

## **Exploring the validity of teachers' judgements of pupils' attainments**

Lillian Munro, Scottish Qualifications Authority  
Sandra Johnson, Assessment Europe, Scotland

*Scottish education for pupils aged 5 to 14 is currently based on a set of National Curriculum guidelines, published in the early 1990s. These attempt to describe systematic subject progression in knowledge, understanding and skills acquisition as pupils move through their schooling, against a common level-based attainment framework. Levels run from A to F, and there are explicit expectations about the levels that pupils in the different stages of schooling can be expected to have attained. Most teachers use this level framework when evaluating their pupils' progress, and many use level-based 'national tests' in reading, writing and mathematics to confirm and support their judgements. There is no statutory testing of pupils, population attainment being monitored through the sample-based Scottish Survey of Achievement (SSA). Pupils are tested in the SSA, but their teachers are also invited to submit level judgements for them. As a result, teachers' level judgements and test-based level classifications are now available for large national pupil samples in key curricular areas. The paper explores relationships between these two sets of assessment data for mathematics and for science, highlighting similarities and differences between the two areas and considering implications for teachers and policy makers in Scotland and further afield.*

### **The Scottish Survey of Achievement**

For over a decade teachers in Scottish schools have been delivering the 5-14 National Curriculum. The National Curriculum Guidelines, published in the early 1990s with some later updates, describe pupils' expected learning experience in each subject in terms of a framework of outcomes, strands, targets and levels (e.g. SOED 1991, SOEID 1999 for mathematics; Scottish Executive 2000 for science). The framework is essentially a 2-dimensional grid, comprising 'strands' (and 'targets') within 'outcomes' as one dimension – the abilities, skills and, in some cases, attitudes to be developed – and attainment 'levels' as the other. While the outcomes and their subdivisions naturally differ from one subject to another, the 6-level progression framework is common to all. Levels run from A to F, and there are explicit expectations about the levels that pupils in the different stages of schooling can be expected to have attained, whatever the subject. Most teachers use this level framework when evaluating their pupils' progress, and many use level-based 'national tests' (known as National Assessments) in reading, writing and mathematics to confirm and support their level judgements (Munro 2003). There is no statutory testing of pupils in Scotland, population attainment being monitored through a sample-based survey programme – the Scottish Survey of Achievement.

The Scottish Survey of Achievement (SSA) is an annual survey of pupil attainment at four stages of schooling: Primary 3 (P3: 7-8 year olds), Primary 5 (P5: 9-10 year olds), Primary 7 (P7: 11-12 year olds) and Secondary 2 (S2: 13-14 year olds). Surveys take place in late April/May, when pupils are close to completing their programmes of work for the school year. In the period 2005-2008 the subject focus of the survey has been, respectively, English language, social subjects (enquiry skills), science and mathematics (see Scottish Government 2006, 2007, 2008 for the reports on the surveys of 2005, 2006 and 2007). The core skills of ICT, problem solving and working with others have also been assessed each year, in the context of the curricular focus of the survey in that year.

The SSA is an integral part of a coherent framework of assessment in Scotland (see Hayward 2007 for details). Its launch in 2005 was set in the context of the much wider Assessment is for Learning (AifL) programme, a programme that explicitly separates the external/summative aspects of assessment from the internal/formative purposes (Hutchinson & Hayward 2005). With the introduction of the SSA, the then Scottish Executive<sup>1</sup> not only extended the previous monitoring role of the Assessment of Achievement Programme (AAP), which had been running since the mid-1980s (Condie et al. 2003), but simultaneously discontinued the annual National 5-14 Survey (also

---

<sup>1</sup> In 2007 the Scottish Parliament changed the name from Scottish Executive to Scottish Government

known as the 'National Audit'), which had run over the previous decade (see, for example, Scottish Government 2004a). The National Audit required schools to provide their local authorities with attainment data for every pupil in the country, at each stage of primary and the first two years of secondary schooling, in the form of teachers' level judgments for reading, writing and mathematics. The information was used by authorities to set performance targets for individual schools, and in turn forwarded to the Scottish Executive, where it was used for national monitoring purposes.

