

Baseline assessment and monitoring in primary schools

Peter Tymms and Monika Wylde

Paper presented at the Symposium Connectable Processes in Elementary and Primary Section

Bamberg April 2003

Abstract

Monitoring Systems

There has been an explosion of interest in monitoring systems around the world and although the evidence for the positive impact of such systems has yet to be established (Visscher and Coe 2002) there can be little doubt that well-constructed, and well-used monitoring systems can be of enormous benefit to the educational provision of children. This article will distinguish different types of monitoring systems and argue that a particular way of monitoring is likely to be most beneficial to pupils. In doing so, it focuses on monitoring which is directly designed to help to improve the educational system by working with teachers, but it starts by looking at the layers of information, and at a Centre which has devoted itself to creating systems of the sort which are of interest to us.

Layers of information

At the most fundamental level of the education system is the child. It is information about the child and his or her progress, attainment levels, attitudes, behaviour and so on that are the very stuff of pedagogy. This is most likely to be known, at least in an informal way, by the teacher. Any system that is put in place to monitor these aspects of the child is simply formalising something that has been implicit to the actions in the classroom for years. But a teacher working with a whole class would find it very difficult to keep all of the information in his or her head, and so some kind of recording system, traditionally a mark book, is usually put in place. A modern monitoring system extends this idea by recording data that are common to all teachers. This enables one to spot children who are falling behind and watch for those who have special needs or consistently extreme levels of attainment. Once this becomes integrated with professional knowledge a powerful pointer for action has been created.

At the next level up, the head of department within a school, or perhaps the head within a school, can benefit from data that have been aggregated to the class or perhaps the department level to provide an overview. Again, this is something that

the principals of institutions have done for years, but putting it onto a systematic basis so that the head can operate a kind of observation post makes considerable intuitive sense. This was written about in Gene Glass's 1979 paper 'Policy for the unpredictable', where he suggests that our inability to know for certain the long-term impact of our actions, even in the light of high quality research, suggests a parallel with forest rangers who are ever watchful for fires. We need, he argues systems that allow us to spot difficulties whilst they are still small enough to be dealt with. Occasionally a fire will break out and somebody needs to do something about that. Spotting the smoke before it becomes a conflagration is vital.

At the next level up data on a series of schools in a district or a state become important, and again the same kinds of ideas apply. Aggregation can continue in this way until it reaches the level of the whole country so that one can see the states or the schools within it. Finally it is even possible to think about monitoring across countries, for example using TIMSS (Third International Maths and Science Study; see for example Howie 2001) or PISA (Programme for International School Assessment; <http://www.pisa.oecd.org/>).

The Curriculum, Evaluation and Management Centre

The Curriculum, Evaluation and Management Centre (CEM Centre) in Durham England, has evolved a particular way of working with schools using monitoring data that has been highly valued by schools. The work started with a project in 1983 known as COMBSE (Confidential Measurement Based Self Evaluation; Fitz-Gibbon 1995). It set out to provide schools with information about their own performance which would not be seen by others, but rather would be available to them so that they could improve themselves. This way of working, where schools could see the others' data but only with codenames became an important feature of the system. The CEM Centre now runs a number of large projects in secondary and primary schools. With 65 full-time members of staff, the participation of thousands of schools in the UK, and with a centre in New Zealand, many schools in Australia, and embryonic systems in other parts of the world, the CEM Centre is probably the largest educational monitoring set up in the world (see for example Fitz-Gibbon and Tymms 2002). Within that family of monitoring systems we focus in this paper on one particular part of it and that is the topic of the next section.

Performance Indicators in Primary Schools

The Performance Indicators in Primary Schools project (PIPS) deals with, as is indicated by the name, the progress and attitude of children in primary schools (see for example Tymms 1999; Tymms and Albone 2002). It runs in a way, which is designed to provide information for teachers about their pupils, and information for heads at elementary schools about the running of their school. It integrates with the statutory test data within England, or within any of the countries where it is operating. Pupils are assessed in the schools, the data are processed by the CEM Centre and these processed data are fed back to the school. At the heart of the system are baseline assessments which assess where the children start from. In this way attainment in the curriculum can be put into context. In each assessment we also pick up data on the developed ability of children by looking at vocabulary and non-verbal ability, as well as the children's attitudes and self-esteem.

