

**National Assessment  
informing policy and practice:  
the Scottish Survey of Achievement**

Sandra Johnson - Assessment Europe  
AEA-Europe, Stockholm 2007

For a copy of the full paper, please send an email request to  
[Sandra.Johnson@Assessment-Europe.com](mailto:Sandra.Johnson@Assessment-Europe.com)

## The SSA: nature and purpose

- Programme of sample-based surveys of pupil achievement at selected stages in primary and secondary education
- Evolved from the sample-based Assessment of Achievement Programme (AAP), and replaced the annual 5-14 survey (census survey based on teacher judgement)
- To provide information about national and authority system outcomes, along with contextualising information about teaching and learning

Sandra Johnson - Assessment Europe  
AEA-Europe 2007

## Stakeholder groups

- National politicians, policy makers and school inspectors
- Local authority managers, policy makers and advisers
- Teachers and teacher trainers
- Educational researchers
- Parents and pupils

Sandra Johnson - Assessment Europe  
AEA-Europe 2007

## Principal objectives of the surveys to date:

- to produce national and authority estimates of pupils' subject achievement at P3, P5, P7 and S2
- to offer subgroup attainment comparisons
- to compare attainment over time (\*)
- to provide information about pupils' learning contexts

Sandra Johnson - Assessment Europe  
AEA-Europe 2007

## The scale of the surveys

Just under 30,000 pupils in total were tested in each of 2005 and 2006 (over 12% of each stage cohort), and almost 45,000 pupils in 2007 (over 18% of each stage cohort).

Sandra Johnson - Assessment Europe  
AEA-Europe 2007

## The scope of the surveys

- in 2005/2006, pupils took reading tests *or* numeracy tests
- in 2007 all pupils took science tests
- class-based writing was submitted for a subsample of pupils
- smaller subsamples engaged in practical assessments
- pupils and teachers completed questionnaires
- schools submitted teachers' level judgements for each of their sampled pupils

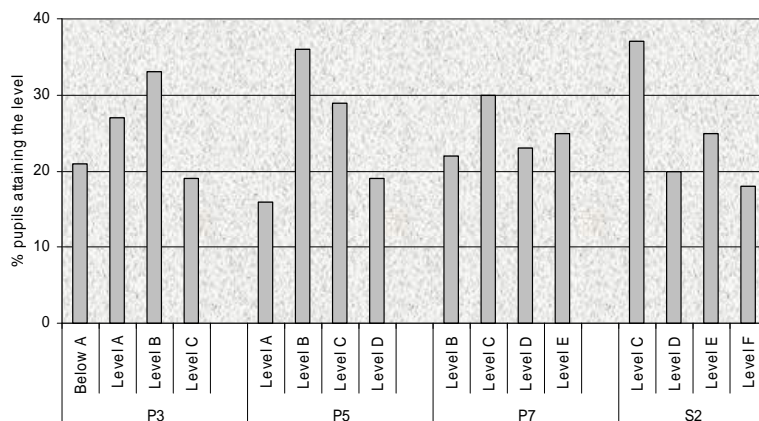
Sandra Johnson - Assessment Europe  
AEA-Europe 2007

## Assessing reading & numeracy

- Several tests were created at each level (A to F), allowing broad curriculum coverage
- Reading tests (single-level) took the form of a source text with related test questions
- Numeracy tests at a level comprised 'atomistic' items at that level distributed over two mixed-level test booklets.
- Items in the numeracy booklets were randomly ordered (not ordered by level); each booklet had two variants, differing in item order
- Booklets were randomly allocated to pupils using multiple matrix sampling; each pupil took items/tests at three different levels
- Cut-off scores were applied to make level attainment decisions

Sandra Johnson - Assessment Europe  
AEA-Europe 2007

## Reading attainment at four stages: the national picture (2006)



Sandra Johnson - Assessment Europe  
AEA-Europe 2007

Here we have the full set of level distributions for reading - for the four stages (8 year olds, 10 year olds, 12 year olds and 14 year olds, respectively, at the time of the survey).