The AAP had already been reporting pupil attainment in the main subjects with reference to the levels of the National Curriculum Guidelines, the same level framework that teachers' judgements relate to. Specifically, the surveys reported the estimated percentages of pupils in the stage population who were deemed on the evidence to be 'working at' one or other of a range of levels assessed at that stage, for example Levels A to C at P3. While there are measurement challenges associated with this form of reporting, the benefits are that all stakeholders – policy makers, teachers, pupils and parents – are able fully to understand the reported findings, given their shared understanding of the 5-14 level framework. The SSA continued this reporting strategy. This meant that it could in principle, through testing *samples* of pupils, adequately provide the now Scottish Government with the same kind of attainment information previously provided by teachers for *every* pupil in the country. The fact that the pupils in the survey samples, and indeed their schools, would remain anonymous would serve to neutralise any fear of the survey results being used for inter-school comparisons, as had been the case with the National Audit.

The Scottish Government introduced the SSA in order to create a 'more robust and reliable system for assessing levels of attainment', more robust, that is, than the collection of teachers' level judgements (Scottish Government, 2004b). Yet there had never been any empirical evidence to allow comparison of teachers' judgments with other 'independent' measures, to inform the decision to abandon the National Audit. By asking teachers to provide their own estimate of pupils' attainment levels for pupils taking part in the SSA, we opened up the possibility of making these comparisons and exploring the reasons for any differences, should differences emerge. Schools that participate in the surveys are routinely invited to submit teachers' level judgements for reading, writing and mathematics for the pupils selected for assessment, whatever the subject focus of the survey. In 2007, where the subject focus was science, teachers were invited to submit an additional judgement of level for science. In this paper we present and discuss the survey findings for mathematics and science.

## **Teachers' level judgements for mathematics and science**

The six levels of attainment in the national progression framework for 5-14 are described verbally in each set of subject curriculum guidelines, in terms of profiles of knowledge, skills, abilities and behaviours. Whatever the subject there are a set of expectations about the levels that pupils at certain stages in schooling should have attained:

Level A: should be attainable in the course of P1-P3 by almost all pupils.

Level B: should be attainable by some pupils in P3 or earlier, and by most in P4.

Level C: should be attainable in the course of P4-P6 by most pupils.

Level D: should be attainable by some pupils in P5-P6 or earlier, and by most in P7.

Level E: should be attainable by some pupils in P7-S1, but certainly by most in S2.

Level F: should be attainable in part by some pupils, and completed by a few pupils in the course of P7-S2.

One could argue about the realism and validity of these official expectations, perhaps more so in some subject areas than others. What is important here is that these are the level expectations that teachers have been offered, to guide their curriculum delivery. Is it surprising, then, that in the majority of cases teachers' level judgements correspond closely with this picture? Tables 1 and 2 give the evidence for mathematics and science based on teachers' judgements submitted in the surveys of 2006 and 2007, respectively (the data are population estimates – see the survey report, Scottish Government 2007, 2008, for details of the estimation procedure).

Table 1  
 Teachers' level judgements for mathematics in 2006  
 (% pupils categorised into each level – population estimates)

Stage	No. of pupils	<A	A	B	C	D	E	F
P3	5,118	2	74	23	0	0	0	0
P5	5,294	0	4	41	52	2	0	0
P7	5,378	0	0	4	21	55	19	1
S2	5,202	0	0	2	10	23	44	20

Source: Extract from Table D4, Scottish Government 2007a, Chapter D

As Table 1 shows, the majority of P3 pupils, around three-quarters, were judged by their teachers to be working at Level A in mathematics, with most of the rest judged to be at Level B. At P5 the majority of pupils were judged at Level B or C. For P7 and S2 the expected levels are Level D and Level E, respectively. We see that these were indeed the modal levels in each case.

Teachers in Scotland have never been required officially to submit level judgements for their pupils in science, either to their authorities or to central government. And neither have there been National Assessments available to them with which they could confirm any such judgements that they made. Asking teachers in the SSA to provide judgements for their survey pupils in this subject area was therefore to be of special interest. How would teachers react? How many would indeed submit level judgements? These were unknowns. In the event, level judgements were submitted for around 90% of the survey pupils. Table 2 provides the resulting level profiles.

