children in care with a missing episode from care was not included in the final model because it was a variable with data provided for by only half of the sample (i.e. forty-four Local Authorities) and therefore, it would limit the parameter estimates in the model.

First, the spending data we used is likely not entirely accurate. Different local authorities report their spending in different ways and while we have tried to take account of those differences when calculating 'spend per child' figures, we inevitably had to make some subjective decisions. It is possible that someone else looking at the same data might arrive at a different figure for at least some of the authorities. However, until there exists more accurate and consistent data, the approach we used is at least reasonably indicative. It is also a limitation that our 'spend per child' figures relate to the overall child population in each authority and not specifically to children referred to social services. Having said this, councils do spend money on the whole child population, for example, via community projects, and some of this investment is intended to prevent families from needing more specialist services. Secondly, our deprivation data relates to entire local authority areas. This approach masks within-authority variation (Bywaters et al., 2014a). As with the spending data, this means a degree of caution is warranted when interpreting our findings. Thirdly, not all local authorities provided a complete set of data to the DfE regarding the key variables we used. For example, only forty-four of the eighty-seven authorities provided data on the number of children in care with five or more 'good' GCSEs.

Finally, it is important to be clear about our use of the word 'predictor' in this context. We did not undertake a longitudinal study, so we are not making any claims about prediction in the sense of future performance. Rather, knowing the deprivation level of the local authorities included in this study allowed us to predict their membership of the different Ofsted categories we used but no causal relationship is implied, and our analyses were not intended to test or confirm causality. Strengths of the study include the relatively large number of local authority variables we considered (fifteen in total) and the combination of performance, practice or outcome, spending and deprivation data. Finally, the eighty-seven authorities included in the sample represent >50 per cent of the total number of authorities in England and thus, the patterns of association we report are likely to represent what happens in the remaining authorities as well.

Discussion

We examined the relationship between a range of key local authority variables, levels of spending, levels of deprivation and Ofsted judgements. It would be reasonable to hypothesise that authorities with better Ofsted judgements would consistently outperform authorities with worse Ofsted judgements, at least for some of these variables if not for all of them. For example, one might expect that 'good and outstanding' authorities have lower rates of re-referral and repeat CP plans than other authorities—but in the sample we considered, they did not. However, 'good and outstanding'

authorities did outperform other authorities in terms of procedural compliance, for example, having fewer overdue assessments, CP visits and CP conferences.

It is important to note that Ofsted do not claim to base their judgements on the variables we have considered in this article—and there is no suggestion by us (or by anyone else) that they should. A reliance on easily-measured performance indicators has been roundly criticised within the profession (Wastell et al., 2010) and our findings indicate that Ofsted judgements do not, in any case, correlate clearly with key performance data. However, this finding is itself confusing. If Ofsted were 'solely' focused on performance, authorities who held more meetings on time might consistently achieve better judgements—and we did not find this to be the case. But on the other hand, if Ofsted were 'solely' focused on outcomes, perhaps those authorities in which children in care achieved better school performance would consistently achieve better judgements (at least in relation to their looked after children services)— but we did not find this to be the case either. In relation to spending, one might hypothesise that 'good and outstanding' authorities simply outspend other authorities. In fact, from an initial consideration of 'face validity', our figures suggest that higher performing authorities ('good and outstanding') actually spend less money per child than underperforming authorities ('in need of improvement' or 'inadequate'). This finding echoes an earlier judgement by the National Audit Office (2016) that there is 'no correlation between local authorities' spending on children in need and the quantity and quality of their services' (p. 26). However, when we compared mean spend per child for levels of deprivation 'within' each Ofsted category, it became clear that 'good and outstanding' authorities do spend considerably more when they have higher levels of deprivation and much less when they have lower levels of deprivation than authorities in the 'inadequate' or the 'in need of improvement' categories.

Of all the variables we considered, we found only one that by itself predicted membership of the 'good and outstanding' category. The more deprived the authority, the more likely it was to be judged 'in need of improvement' or 'inadequate' and the less deprived the authority, the more likely it was to be judged 'good and outstanding'. Although some of the other variables we considered have some individual predictive power, the level of deprivation is the only variable to retain predictive power when combined with other variables. Recently, Ofsted themselves have become aware of this potential relationship, stating in their latest annual report, 'We have begun looking at these [data] and we found some correlation between the level of deprivation in the area and overall effectiveness' (2017b, p. 70). This represents an apparent shift from the previous year's report, in which it was stated that 'inadequacy is not a function of size, deprivation or funding (Ofsted, 2016,

p. 5, emphasis added). Ofsted say they will publish the results of their own investigations regarding the relationship between performance and deprivation in early 2018.

Another approach?