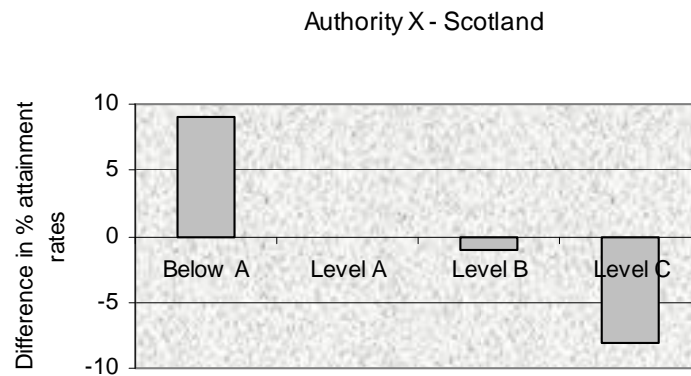
In total, 12 different tests at each of three levels were administered in the survey at each stage (e.g. levels A, B and C at P3), with each pupil randomly allocated a triplet covering the three levels. Each test took the form of a source text with associated test questions; text length and difficulty increased with level, as did the number of test questions (from 21 to 30 or more). Pupils were classified as having 'attained' a level if they answered at least 65% of the test questions correctly in their test at that level. Over 3000 pupils were tested in reading at each stage, and the standard error attached to each estimated proportion is around one percentage point (jackknife technique).

This is the 'bottom line' information that will be of particular interest to national and local policy makers.

The fact that three of the four levels that appear in any one stage distribution appear also in the distribution for the next stage up, means that there is great potential for exploring stage progression in reading attainment. Again, this is important information for policy makers, but also for other education professionals, including subject advisers, teacher trainers, educational researchers and teachers.

A corresponding chart is available for numeracy.

## Reading attainment at P3 (2006): an authority v nation comparison

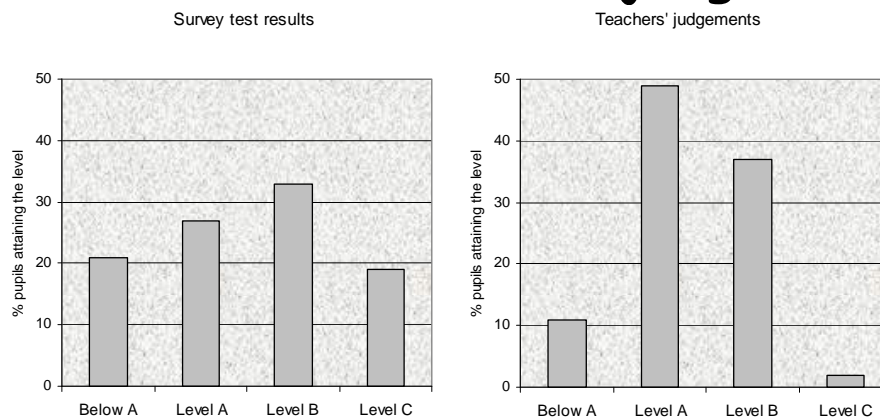


Sandra Johnson - Assessment Europe  
AEA-Europe 2007

The information presented in the previous chart at national level is also available, in the same form, for individual local authorities. This means that system managers and policy makers in the authorities now benefit from the same powerful system monitoring tool that hitherto has only been available to their colleagues in central government, albeit based on very much smaller samples of pupils - around 190 P3 pupils were tested in reading in authority X compared with over 3000 nationally.

In addition to reviewing the authority results in isolation, there is also the possibility of comparing the authority system with the national system. This chart illustrates just one way of making such a comparison in a form that all stakeholders could immediately understand. The chart shows that in authority X there are proportionally more P3 pupils at the bottom end of the level distribution, 'below level A', and proportional fewer at the top end, at level C. In other words, authority X pupils have done less well than pupils nationally at this stage. The difference in the general pattern is statistically significant. Most importantly, those authorities that elect in future surveys for separate attainment estimation in this skill area will be able to look back to see how, if at all, things might have changed - perhaps in response to educational initiatives launched within the authority in the interim.