Table 2  
 Teachers' level judgements for science in 2007  
 (% pupils categorised into each level – population estimates)

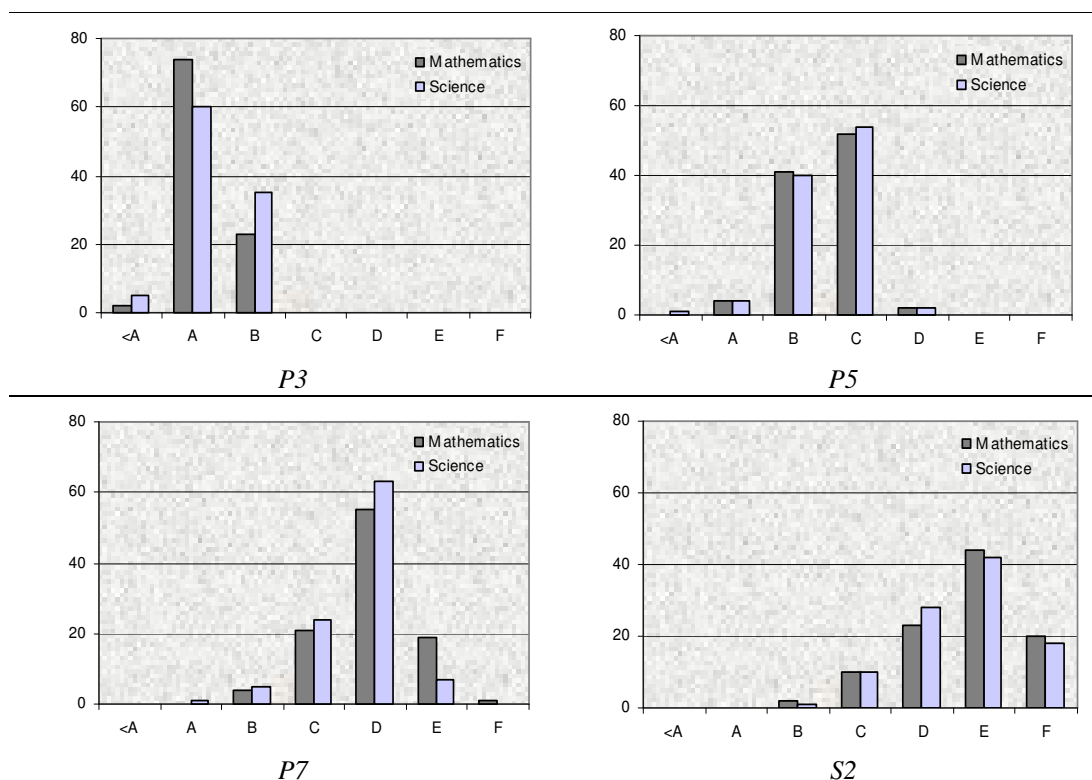
Stage	No. of pupils	<A	A	B	C	D	E	F
P3	9,656	5	60	35	0	0	0	0
P5	9,826	1	4	40	54	2	0	0
P7	10,154	0	1	5	24	63	7	0
S2	10,496	0	0	1	10	28	42	18

Source: Extract from Table B4, Scottish Government 2008, Chapter B

From Table 2 we see that, although the proportion, at 60%, is lower than that for mathematics, at 74%, the majority of P3 pupils were judged at Level A in science by teachers, with over a third more put at Level B. At P5, P7 and S2 the modal levels are Levels C, D and E, respectively, as for mathematics. In general, teachers' judgements were more positive for science than for mathematics at P3 and slightly less positive at P7, with virtually no difference between the two subjects at the other stages (see Figure 1).

The question is: Do these results validate the expected levels as given in the curriculum guidelines, or do teachers simply interpret the level descriptions to fit with the official expectations for the pupils they teach? We cannot know this. But we can at least now compare the teachers' judgements with test-based evidence, something that has not previously been possible, even if the test-based evidence available for mathematics actually relates to 'numeracy' (assessed in 2005 and 2006) and not to mathematics more generally.

Figure 1  
A comparison of level distributions for mathematics and science based on teachers' judgements  
(% pupils classified at each level)



### Teachers' judgements compared with test results

The decision to assess numeracy in the first SSA surveys rather than the wider mathematics, which had been the subject of previous AAP surveys in 2004 and earlier, was related to a policy requirement to assess core skills in the new programme. However, the term 'numeracy' can be used in different ways and there was no accepted definition in relation to the 5-14 curriculum guidelines for mathematics followed within Scottish schools<sup>2</sup>. The definition which was adopted for the SSA was based on the functional understanding of numeracy used in the context of core skills within Scottish National Qualifications for post-14 year old pupils, which emphasises the need for individuals to be 'comfortable' with numbers, graphs, symbols and diagrams in order to meet the demands of day to day living, including both work and study (Scottish Qualifications Authority, 2003). For SSA purposes numeracy was taken to comprise two distinct aspects: 'using number', defined in terms of the four number processes, and 'using graphical information', defined in terms of information handling. Individual assessment items were therefore drawn from two of the four outcomes defined in the 5-14 Mathematics Guidelines, namely *Number, money and measurement* and *Information handling*. The other two outcomes of *Shape, position and movement* and *Problem solving* were not included.