Accountability systems and professional monitoring

As was hinted earlier, we make an important distinction between official accountability systems and professional monitoring systems. An official accountability system is put into place with the express purpose of holding schools and/or teachers to account. On the other hand, a professional monitoring system is put into place in order to generate data for the professionals to run a better system. (Tymms 1999) An example of an accountability system may be seen in England where a National Curriculum was introduced so that it would be clear what the teachers were supposed to do and then pupils are tested at the end of what were called Key Stages so that their progress in the curriculum can be judged. The data from statutory assessments given at the end of primary school and the end of secondary school are released to newspapers and school league tables appear on a yearly basis. Further an authoritarian inspection system, known as Ofsted, was put into place, and their judgements may not be challenged by anybody except Ofsted. The results of these inspections can have major consequences for the life of a school and the professional career of heads and teachers (see for example Fitz-Gibbon 1998; Shaw et al 2003).

This national system is, by comparison with other systems around the world, an extreme system and it is an accountability system. The tests at the statutory case are there so that the parents can check that the schools are doing a good job, so that Ofsted can make judgements about the schools, and so that market forces can operate within the system. Of course there are other uses of the system other than primary accountability purposes, but its main *raison d'être* is clear.

At another extreme are professional monitoring systems, which are designed to provide data to help the schools help themselves. In other words - to see which pupil is falling behind, which pupil is progressing rapidly, which class has difficulty with maths and so on. A very important element of this is the notion of value added, that is to say, the progress of one pupil compared to other similar pupils in other schools. By looking at pupils with similar starting points, fair comparisons are made and like is compared with like. The technique employed is regression analysis and the value-added data are residuals from regression analysis.

Although the accountability and professional models have been presented as extremes they are, in fact, at two ends of a continuum and one would expect the data in accountability systems to be used to help pupils in some circumstances and it is inevitable that the kind of data developed for professional monitoring is used to hold some teachers to account on occasions. It is nevertheless useful to make the distinction and also to think about what happens if there is a problem in a school. For if there is a difficulty in a school then in an accountability system the school or the teacher will want to hide that problem and to disguise what is going on. On the other hand, if they are part of a professional monitoring system, which they might have paid to join, they want to find problems. These are very different mindsets and have very different implications for the way the projects run and also for the quality of the data that are developed as time goes on.

On-entry baseline assessment

When children start school it is important for a monitoring system to know the levels at which they start. This section is about on-entry baseline assessments and it begins by looking at reasons why one might have them, in addition to monitoring, but

also the forms of assessment that might be used and the difficulty of assessing young children.

Baseline assessments have been around as long as there has been teaching and they have many purposes. One might be to get to know the children when they start; another might be to help to plan the curriculum; another to identify the special needs of children at an early stage so remediation programmes can be put in place. The principle reason for the set-up of the PIPS baseline assessment was to establish a baseline so that future progress may be looked at, not only for individual children, but also for whole classes and schools. These systems, of course, get intermingled with one another, but there are many purposes to a baseline assessment (see for example Wolfendale 1993; Tymms 1999).

Forms of assessment

Assessments used with young children can be broadly divided into two types. One is based on observation and the other on gathering objective data from specific tasks given to the child. In some quarters the choice between these forms is seen as a dichotomy between which a choice must be made. We see this is a false dichotomy. Teacher assessment is a given – it is going on all the time and it may well be that the more effective teachers are those that make better assessments (Tymms 1996). But on the other hand it is essential also to have objective assessments since teachers vary in their judgements. There are some things that observation cannot do. For example, it cannot provide fair comparative data across classrooms, nor can it probe indications of special needs in the way that objective data can. A child with precocious mathematical development may give no hint of it when observed, similarly incipient dyslexia may only show up with carefully chosen questions.

Difficulties of assessing young children

A good test for 16 year olds in mathematics is a long test. It is long because the more items it includes the more reliable it will be. Further, in order to get something that has good predictive validity it must have good reliability. When assessing a young child the same principles present problems. This is because young children are typically slow to respond and that restricts the number of items that can be asked. If one attempted to assess the child over a long period the child would lose interest and the validity of the assessment would then be open to question.

It may also be difficult because a young child cannot easily be assessed as part of a group. The need for individual assessment puts constraints in the classroom and may require extra finances. Further, when the child arrives at school for the first time there is no a curriculum on which to base the assessment.

Taken together these obstacles seem formidable but we believe that we have overcome most of them within the PIPS project. Much of the advance has resulted from taking an adaptive approach to the assessment (Tymms 2001). That is to say as children answer questions we decide whether to ask further questions or not on the basis of their responses. The assessment has been adjusted to run on computer and the choices are made automatically. This means that a very long assessment can be very short for an individual child. The PIPS on-entry baseline assessment takes about 20 minutes per pupil and this has proved to be easily manageable by schools although it is certainly something that they have to think about very carefully.

