Evaluating the impact of the National Healthy School Standard: using national datasets

Sandie Schagen¹,³, Sarah Blenkinsop¹, Ian Schagen¹, Emma Scott¹, Michelle Eggers¹, Ian Warwick², Elaine Chase² and Peter Aggleton²

Abstract

An evaluation of the National Healthy School Standard (NHSS) was undertaken by the authors on behalf of the Department of Health and the Department for Education and Skills. One part of the evaluation involved gaining access to a number of datasets derived from previous research and analysing the health-related outcomes of schools which had attained Level 3 of the NHSS, compared with those of other schools. The sources which provided the most interesting findings were the Health-Related Behaviour Questionnaire (HRBQ) survey and the Ofsted database of school inspection ratings. This paper describes the statistical methods used, and the results of the HRBQ and Ofsted analyses. Using HRBQ data, many pupil-level outcomes were explored, but relatively few indicated significant differences and even those tended to be quite small. The Ofsted school-level data yielded stronger evidence of NHSS impact. The paper concludes by suggesting possible reasons for these findings.

Introduction

The National Healthy School Standard (NHSS) was established in England in 1999 as part of the Government’s drive to reduce health inequalities, promote social inclusion and raise educational standards. Its overall goal is to help schools become healthier environments for their pupils and staff, in the belief that a healthy school succeeds in helping pupils to do their best and build on their achievements. The NHSS aims to promote physical and emotional health by providing accessible and relevant information, and equipping pupils with the skills and attitudes to make informed decisions about their health (Health Development Agency, 2004).

Participation in the NHSS is open to maintained and independent nursery, primary, middle and secondary schools, as well as special schools and pupil referral units. A whole-school approach is taken to the identification of healthy schools activities, which are locally determined, but take national priorities into account. There are three levels at which a school can be involved in the NHSS:

- Level 1 indicates a general awareness of the NHSS and its goals.
- Level 2 requires schools to have accessed training and/or support through the scheme.
- Level 3 requires schools, in addition, to have begun the detailed process of auditing, target setting and action planning. (During the course of the evaluation, new guidance about Level 3 status was produced. For further information about these criteria, see: http://www.wiredforhealth.gov.uk/cat.php?catid=1080.)

In December 2002, the Department of Health and the Department for Education and Skills commissioned the Thomas Coram Research Unit at the Institute of Education, University of London and the National Foundation for Educational Research to jointly conduct an evaluation of the impact of the

¹National Foundation for Educational Research, Slough SL1 2DQ and ²Thomas Coram Research Unit, Institute of Education, University of London, London WC1H 0AA, UK
³Correspondence to: S. Schagen;
E-mail: s.schagen@nfer.ac.uk
NHSS. The evaluation had two components: the first examined processes of implementation at national, regional, local and school-based levels through in-depth interviews; the second aimed to identify pupil-level outcomes through an analysis of secondary datasets.

This paper reports on findings drawn from this second element of the evaluation. Secondary data sources were analysed in order to determine whether, and to what extent, the NHSS was achieving its three strategic aims, and to develop a set of quantitative national outcome indicators that could be used to analyse and monitor the future progress of the NHSS. In order to do this, Level 3 schools were compared with other schools in terms of pupil-level outcomes. A parallel paper (Warwick et al., 2005) reports in greater detail on school-level perceptions of programme and its effects.

Method

The first task was to identify possible outcome indicators to be ‘tested’ during the course of the evaluation, in order to determine whether, and to what extent, the NHSS was achieving its three strategic aims, and to develop a set of quantitative national outcome indicators that could be used to analyse and monitor the future progress of the NHSS. In order to do this, Level 3 schools were compared with other schools in terms of pupil-level outcomes. A parallel paper (Warwick et al., 2005) reports in greater detail on school-level perceptions of programme and its effects.

The following variables were defined to investigate the possible impact of Level 3 status:

- An indicator of whether or not the school has Level 3 status, regardless of when it achieved this status.
- An indicator to mark the length of time at Level 3.