Domain sampling was employed to select numeracy items from the 5-14 National Assessment Bank<sup>3</sup> for survey use, supplemented through new item development when necessary. Assessment items had been developed and judgementally validated (by practising teachers) to tap the knowledge, abilities and skills at a single level. In total, almost 600 numeracy items were administered in each of the surveys of 2005 and 2006, over half being incidentally common to both. In 2005 the items were distributed among 40 different test booklets, 10 per stage, while in 2006

<sup>2</sup> Since then, the development of the new 'Curriculum for Excellence' has defined numeracy as a sub-set of mathematics. These new guidelines are currently out for consultation and have not yet been implemented within Scottish schools.

<sup>3</sup> <http://www.aifl-na.net/>

there were 48 different booklets, 12 per stage. Test booklets all had the same general make-up in terms of content coverage appropriate to the stage concerned (no empirical constraints were imposed during booklet creation), and each came in two versions, the second simply presenting the items in reverse order. Multiple matrix sampling was used to allocate items to test booklets, and test booklets to pupils. Over 350 assessment tasks (comprising one or more items) featured in the 2007 science survey. Again, every task was judgmentally validated to be at a particular level in the curriculum framework for science (environmental studies). The tasks were distributed among a total of 48 different test booklets, 12 at each stage. Multiple matrix sampling was used to allocate test booklets to pupils.

In neither subject was a test booklet a test. Every booklet contained items at three consecutive levels, with the exception of P3 in science, for which booklets contained tasks at two levels only, *viz.* A and B. This was to ensure that no pupil was faced with a test booklet entirely composed of items/tasks from one level only, since the level concerned might be far below their capabilities or dauntingly above them. In numeracy, items were presented in a randomized order throughout each booklet (there was no ordering by level), and every booklet came in two versions, one presenting items in the reverse order of the other. In science, tasks were grouped by theme (e.g. 'photosynthesis') and there were three versions differing by order of task presentation. Individual pupils took two mixed-level test booklets, randomly assigned. Single-level tests, or 'testlets', comprised those items or tasks at the relevant level across the booklet pair. Each booklet pair therefore furnished data relating to three different single-level testlets, one at each of three consecutive levels (two levels at P3 for science).

In both subjects, pupils were classified into attainment bands at each level on the basis of their testlet performances, specifically in terms of the proportions of items they successfully answered at that level in their two booklets<sup>4</sup>. Pupils achieving 80% or more of the available marks at a level were deemed to have shown 'very good' attainment at that level. Pupils achieving 65% or more of the marks at a level correctly but fewer than 80% were classified as having 'well-established' skills at the level (i.e. 'mastery'). Pupils achieving 50% or more of the marks at a level correctly, but not as many as 65% were deemed to have made a 'good start' at the level. Weighted proportions of pupils in each band at each level in each authority and nationally were computed, and margins of error estimated using the jackknife technique. The results are presented in some detail in the survey reports (Scottish Government 2007, 2008).

For present purposes the published attainment data have been re-organised, to produce level profiles at each stage that can be readily compared with those given in Tables 1 and 2 for teachers' judgements. For example, a P3 pupil demonstrating 'mastery' (testlet score of 65% or higher) at, say, Level A, but not doing so at Levels B or C would be classified as 'Level A'. A P3 pupil reaching the 65% cut-off score at both Levels A and B but not C would be deemed to be at Level B (to give an indication of the reliability of the testlets for classifying pupils by level, the overall percentages of pupils who reached the cut-off score at one level having failed to reach it at a lower level were 1.5% for numeracy and 4% for science). The resulting level profiles are given in Tables 3 and 4 for numeracy and science, respectively.