When evaluating the performance of secondary schools, the DfE take into account that different schools have different pupil intakes. They do so by calculating the 'value added' by each school, assessing the progress made by pupils from key stage two (Year 6) to key stage four (Year 11). This approach recognises that although all pupils can make progress, some pupils experience more disadvantage. Thus, achieving good outcomes is more difficult for some schools than for others. It also recognises that such difficulties are not directly related to the quality of teaching, important though this is. For example, pupils in school A achieve better GCSE results at key stage four than pupils in school B. But pupils in school B make greater progress between key stage two and key stage four than pupils in school A. As a result, school B will have a higher 'value added' score than school A. According to the DfE, this 'value added' measure is the best indication of a school's overall effectiveness.

However, measuring the performance of children's services is much more difficult than measuring the performance of schools (Malley and Fernandez, 2010; Forrester, 2017). Pupils can be asked to take standardised exams, not only at GCSE and A-level but theoretically at any point in their school career. The results of these tests can be compared at local, regional, national and even international level. Evaluating the performance of children's services is much more complicated. When a child comes into care, this might represent the best possible outcome for a child who would otherwise experience significant harm at home with their family. Or it might represent a failure on the part of children's services—and on the part of society as a whole—to adequately support the child's family. This complexity may explain why for so long the sector has relied on measuring things that are easy to count, such as the number of assessments completed on time, rather than the things that really matter but which are harder to quantify, such as the quality of those assessments.

Conclusion

Much of the publicly available data provided by local authorities to the DfE relates not to the quality of practice or outcomes for children but to the output. In the past, Ofsted inspections have been criticised for being overly concerned with compliance and insufficiently focused on the quality of practice and outcomes. In the time since we completed our analysis for this article, Ofsted (2017a) published a revised inspection framework and explicitly identified the need to strike a balance

between a rigorous and objective inspection and the need to help authorities improve rather than risk hindering their progress (Schooling, 2017). It would no doubt make for an interesting future analysis to consider a new set of inspection judgements and how they relate to an even broader set of performance indicators.

Understanding whether (and how) children's services are of sufficient quality is a key question for children, families, professionals and policy-makers. Yet it is important not to ignore the importance, significance and influence of wider social factors too. As Bywaters et al. (2015, 2017) have found, there is a strong relationship between deprivation and intervention rates and large inequities between ethnic categories. It is almost inconceivable that this relationship would not also be apparent within the pattern of Ofsted judgements across the country. No one is suggesting we should lower our aspirations for children's services, or that high deprivation should offer an easy excuse for poor practice. Yet, it is undoubtedly the case that the well-being and welfare of children and their families do not result simply from the quality of help provided by children's services. Put simply, providing effective help for children and families is harder when levels of deprivation are high—and a holistic inspection regime should take this into account.

References

- Bilson, A. and Martin, K. (2017) 'Referrals and child protection in England: One in five children referred to children's services and one in nineteen investigated before the age of five', *British Journal of Social Work*, 47(3), pp. 793–811.
- Bywaters, P., Brady, G., Sparks, T. and Bos, E. (2014a) 'Inequalities in child welfare intervention rates: the intersection of deprivation and identity', *Child and Family Social Work*, 21(4), pp. 452–63.
- Bywaters, P., Brady, G., Sparks, T. and Bos, E. (2014b) 'Child welfare inequalities: New evidence, further questions', *Child & Family Social Work*, 21(3), pp. 369–80.
- Bywaters, P., Brady, G., Sparks, T., Bos, E., Bunting, L., Daniel, B., Featherstone, B., Morris, K. and Scourfield, J. (2015) 'Exploring inequities in child welfare and child protection services: explaining the 'inverse intervention law'', *Children and Youth Services Review*, 57, pp. 98–105.
- Bywaters, P., Brady, G., Bunting, L., Daniel, B., Featherstone, B., Jones, C., Morris, K., Scourfield, J., Sparks, T. and Webb, C. (2018) 'Inequalities in English child protection practice under austerity: A universal challenge?', *Child and Family Social Work*, 23(1), pp. 53–61.