These variables fulfilled different, but related, functions. If Level 3 schools perform significantly better (or worse) than other schools on a particular outcome, the association is not necessarily causal; it could mean that those schools have something in common (other than belonging to Level 3) which has influenced the outcome. In order to test whether the difference is due to Level 3 itself, it is important to take into account the length of time individual schools have had that status. Clearly, if a school became Level 3 only in 2002, for example, that could not have influenced the results of a survey undertaken in 2001. Conversely, if Level 3 status was having a positive impact, we would expect the variable denoting length of time to have positive statistical significance; if that is not the case, it is likely that the enhanced performance of Level 3 schools is due to other factors beyond the scope of the analysis.

The main analysis techniques used were linear (or logistic) regression and multilevel modelling. When dealing with pupil data, multilevel modelling is the preferred method of analysis, as it takes
account of the fact that pupils in the same cohort and/or school may have more in common than pupils in different schools [see (Goldstein, 2003)]. However, multilevel modelling is technically complex and time consuming, and as we were examining a wide range of outcomes on a variety of different datasets, it would not have been feasible to construct a model for every individual outcome.

A two-stage process was therefore adopted, whereby each outcome was first investigated by using linear or logistic regression. This is a simpler and quicker process, but (in comparison with multilevel modelling) may tend to exaggerate the significance of differences identified. Therefore, any factors identified as significant in a regression analysis need to be further explored by multilevel modelling, in order to confirm whether the difference is genuine; but if factors do not appear as significant in regression analysis, we can be confident that they are not significant, and no multilevel modelling is necessary.

Thus, outcomes where regression analysis suggested a possible effect of NHSS were subject to further analysis by multilevel modelling. In some cases, this confirmed that there was indeed an impact of NHSS (because Level 3 schools were significantly different from others, after controlling for other relevant variables); in other cases, it showed that the difference identified by regression was apparent rather than real. The results reported as significant below are those which have been confirmed by multilevel modelling. It was decided to identify findings significant at the 10% level, which increases the probability of detecting Level 3 effects, although it also increases the probability that effects detected may be due to chance. It was felt that borderline significance could indicate areas of potential, if not actual, NHSS impact, which would therefore be worth monitoring.

### Results

Below we describe the results from the analysis of the two sources which produced the most interesting findings.

#### Health-related behaviour questionnaire (HRBQ)

The HRBQ, developed by the Schools Health Education Unit at Exeter, provides an evaluation of current patterns in health-related behaviour of primary and secondary pupils (Years 5, 6, 8 and 10) based on self-reported data [see, e.g. (Balding, 2002)]. It should be noted that the survey does not cover a nationally representative sample of schools, as individual schools across the country opt to participate each year. (In some cases, organizations such as local education authorities, primary care trusts and/or drug action teams may fund participation.) However, for our purposes, the important point was that the sample should include a balanced mix of Level 3 and other schools, and that we could control statistically for other factors which might influence the outcomes.

Primary and secondary HRBQ data was available for 2000, 2001 and 2002, covering 361 schools (of which 227 were Level 3; 63%). For the purpose of the evaluation, 2000–2002 data was combined. The proportion of pupils in Level 3 schools is shown in Table I.

Primary and secondary school data was analysed separately. The following background factors were included at pupil level:

- Sex (boy/girl).
- Age.
- Ethnicity.

The following background factors were included at school level:

- Percentage eligible for FSMs (for analysis purposes, FSM eligibility is used as a proxy measure of socioeconomic status).

<table>
<thead>
<tr>
<th>Table I. The proportion of pupils in Level 3 schools responding to the HRBQ surveys (2000–2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of pupils who responded to HRBQ survey</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
</tbody>
</table>
The school’s overall KS2 (primary) or KS3 (secondary) results.
Whether or not the school was known to be Level 3.
Whether or not the school was known to be Level 3 prior to the survey.
Focus of NHSS activity.