Note that the administration of just three single-level tests to individual pupils (two at P3 in science) means that a pupil could only be classified at one of four levels (three at P3 in science). Thus, in mathematics a P3 pupil could be classified as 'below Level A' or at one or other of Levels A to C. At P3 and P5 this presents no difficulty for comparison with teachers' judgements, since these anyway span a similarly narrow range. At P7 and S2, where the levels spanned by the teachers' judgements fall beyond the range available from the testing, the judgements for the small percentages of pupils affected were adjusted: thus, for example, the 2% of S2 pupils judged to be at Level B by their teachers were put at Level C for the sake of the comparison.

---

<sup>4</sup> The banding criteria, while in principle arbitrary, are based on the professional judgment of subject specialists. They were first used for reporting reading attainment in the 2001 AAP English Language survey, and have continued into the SSA.

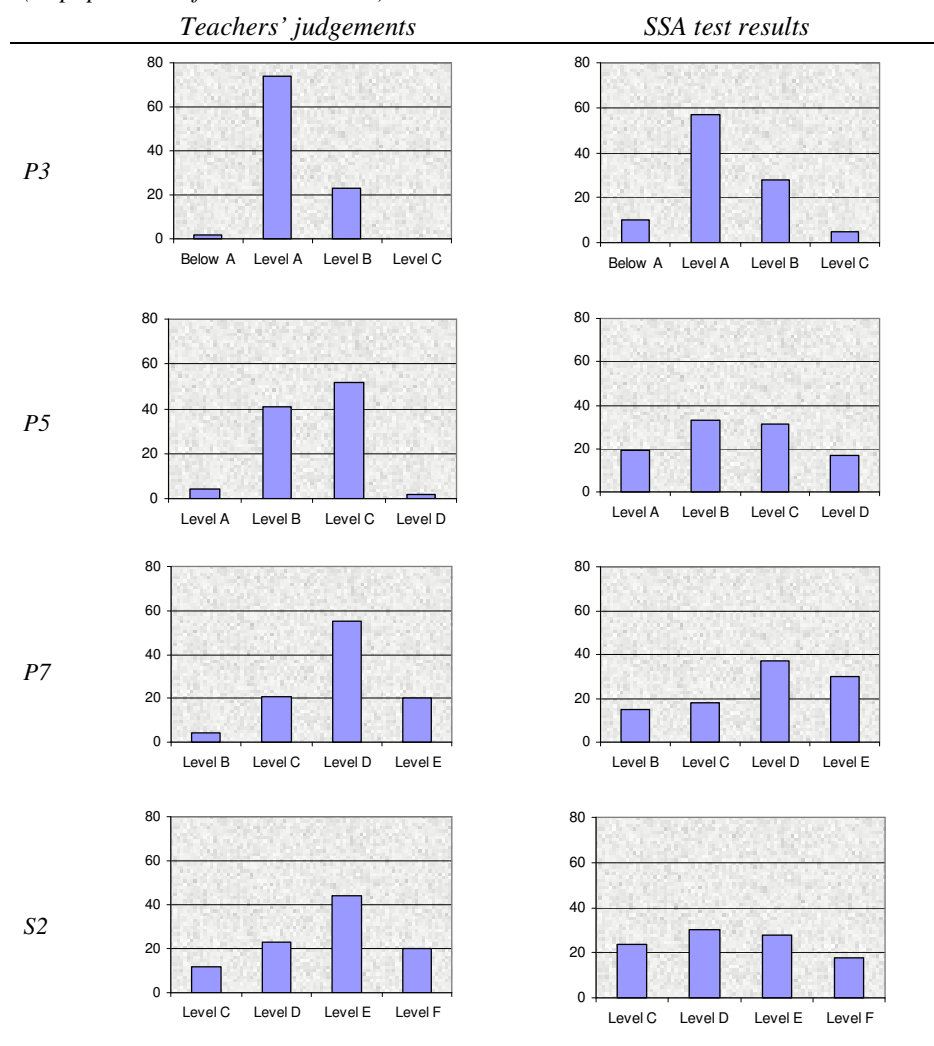
Table 3  
 SSA test-based level profiles for numeracy in 2006  
 (% pupils categorised into each level – population estimates)

Stage	No. of pupils	<A	A	B	C	D	E	F
P3	3,316	10	57	28	5	0	0	0
P5	3,221	0	19	33	31	17	0	0
P7	3,500	0	0	15	18	37	30	0
S2	3,011	0	0	0	24	30	28	18

Source: These data have been produced by sequential subtraction of the ‘well-established’ proportions shown in Table D1, Scottish Government 2007, Chapter C.