## Reading attainment at P3 (2006): test results v teachers' judgements

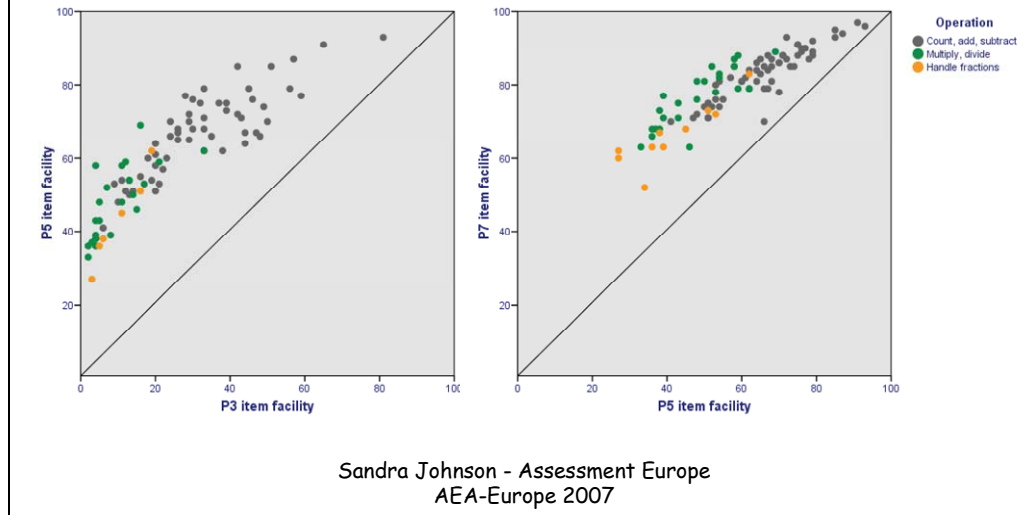


Sandra Johnson - Assessment Europe  
AEA-Europe 2007

The two charts shown above compare the test-based level distribution at P3 for 2006 (left-hand chart) with that based on submitted teacher judgements for the survey pupils (both distributions are based on weighted estimates of population proportions). The most immediately obvious feature in these charts is the difference in shape between one and the other. The test-based distribution is 'flatter' than the teachers' judgement distribution. While the teachers put most pupils at level A or level B - the levels they would expect P3 pupils to be at, according to the progressive level framework in Scotland's National Curriculum Guidelines - the SSA test-based results put proportionally more pupils below level A or up at level C. The almost 20% of P3 pupils demonstrating level C attainment on their reading tests surprised many education professionals. So, too, did the 5% of P3 pupils that demonstrated level C attainment in numeracy, when a mere handful of the 3000+ pupils tested at this stage in numeracy were put at this level by their teachers.

The same general picture emerged at P5. At P7 and S2, however, the pattern changed. At these stages the test-based results did not produce a flatter distribution, but rather shifted the teachers' distribution downwards. Some might argue that this is evidence that teachers are over-estimating their pupils' attainment levels at these stages, and reasons can be offered for why this might be the case. On the other hand, others could insist that teachers were right, and that it is the survey tests that have not done pupils justice. This is one of the issues that comparisons of this kind have inevitably raised - and it hinges critically on the very meaning of a 'level' in this context. This is but one example of the kinds of important debate that the SSA stimulates.

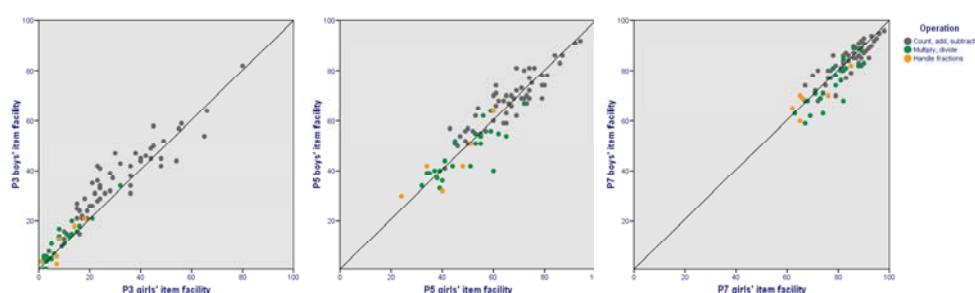
## Stage progression in numeracy, by test item (level C, 2006)



Over the two reading/numeracy surveys of 2005 and 2006, a total of around 100 reading tests, containing over 2000 individual test items, were administered, along with around 1000 different 'atomistic' numeracy test items. These items are equally spread over the six 5-14 levels and cover the curriculum concerned quite comprehensively. Moreover, each item was attempted by a representative sample of at least 300 pupils, and often more. The result of this scale of testing is a rich database of item-level data that has enormous potential for further informing policy formulation, in particular in terms of future curriculum development and teaching focus.

The charts shown here illustrate the picture of stage progression in numeracy at level C (around 90 test items administered at P3, P5 and P7), but many others could have been produced, both for numeracy and for reading. In this particular pair of charts, note the general increase in facility levels from one stage to another (P3 to P5 in the left-hand chart, P5 to P7 on the right), along with the age-related reduction in the range of item facility values. Note also the scatter in both charts, illustrating quite strong stage by item interaction effects (reflecting, among other things, the impact of changing curriculum focus, increasing familiarity with the different operations, changes in opportunities to practice applying the different operations, and so on). At all three stages, items focusing on multiplication, division or fractions proved harder than items demanding application of counting, adding or subtracting skills (whether featuring graphical displays or not).