The assessment is designed to act as a basis for monitoring and as such it must have good predictive validity. To achieve this an empirical position was taken which started by looking at the research literature. This established what it is about a young child starting school that best predicts difficulties or successes later on. The assessment was then constructed from a series of units, each of which dealt with a major predictor. When put together this made an assessment with considerable depth.

PIPS on-entry baseline

The PIPS on-entry baseline is available in text format and it also runs on computer. It is being used in several different countries and in a number of different languages. It includes sections on name writing, vocabulary, concepts about print, phonological awareness, letter identification, reading, concepts about mathematics, simple sums, simple counting, digit identification, formal sums, and short-term memory (see for example Tymms 1999b). Together these constitute a powerful body of data when a child is assessed. The result is a test-retest reliability of 0.98. This was established by selecting a sample of pupils across several schools after teachers had carried out the assessment and then being repeated by a researcher. When the children take a reading or a mathematics test three years later the predictive validity of that is around 0.7 and it has even been shown to predict later performance at the age of 11 at the 0.7 level.

Examples of feedback

Often it is the teacher who does the assessment shortly after the child starts school although it does not have to be the teacher. The data are returned to the CEM Centre for processing and the results are returned to the schools. This feedback includes standardised scores in early reading, early mathematics and phonological awareness. The data are also presented graphically as box and whisker plots as shown below which give an indication of the spread of scores within a school. A stacked bar

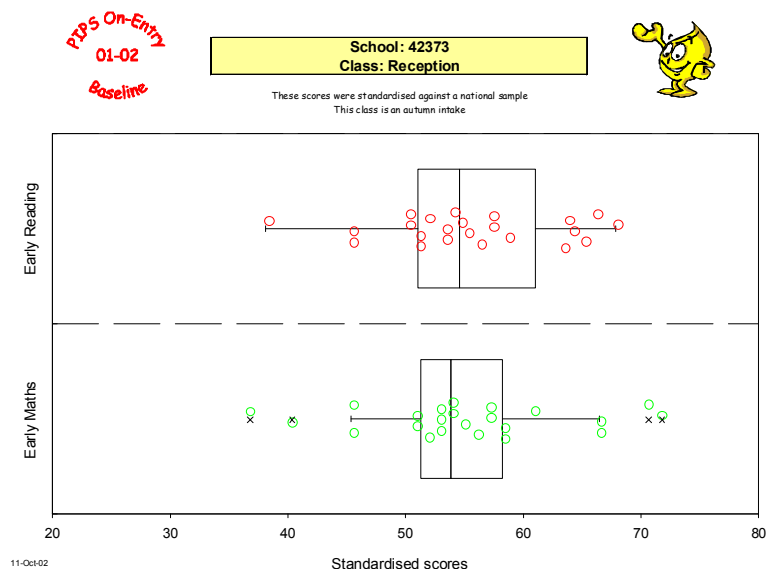
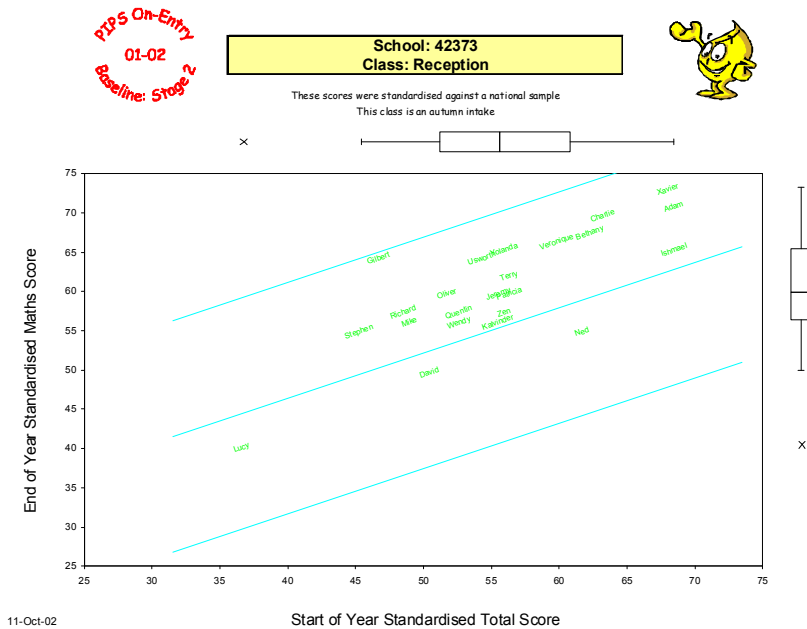


chart allows the teacher to spot strengths and weaknesses at an early stage.