The following analysis was undertaken:

- Comparison of Level 3 schools with other schools in relation to HRBQ individual outcomes, e.g. healthy eating, general sickness, participation in activities.
- Exploration of NHSS focus areas, to investigate whether the focus of NHSS activity has a significant impact on any relevant outcomes.
- Comparison of total scores. Given the broad scope of the NHSS and the fact that schools may choose to emphasize some areas rather than others, it is possible that the impact of the NHSS might not be measurable for any individual outcome, but that Level 3 schools might achieve a higher overall rating on a composite indicator. To test this hypothesis, a number of outcome measures included in the HRBQ survey were used to calculate a total score. The scores of pupils in Level 3 schools were then compared with the scores of pupils in other schools.
- Comparison of change in scores over time. For schools which had data available for 2000 and 2002, an average total score was calculated for each year, and the difference between the two scores was compared, in order to see whether Level 3 schools had improved more than others.
- Exploration of the differential impact of Level 3 status on schools with different proportions of pupils eligible for FSM. It was suggested that the NHSS might be having a greater impact on schools with a higher proportion of pupils eligible for FSM. An interaction term was included in the model, in order to test this hypothesis.

The summary of findings below is based on the multilevel modelling analysis of the combined 2000–2002 data. In order to investigate the possible impact of the NHSS, schools that were Level 3 prior to the survey have been compared with other schools. (‘Level 3 prior to the survey’ means that a school had Level 3 status prior to the survey being analysed, e.g. a school which achieved Level 3 status in 2001 would be counted as ‘Level 3 before the survey’ with reference to the 2002 data, but not with reference to the 2000 data.)

**Key findings for primary schools**
Pupils who were in Level 3 schools prior to the survey were significantly different from those in other schools in only two of the 12 outcomes analysed:

- They were less likely to be afraid of school due to bullying, but
- They were also less likely to eat fresh fruit.

There were no significant effects relating to the focus of NHSS activities in primary schools.

Twelve different outcome measures from the HRBQ primary school questionnaire (one numeric and 11 binary) were used to calculate a total score for each pupil (sufficient data was available for approximately 36,000 pupils). (The only non-binary outcome used in the primary score related to self-esteem, where a score was calculated based on pupil response to a series of statements. This was rescaled to a mean of zero and combined with the binary measures. They were included in the total score as +1 or −1, depending on whether they were felt to be positive or negative indicators of healthy behaviour.) The outcomes are listed in Table II. The total score was used as the outcome in regression analysis, controlling for pupil background factors, including being in a school which had become a Level 3 school prior to the survey.

There was no significant difference between the total scores obtained by primary pupils in Level 3 schools and those obtained by pupils elsewhere. This indicates that the NHSS was not having an impact on primary pupils in terms of health-related behaviour overall.

Sixty-eight primary schools participated in both the 2000 and 2002 surveys. For each, an average
total score was calculated, and change over time was investigated. Figure 1 plots the average scores for 2000 against the average scores for 2002. As shown, some schools (both Level 3 and others) had improved scores and others did not. Level 3 schools were no more likely than other schools to improve their scores between 2000 and 2002.

One aim of the NHSS is for all schools with more than 20% of pupils eligible for FSM to reach NHSS Level 3 status by 2006. The Level 3 primary schools included in the above analysis had a higher average percentage of pupils eligible for FSM than the other primary schools (22% compared to 15% eligibility). This suggests that the NHSS is succeeding in its target of reaching schools with high FSM eligibility.

It has been suggested that the NHSS might be having a stronger impact on schools with high proportions of pupils eligible for FSM. To explore this hypothesis, an interaction term was included in the statistical model which would indicate if the impact of the NHSS varied according to level of FSM eligibility. This proved to be the case in relation to some outcomes, but the findings were not always as might be expected.

The analysis revealed that pupils in Level 3 primary schools with higher FSM eligibility were (slightly) more likely to drink water than would otherwise be predicted, but also more likely to smoke cigarettes, drink alcohol and have nothing for breakfast, and less likely to have regular dental checkups and eat vegetables. These findings suggest that, overall, the NHSS has had a greater impact in primary schools with lower FSM eligibility, although the differences were very small.