If we compare the test-based results for numeracy with the teachers’ judgements for mathematics (Figure 2) we see that the two sets of information are discrepant to some extent. At all stages, the SSA test results tended to produce ‘flatter’ level distributions for numeracy than the teachers’ judgements for mathematics.

Figure 2  
 A comparison of level distributions for SSA numeracy testing and teachers’ judgements for mathematics  
 (% pupils classified at each level)



For example, as we see from Table 3, the SSA test results put fully 10% of P3 pupils below Level A for numeracy skills, almost 20% of P5 pupils below Level B, and 15% of P7 pupils below Level C, when teachers at the different stages put fewer than 5% at these lowest levels (Table 1). The SSA tests equally put 5% of P3 pupils at Level C, over 15% of P5 pupils at Level D, and 30% of P7 pupils at Level E, when again the teachers had rated relatively few pupils so highly: virtually none at P3, 2% at P5 and under 20% at P7. At S2, teachers and tests agreed that around 20% of pupils were at Level F. But the SSA tests tended to pull the rest of the level distribution downwards, i.e. the tests tended to credit very many S2 pupils with lower levels of attainment in numeracy than the teachers had given them for mathematics as a whole. Figure 2 illustrates the picture across the stages. Where both a test result for numeracy and a teacher's judgement for mathematics were available for individual pupils, the level classifications agreed in 50% of cases, the highest agreement rate emerging for P3, at 58%, and the lowest for P5 and S2, at 46% and 47% respectively.

Table 4 reveals a picture of test-based attainment in science that is very much less positive than that for mathematics/numeracy. We no longer see modal levels coinciding with expected levels for the various stages. Indeed, at P7 and S2 the majority of pupils were classified on the basis of their test results two levels below where they would be expected to be on the basis of the Curriculum Guidelines, and where their teachers actually put them. Among those individual pupils for whom a teacher's judgement and a test result were available just over one in five (22%) received the same level classification from both sources: the highest rate of individual agreement was again for P3, at 34%, with the lowest for P7, at 11%.

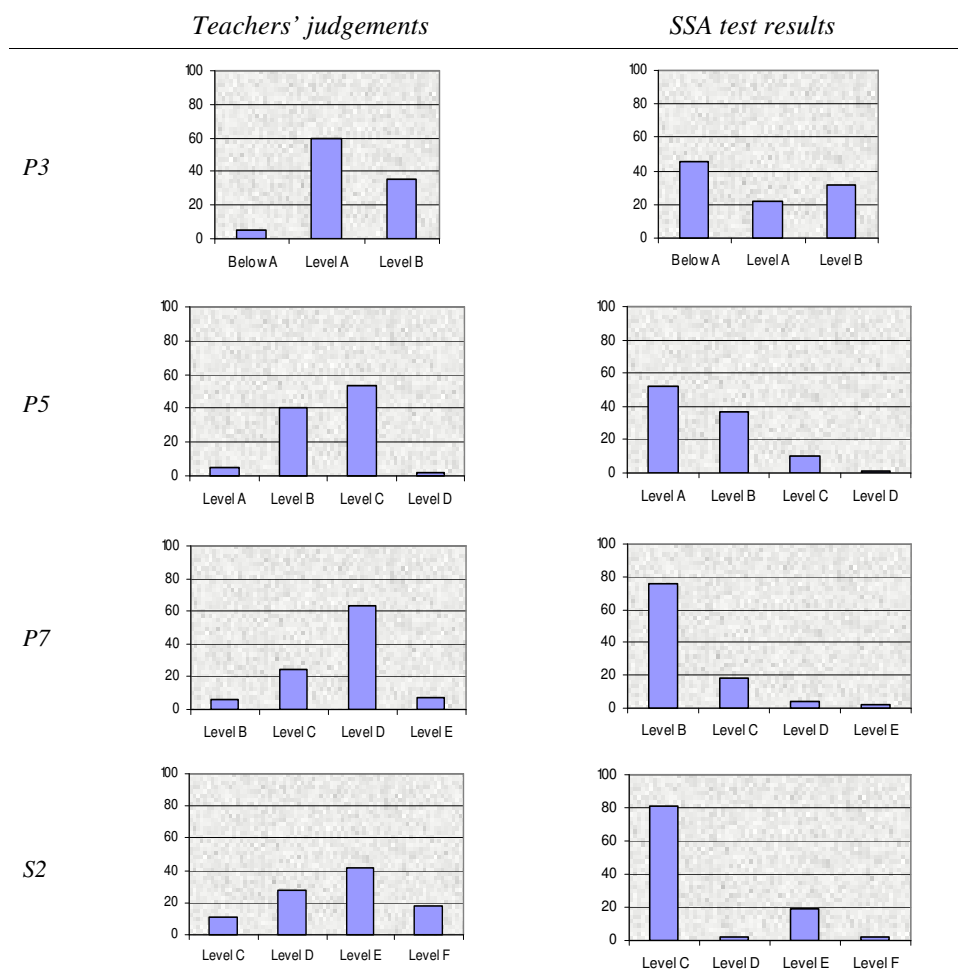
Table 4  
SSA test-based level profiles for science knowledge and understanding  
in 2007  
(% pupils categorised into each level – population estimates)