- Bywaters, P., Webb, C. and Sparks, T. (2017) Ofsted ratings do reflect local authority deprivation and spending, available online at: <http://www.communitycare.co.uk/2017/01/18/ofsted-ratings-reflect-local-authority-deprivation-spending/> (accessed 27 February 2019).
- Campbell, J. and Husbands, C. (2000) 'On the reliability of OFSTED inspection of initial teacher training: A case study', *British Educational Research Journal*, 26(1), pp. 39–48.
- Field, A. (2013) *Discovering Statistics Using SPSS*. 4th edn. Newbury Park, California, Sage Publications Ltd.
- Forrester, D. (2017) 'Outcomes in children's social care', *Journal of Children's Services*, <https://doi.org/10.1108/JCS-08-2017-0036> (accessed 27 February 2019).
- HM Government (2015) PM: We will not stand by – failing children's services will be taken over, available online at: <https://www.gov.uk/government/news/pm-we-will-not-stand-by-failing-childrens-services-will-be-taken-over> (accessed 27 February 2019).
- Hood, R., Grant, R., Jones, R. and Goldacre, A. (2016) 'A study of performance indicators and Ofsted ratings in English child protection services', *Children and Youth Services Review*, 67, pp. 50–6.
- Hood, R., Goldacre, A., Grant, R. and Jones, R. (2016) 'Exploring demand and provision in English child protection services', *British Journal of Social Work*, 46(4), pp. 923–41.
- Impower (2015) 'A brave new world: Is inspection improving Children's Services?', available online at: <http://www.impower.co.uk/insights/a-brave-new-world-is-inspection-improving-childrens-services> (accessed 27 February 2019).
- Jones, R. (2015) 'The end game: The marketisation and privatisation of children's social work and child protection', *Critical Social Policy*, 35(4), pp. 447–69.
- Jones, R. (2015b) 'Ofsted doesn't recognise the practice chaos and professional carnage it leaves in its wake', available online at: <http://www.communitycare.co.uk/2015/05/07/ofsted-doesnt-recognise-practice-chaos-professional-carnage-leaves-wake/> (accessed 27 February 2019).
- La Valle, I., Holmes, L., Gill, C., Brown, R., Hart, D. and Barnard, M. (2016) 'Improving Children's Social Care Services. Results of a feasibility study', available online at: http://www.corc.uk.net/media/1415/201611improving_childrens_social_care_services.pdf (accessed 27 February 2019).
- Malley, J. and Fernandez, J. (2010) 'Measuring quality in social care services: Theory and practice', *Annals of Public and Cooperative Economics*, 81(4), pp. 559–82.

- Munro, E. (2011) 'The Munro review of child protection: Final report', A Child-Centred System, available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175391/Munro-Review.pdf (accessed 27 February 2019).
- Munro, E. (2014) Review of first eleven Ofsted inspections of services for children in need of help and protection, children looked after and care leavers, and Local Safeguarding Children's Boards, available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/396378/Review_by_Professor_Eileen_Munro_of_the_first_eleven_Ofsted_inspections_of_children_s_services_and_reviews_of_LSCBs.pdf (accessed 27 February 2019).
- National Audit Office (2016) 'Children in need of help or protection', available online at: <https://www.nao.org.uk/wp-content/uploads/2016/10/Children-in-need-of-help-protection.pdf> (accessed 27 February 2019).
- Ofsted (2014) 'Raising standards, improving lives: The Office for Standards in Education, Children's Services and Skills (Ofsted) Strategic Plan 2014 to 2016', available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/379920/Ofsted_20Strategic_20Plan_202014-16.pdf (accessed 27 February 2019).
- Ofsted (2016) 'Ofsted social care annual report 2016', available online at: <https://www.gov.uk/government/publications/ofsted-social-care-annual-report-2016> (accessed 27 February 2019).
- Ofsted (2017a) 'Framework and evaluation schedule for the inspections of services for children in need of help and protection, children looked after and care leavers', available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/662790/Inspection_handbook_inspections_of_services_for_children_in_need_of_help_and_protection.pdf (accessed 27 February 2019).
- Ofsted (2017b) 'The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2016/17', available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/666871/Ofsted_Annual_Report_2016-17_Accessible.pdf (accessed 27 February 2019).
- Ouston, J., Fidler, B. and Earley, P. (1997) 'What do schools do after OFSTED school inspections – or before?', *School Leadership and Management*, 17(1), pp. 95–104.
- Rosenthal, L. (2004) 'Do school inspections improve school quality? Ofsted inspections and school examination results in the UK', *Economics of Education Review*, 23(2), pp. 143–51.

- Rowlands, J. (2010) 'Services are not enough: child well-being in a very unequal society', *Journal of Children's Services*, 5(3), pp. 80–8.
- Schooling, E. (2017) 'A new inspection system', available online at: <https://socialcareinspection.blog.gov.uk/2017/11/29/a-new-inspection-system/> (accessed 27 February 2019).
- Shaw, I., Newton, P., Aitkin, M. and Darnell, R. (2003) 'Do OSFTED inspections of Secondary Schools Make a Difference to GCSE Results? ', *British Educational Research Journal*, 29(1), pp. 63–75.
- Tilbury, C. (2004) 'The influence of performance measurement on child welfare policy and practice', *British Journal of Social Work*, 34(2), pp. 225–41.
- Wastell, D., White, S., Broadhurst, K., Peckover, S. and Pithouse, A. (2010) 'Children's services in the iron cage of performance management: street-level bureaucracy and the spectre of Svejkism', *International Journal of Social Welfare*, 19(3), pp. 1–20.