### Key findings for secondary schools

Secondary school pupils in schools which were Level 3 prior to the survey:

- Were less likely to have used drugs.
- Were more likely to feel at ease when visiting a doctor.
- Had higher self-esteem scores.
- Were more likely to know where to get free condoms.

In addition, students in Level 3 schools with a focus on physical activity were less likely to watch TV for more than 1 hour after school.

Nineteen different outcomes (four numeric outcomes and 15 binary) from the secondary questionnaire were used to create a total score for pupils (sufficient data was available for approximately 52,000 pupils). The outcomes included are listed in Table III.

There was a significant positive relationship between the total score and being Level 3 prior to the survey. It appears that for secondary pupils (unlike primary pupils) attendance at a Level 3 school did have a positive impact on knowledge and behaviour related to health and social inclusion.

The average total score for secondary schools participating in 2000 and 2002 (37 schools) was calculated for each year. Figure 2 plots the average scores for 2000 against the average scores for 2002.
It shows that Level 3 secondary schools were more likely than other schools to have improved scores over time. Again, it suggests that Level 3 secondary schools were having a positive influence on pupils’ health-related behaviour.

Level 3 secondary schools included in the above analysis had a higher average percentage of FSM eligibility (20%) than other secondary schools (17%), although the difference was not as striking as for primary schools.

Students in Level 3 secondary schools with higher FSM eligibility were more likely to eat fruit than would otherwise be predicted. However, they were more likely to have been offered cannabis and to play computer games, and less likely to eat vegetables; they also had lower self-esteem. Again, these findings suggest that the NHSS has had a greater overall impact in schools with lower FSM eligibility, although it should be noted that (as with primary schools) the differences were very small.

**Summary of HRBQ findings**

Findings from the analysis of the HRBQ datasets were, overall, very positive for secondary schools and less so for primary schools. There was a significant difference between Level 3 and other schools for only two of the primary outcomes, and one of these was not in a positive direction. There were no significant differences relating to focus areas, total score or change over time.

With secondary schools, however, there was a significant difference on four individual outcomes, all in a positive direction. Further, students in Level 3 schools with a focus on physical activity were less likely to spend a lot of time watching television. Students in Level 3 schools scored higher on a composite measure of health-related behaviour and Level 3 schools were more likely to improve their average scores over time.

**Ofsted inspection data**

Data was received from Ofsted based on school inspections carried out during the academic years 1999/2000 to 2002/2003 [see (Ofsted, 2002)]. (Schools are normally inspected on a 6-year cycle. A team of inspectors spends a few days in a school observing lessons and speaking to teachers and pupils, to gather evidence of school performance. Inspectors are given frameworks which spell out principles of inspection and handbooks which give detailed guidance on inspection criteria. All aspects of school life are rated on seven-point scales.) Usable data was received for 7666 primary schools (2518 of which were Level 3; 33%) and 1402
secondary schools (545 of which were Level 3; 39%).

The Ofsted scales received related to a number of social inclusion indicators relevant to the evaluation of the NHSS, in particular emotional well-being, behaviour, participation and attitudes to school. Specifically, the 11 scales used were:

1. Attitudes to school.
2. Behaviour including exclusions.
3. Personal development and relationships.
4. Enthusiasm for school.
5. Interest and involvement in activities.
7. Absence of oppressive behaviour (e.g. bullying).
8. Provision for Personal, Social and Health Education (PSHE).
9. Monitoring and promoting good behaviour.
10. Monitoring and eliminating oppressive behaviour.
11. Monitoring and supporting personal development.

It should be noted that these are school-level rather than pupil-level scales, i.e. Ofsted inspectors rate schools rather than individual pupils. It is also important to note that the Ofsted scales are based on the judgement and discretion of teams of school inspectors, which may vary.