Stage	No. of pupils	<A	A	B	C	D	E	F
P3	8,809	46	22	32	0	0	0	0
P5	8,832	0	52	37	10	1	0	0
P7	8,978	0	0	76	18	4	2	0
S2	8,522	0	0	0	81	2	15	2

Source: These data have been produced by sequential subtraction of the 'well-established' proportions shown in Table B1, Scottish Government 2008, Chapter B.

Figure 3 illustrates the comparative picture for science across the four stages. At P3 the main shift in level profile is a downward pull from Level A to below Level A: the SSA tests classified over 45% of the pupils as being below Level A, whereas just 5% of the pupils were judged to be this weak in science by their teachers (Table 2). At all other stages the distributional shift is universally downwards. In particular, while over half of P5 pupils were judged by their teachers to be at Level C for science the tests put just 10% at that level, and while the teachers put just 5% of their pupils at Level A or below the tests put over half of them there. The same kind of picture, but even more extreme, emerged for P7: over 60% of P7 pupils were judged to be at Level D by their teachers compared with fewer than 5% by the SSA tests, and just 5% of the pupils were judged at Level B by the teachers compared with over 75% by the tests. At S2 we see a smoother distribution for teachers' level judgements, with a strikingly contrasting bipolar distribution for the tests. In particular the tests put the great majority (80%) of S2 pupils at Level C (or below) compared with just 11% for the teachers.

Figure 3  
 A comparison of level distributions for SSA testing of ‘science knowledge and understanding’ and teachers’ judgements for ‘science’  
 (% pupils classified at each level)



### Reflections on the findings

There are many reasons why we might expect a degree of difference between teacher judgments and test results. Teachers have at their disposal a broad range of assessment evidence collected over an entire session or longer, and that evidence potentially covers a wide variety of learning experiences drawn from all aspects of the curriculum and in different contexts. The test results inevitably comprise a more limited range of assessment evidence collected in two relatively short test sessions, and sampling only those aspects of the curriculum which are amenable to pencil and paper assessment. Moreover, because we have no prior knowledge of the ability of the pupils, the assessments have to cover a range of levels and cannot focus in depth on any particular one. In such circumstances we might reasonably expect some divergence between the test results and teachers' judgements. Perhaps, though, our expectation would be that the tests might rate pupils more highly than the teachers, given their narrower scope. This is not the case. And so we have to try to explain the direction of the differences as well as the contrasting assessment profiles.

In both mathematics and science (and indeed reading) teachers place more children at the expected levels for a stage than the test results do. Is it the case that teachers are restricting their expectations and hence their assessment of some pupils to the expected levels for the stage? Or is it that teachers are assessing different aspects of subject performance than the tests? For science, the difference between the test results and teachers' judgements are extremely marked. Indeed, teachers' judgements for science are more similar to their judgements for mathematics than they are to the

science test results. Perhaps in science teachers were judging process or 'investigation' skills more than the factual knowledge and understanding that the SSA tests were designed to assess?