## Gender and numeracy (level C, 2006)



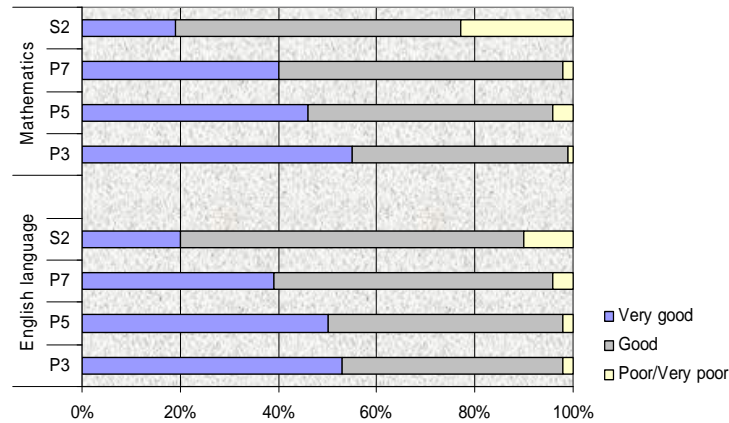
Sandra Johnson - Assessment Europe  
AEA-Europe 2007

This is another example of how item-level data can prove informative. These charts illustrate the picture of gender difference (or similarity) for level C numeracy at the three stages at which level C items were administered, *viz.* P3, P5 and P7.

A clear feature in the charts, as in the previous ones, is the rise in general facility values with age - not surprisingly the older the pupils the easier they find the items. Here also we see that items involving multiplication and division, along with those involving fractions, proved the most challenging at all stages, and for both genders. A third feature to note is a tendency for boys to have produced better performances than girls at P3 on most of the items at this level (level C is the highest level assessed at P3, and no gender difference emerged at levels A or B). The gender difference seems to disappear by P5: while there is quite a bit of scatter here, there is no preponderance of items one side of the line as opposed to the other. By P7, where level C was the lowest level assessed, there is, if anything, a suggestion that many of the multiplication and division items were slightly better done by girls than boys. We have evidence in these charts not only of age by item interaction effects but also of gender by item interaction effects.

It remains to explore pupil/item behaviour in much more depth, to draw out messages for policy makers, curriculum developers and teachers.

## Pupils' learning motivation (teachers' opinions, 2005)



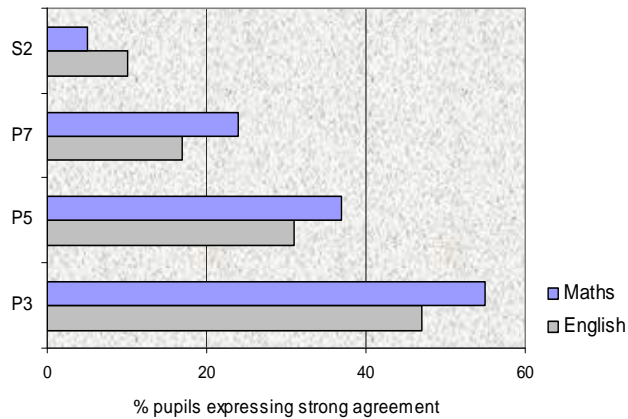
Sandra Johnson - Assessment Europe  
AEA-Europe 2007

The teacher questionnaire used in its various forms in the 2005 and 2006 surveys has provided a wealth of information about the context of teaching and learning, in English language, in mathematics and in social subjects (i.e. social subjects, environmental studies, history, geography, modern studies); the 2007 survey will soon add corresponding information for science.

In this chart we look only at English language and mathematics, and show the results for one aspect of the invitation to teachers to *Please indicate your opinion about your pupils' motivation to learn, behaviour and lesson attendance (in general)*. As the chart clearly shows, the proportion of teachers rating their pupils' motivation to learn as 'very good' steadily declines from P3 to S2, for both subjects. The only difference of note between the two subjects is the higher proportion of S2 teachers rating their pupils' learning motivation as poor or very poor in mathematics compared with English.