It should be noted that Ofsted scales are ordered with 1 being the best rating and 7 being the worst. If the analysis were carried out using these scales as they stand, then a positive relationship would be indicated by a negative coefficient and vice versa. To avoid confusion, each scale was reversed so that the best rating was 6 and the worst 0.

Analysis was carried out separately for primary and secondary schools, and in each case the following school-level background factors were included in the model:

- An indicator of the year in which the inspection was conducted.
- Whether or not the school is currently Level 3.
- Whether the school was Level 3 before the inspection was carried out.
- The focus of NHSS school activity.

For the reasons explained above, the key variable to consider was not Level 3 status per se, but whether the schools concerned had achieved Level 3 status prior to the Ofsted inspection which had generated the scales.

**Key findings for primary schools**

Table IV summarizes the results of the analysis, carried out using multiple regression, for the primary schools. The values tabulated are β coefficients (represented as percentages) which indicate the ‘strength’ of the relationship between the given Ofsted scale and each background factor, when all other factors are taken into account (i.e. a positive

<table>
<thead>
<tr>
<th>Table IV. Analysis of Ofsted scales (primary schools)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% FSM (%)</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Attitudes to school</td>
</tr>
<tr>
<td>Behaviour including exclusions</td>
</tr>
<tr>
<td>Personal development and relationships</td>
</tr>
<tr>
<td>Enthusiasm for school</td>
</tr>
<tr>
<td>Interest and involvement in activities</td>
</tr>
<tr>
<td>Behaviour</td>
</tr>
<tr>
<td>Absence of oppressive behaviour</td>
</tr>
<tr>
<td>Provision for PHSE</td>
</tr>
<tr>
<td>Monitoring and promoting good behaviour</td>
</tr>
<tr>
<td>Monitoring and eliminating oppressive behaviour</td>
</tr>
<tr>
<td>Monitoring and supporting personal development</td>
</tr>
</tbody>
</table>

aPerformance is based on an indicator (on a scale of 1–5) of the schools’ overall performance as judged by the latest available national curriculum assessment results.
figure illustrates a positive relationship between the scale and the background factor). Only those coefficients which are significant at the 10% level are shown.

The relationship with FSM eligibility was mostly negative for primary schools overall; as might be expected, schools with higher proportions of FSM pupils had lower ratings on most criteria. However, those with a high percentage of pupils eligible had positive relationships with monitoring and eliminating oppressive behaviour and monitoring and supporting personal development. There was a positive relationship between all 11 scales and performance. Larger schools had more positive relationships with five of the 11 scales (illustrated in the ‘size’ column).

Allowing for these background factors, there was a significant positive relationship between being a Level 3 school prior to inspection and 10 of the 11 scales. The relationship with provision for PSHE was particularly positive. The only scale where no significant relationship was evident was ‘Monitoring and supporting personal development’.

**Key findings for secondary schools**

Table V summarizes the results of the Ofsted analysis for secondary schools. In contrast to primary schools, in schools with higher FSM eligibility there were positive relationships with all 11 scales. However, the relationship with size was largely negative (i.e. larger schools had lower ratings). The relationship with school examination performance was much stronger (more positive) than was the case for primary schools.

Allowing for these background factors, the findings regarding Level 3 status were not as impressive as those for primary schools, although there was a positive relationship between being a Level 3 school prior to inspection and six of the 11 scales. Similar to the primary findings, the relationship with provision for PSHE was particularly positive. The relationship between being a Level 3 secondary school prior to inspection and monitoring and supporting personal development was equally positive (this was the only scale where no positive relationship existed for primary schools which were Level 3 prior to inspection).

**Discussion**

With reference to the HRBQ dataset, many possible outcomes were investigated, but relatively few indicated significant differences and even these tended to be quite small (the same was also true for other data sources analysed). It should be noted, however, that the analysis undertaken was effectively a baseline measure, as most schools had been Level 3 for a relatively short period of time when the relevant survey was conducted. Further, the broad nature of the NHSS (with the specific focus determined at local level) reduces the expectation of detecting an overall impact on any single national indicator.