Another aspect worth considering is the manner in which the mathematics and the science curricula are delivered. As a core subject, mathematics in primary school has a minimum time allocation of 15%, whereas science *shares* a similar minimum time allocation with other aspects of environmental studies, including outcomes relating to history, geography and modern studies as well as technology. In secondary schools the picture is different; mathematics has a minimum allocation of 10% and environmental studies 30%. Pupils in primary schools are likely to experience a limited number of science-based topics each year appropriate for their particular stage, and ranging across two or even three levels, than they are to work through the curriculum level by level. For example pupils might study simple concepts about forces in P3, covering Levels A and B, in P6 they might cover topics of friction and air resistance at Levels C and D and in S2 be introduced to forces and gravity. If science teaching is topic based rather than level related, how are teachers assessing a pupil's level? It is not surprising that there is a high correlation between their reported levels for science and those for a core subject like mathematics; a child who is Level B in mathematics (or reading) is likely to be judged as Level B in science perhaps on the basis of limited available class work supporting this judgement rather than there being evidence from across the range of strands and targets at a level. On the other hand, in secondary schools the teacher of science is not the teacher of mathematics, and here we similarly see a close correspondence between mathematics and science judgements, but very different test results. There are clearly other factors at work.

Could another possible reason for the greater disparity in the science results be related to the fact that, unlike mathematics, there are no National Assessments in science for teachers to confirm their judgements? While teachers are not required to use National Assessments (although some schools and local authorities may have a policy to do so), in reading and mathematics they have undoubtedly contributed over the 15 years since their introduction to a shared understanding of levels, as these have been interpreted by the test developers. With no comparable national resource for science, teachers are left to their own devices when interpreting the meaning of the levels in this subject.

Teachers' judgements provide one estimate of pupils' attainment. Tests administered through surveys such as the SSA provide another. Should we expect these to be the same or different? And if different, in what way? These are questions worthy of debate. The SSA has provided some concrete assessment evidence to support that debate. What is fascinating is exploring the different patterns that have emerged, and speculating on possible explanations. In doing so, however, we inevitably raise more questions than we provide answers.

## References

- Condie, R., Robertson, I.J. & Napuk, A. (2003). The Assessment of Achievement Programme. In Bryce, T .G. K. & Humes, W. M. (eds), *Scottish Education*. Edinburgh: Edinburgh University Press.
- Hayward, E. L. (2007). Curriculum, pedagogies and assessment in Scotland: the quest for social justice. 'Ah kent yir faither'. *Assessment in Education*, 14, 251-268.
- Hutchinson, C. & Hayward, E.L. (2005). The journey so far: assessment for learning in Scotland. *The Curriculum Journal*, 16, 225-248.
- Munro, L. (2003). National Testing and National Assessments. In Bryce, T .G. K. & Humes, W. M. (eds), *Scottish Education*. Edinburgh: Edinburgh University Press.
- Scottish Executive (2000). Environmental Studies – Society, Science and Technology. 5-14 National Guidelines. <http://www.ltscotland.org.uk/5to14/htmlguidelines/environmental/page3.htm>

Scottish Government (2004a). *5-14 Attainment in Publicly Funded Schools in 2003/04*. <http://www.scotland.gov.uk/Publications/2004/12/20331/47509>

Scottish Government (2004b). *Assessment, Testing and Reporting 3-14 - Our Response*. <http://www.scotland.gov.uk/Resource/Doc/26800/0023692.pdf>

Scottish Government (2006). *The 2005 Scottish Survey of Achievement (SSA): English Language and Core Skills*. <http://www.scotland.gov.uk/Publications/2006/06/29141936/0>.

Scottish Government (2007). *Scottish Survey of Achievement: 2006 Social Subjects (Enquiry Skills) and Core Skills*. [www.scotland.gov.uk/Publications/2007/08/15104710/0](http://www.scotland.gov.uk/Publications/2007/08/15104710/0).

Scottish Government (2008). *Scottish Survey of Achievement. 2007 science, science literacy and core skills – supporting evidence*. <http://www.scotland.gov.uk/Publications/2008/06/04092011/0>

Scottish Qualifications Authority (2003). *Core skills framework: An introduction – Numeracy*. [http://www.sqa.org.uk/files\\_ccc/CoreSkillsNumeracy%20241106.pdf](http://www.sqa.org.uk/files_ccc/CoreSkillsNumeracy%20241106.pdf)

SOED (1991). *National Guidelines: Mathematics 5-14*. Edinburgh: Scottish Office Education Department. <http://www.ltscotland.org.uk/5to14/guidelines/mathematics.asp>.

SOEID (1999). *National Guidelines: Mathematics 5-14 Level F*. Edinburgh: Scottish Office Education and Industry Department. <http://www.ltscotland.org.uk/5to14/guidelines/mathematics.asp>