However, this is a chart based on teachers' *perceptions* of learning motivation in their classes. And it must be entirely possible that teachers teaching older pupils have more demanding expectations of what learning motivation looks like. What evidence do we have that these perceptions match reality? The answer lies in what pupils said in response to similar questions in their questionnaires (see next chart).

## "I look forward to lessons" (pupils strongly agreeing, 2005)

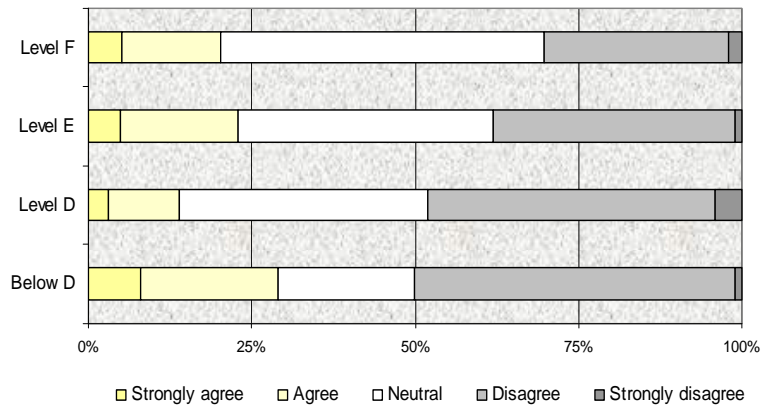


Sandra Johnson - Assessment Europe  
AEA-Europe 2007

Pupils at all stages were invited in their questionnaires to offer information about various aspects of their lives in and out of school. Among the enquiries was one which presented them with a list of statements about subject lessons, for each of which they were invited to express their degree of agreement or disagreement (using a 5-point Likert scale). Here are some examples of the statements presented: "Pupils settle down quickly at the start of our Mathematics lessons", "I find Mathematics easy to understand", "It is hard to catch up in Mathematics if I miss a lesson", "I look forward to Mathematics lessons".

The chart illustrates the results for "I look forward to Mathematics lessons" and, from the complementary questionnaire that focused on English language, "I look forward to English language lessons" (phrased for P3 as "I look forward to the work in language lessons"). The most striking feature in the chart is the very clear decline in enthusiasm for lessons in both subjects as pupils move through the primary school and eventually into S2. A second interesting feature is that while mathematics lessons appeared to be slightly more popular with pupils than English language lessons at every stage in the primary school, lessons in mathematics became less popular than lessons in English language at S2. This finding would seem to support what the teachers reported about pupil motivation to learn at the different stages in these subjects (see previous chart).

## "I look forward to lessons", by attainment level (S2 numeracy, 2005)



Sandra Johnson - Assessment Europe  
AEA-Europe 2007

Whenever surveys provide both attainment and questionnaire data, there is inevitable interest, among policy makers and others, in linking the two.

The chart shown here illustrates just one example: the association between numeracy attainment and enjoyment of mathematics, as indicated in responses to the statement "I look forward to mathematics lessons". Although the size of correlation between agreement ratings and test scores (as opposed to level classifications) was extremely small, well below 0.1, the chart nevertheless suggests that an association does exist. But it isn't the case that the more able the pupils the more they looked forward to their mathematics lessons. It is rather that the less able the pupils the less they looked forward to mathematics lessons.

A broad brush correlation sweep has not produced strong evidence of association in reading or in numeracy between pupil attainment and questionnaire responses. But clearly associations do exist. The challenge is to explore more deeply to detect them. What policy value any discoveries will have remains to be seen.

## The power of the SSA as a system evaluation tool

- Directly relates to Scotland's national curriculum, and focuses on the pupil stages of choice
- Comprehensively covers the curriculum in the surveyed subject
  - through the administration in surveys of hundreds of test items in numerous test booklets
- Reports pupil attainment in a way that stakeholders can understand
- Provides central government *and* local authorities with a valuable system evaluation tool, that as a bonus enables interpretable inter-system comparisons
- Is able to respond quickly to policy needs
- Avoids blanket testing

**But design challenges and measurement issues continue!**

Sandra Johnson - Assessment Europe  
AEA-Europe 2007

The survey reports for 2005 and 2006, respectively, can be accessed at:

[www.scotland.gov.uk/Publications/2006/06/29141936/0](http://www.scotland.gov.uk/Publications/2006/06/29141936/0) and

[www.scotland.gov.uk/Publications/2007/08/15104710/0](http://www.scotland.gov.uk/Publications/2007/08/15104710/0).