**Table V. Analysis of Ofsted scales (secondary schools)**

<table>
<thead>
<tr>
<th>Attitudes to school</th>
<th>% FSM</th>
<th>Size</th>
<th>Performancea</th>
<th>Level 3 before inspection</th>
</tr>
</thead>
<tbody>
<tr>
<td>+12</td>
<td>8</td>
<td>+66</td>
<td>+5</td>
<td></td>
</tr>
<tr>
<td>Behaviour including exclusions</td>
<td>+11</td>
<td>-9</td>
<td>+60</td>
<td></td>
</tr>
<tr>
<td>Personal development and relationships</td>
<td>+18</td>
<td>-6</td>
<td>+62</td>
<td></td>
</tr>
<tr>
<td>Enthusiasm for school</td>
<td>+12</td>
<td>-10</td>
<td>+67</td>
<td></td>
</tr>
<tr>
<td>Interest and involvement in activities</td>
<td>+9</td>
<td>+60</td>
<td>+5</td>
<td></td>
</tr>
<tr>
<td>Behaviour</td>
<td>+10</td>
<td>-9</td>
<td>+59</td>
<td></td>
</tr>
<tr>
<td>Absence of oppressive behaviour</td>
<td>+20</td>
<td>-13</td>
<td>+57</td>
<td></td>
</tr>
<tr>
<td>Provision for PHSE</td>
<td>+12</td>
<td>+16</td>
<td>+10</td>
<td></td>
</tr>
<tr>
<td>Monitoring and promoting good behaviour</td>
<td>+23</td>
<td>-7</td>
<td>+42</td>
<td></td>
</tr>
<tr>
<td>Monitoring and eliminating oppressive behaviour</td>
<td>+26</td>
<td>-8</td>
<td>+46</td>
<td></td>
</tr>
<tr>
<td>Monitoring and supporting personal development</td>
<td>+19</td>
<td>+37</td>
<td>+10</td>
<td></td>
</tr>
</tbody>
</table>

*Performance is based on an indicator (on a scale of 1–5) of the schools’ overall performance as judged by the latest available national curriculum assessment results.
In an attempt to address the latter issue, the focus areas of individual schools were (where possible) included in the analysis, but only in two cases was there evidence of a significant association between focus area and a related outcome. However, it should be borne in mind that schools’ choice of focus area may in some cases at least reflect a known area of weakness. For example, a school which recognized that its pupils were eating unhealthy diets might focus on healthy eating. If they started from a low base, the fact that they were still below average in terms of fruit consumption would not necessarily imply that their intervention had failed.

When the HRBQ findings are compared with those from the analysis of the Ofsted database, some interesting differences emerge. The HRBQ findings were more positive for secondary schools than for primary schools, while the opposite was true of the analysis of Ofsted ratings. In general, however, the analysis of the Ofsted database yielded stronger evidence of NHSS impact than the HRBQ or any of the other sources based on pupil surveys. Why should this be? We should first acknowledge the possibility that awareness of the school’s involvement in the NHSS could have influenced Ofsted assessment on related criteria. For example, knowing that a school had achieved Level 3 status might reasonably lead an inspector to infer that they must be doing good work in PSHE and this knowledge might contribute to a positive assessment.

A more important point to note is that some projects undertaken under the auspices of the NHSS may directly affect only a small proportion of the total student body. The impact of such projects would be ‘diluted’ when outcomes for a whole cohort or a whole student population were being explored. Participation may have a dramatic impact on the attitudes and behaviour of the pupils concerned, but that change may not be enough to significantly increase the school’s average rating in terms of student attitudes.

Considerations such as these highlight the difficulty of evaluating initiatives such as the NHSS and the need for sophisticated analysis to detect their impact.

Acknowledgements

The work described here was funded by the Department of Health and the Department for Education and Skills. The views expressed are those of the authors. We would like to thank Ofsted and the Schools Health Education Unit for allowing us to use their databases for our analysis.

References


Received on October 28, 2004; accepted on February 7, 2005