The assessment is repeated with pupils at the end of the year and it is at this point that feedback can go to the schools revealing the impact of their efforts. The enormous progress of children during their first year in education can be seen and the progress of children can be tracked (Tymms et al 1997). It also becomes possible to

get a measure of the relative effectiveness of different schools. The data are fed back in a variety of formats but the most revealing is the scattergram showing the pupils' starting points and end points along with a line indicating the national trend. This graph can reveal individual progress as well as patterns at the school level. It can show achievement in line with national trends as well as differential effectiveness with sub-groups. Naturally, training is needed to help with the interpretation of the data and a large effort on the part of the PIPS project is directed towards enhancing the skills and knowledge base of teachers.



Some research findings

The data from PIPS is intended to help teachers and headteachers with their schools. This is a kind of concept of distributed research (Fitz-Gibbon 1995b) in which the CEM Centre picks up the information and then feeds it back later on. But the project incidentally picks up a lot of data and it has been possible to look at a number of relationships.

For example, we have been able to look at the progress of children over a year at school and to look at the relative importance of factors associated with that progress. We have been able to show the enormous variation from one school to another (typically from one teacher to another) and the central importance of the teacher (Tymms et al 1997). He or she is far more important, for example, than the home background to the progress that is made. The two key variables in predicating a child's academic standing at the end of one year at school are the child's starting point and who the teacher is. We have also been able to follow children up three years later, and show that that initial impetus, or lack of impetus, during that first year at schools stays with the child two years later (Tymms et al 1999).

Other projects have looked at other issues in the primary years. For example, at the end of the first year we collect data on attention deficit, hyperactivity and impulsivity using the American Psychiatric Association ADHD scale completed by teachers. We have been able to pick out children with ADHD-like characteristics at the end of the first year school. That kind of information is then related to the

progress of the children in later years and we have shown that such children tend to fall further and further behind in the curriculum. The question then arises: “what can be done to help such children?” To start to answer this question we have set up a funded project with a randomised design, which will investigate the value of giving advice to teachers about how to deal with such children and also investigate the possibility of a national screening program for ADHD like characteristics.

FIPS

In 1995 a sudden rise in parental concern about the educational success of their children occurred in the German School London. There was a lack of progress by some pupils and support in class for unusually able children was missing. Teachers agreed that a number of pupils experienced severe problems in more than two subjects often covering the span from languages to sciences. As a result some students were inattentive and disruptive during the lessons and had general behaviour problems.

For many years the School had provided additional tuition for what were regarded as weak children on a subject basis but tied very closely to the curriculum. No attempt had been made to identify the reasons each child was failing or to target the additional tuition to meet individual needs.

The school has an average yearly intake of approximately 100 students a year. About 20% of each annual intake was being tested to identify individual problems. The number of children who needed specialist tuition grew. By looking at the test results Monika Wylde discovered a number of cases where there was a significant gap between the child’s general ability and actual achievement. Out of the large group of children tested, 83 children showed problems, 80 had developmental problems, being mildly up to severely affected. Of the total 25 already attended the grammar school. The difficulties identified included poor achievement in language and maths and in a number of cases social and personality problems. Language difficulties also seemed to be a cause for low achievement in other subjects.

Looking at their educational career and achievements from school entry to date it became obvious to her that most children had experienced problems since their first days at school. The question arose whether the individual needs of those children were correctly recognised and addressed during their school days. She believed that a cognitively oriented on-entry assessment could identify where a child stood as it started school. Teachers could then use the information provided by the test to plan each child’s future educational provision.

The test had to identify children with potential learning difficulties as well as unusually able children quickly and objectively. It also needed to cater for bi-lingual and multi-lingual children who attended the school. In short the requirement was for a clear picture of the strengths and weaknesses of each child and which provided an indicator as to the level each child was ready to be taught. There was no such test available in Germany.

She had seen the PIPS assessment as something that could be valuable in German schools. She contacted the CEM Centre in Durham and set about translating the assessment, which has now been used with German children both in London and also

in Germany where it is known as FIPS. The test itself is administered, whilst the form teacher, to criteria compiled by the school, also observes the child. This dual approach of assessment and observation has provided sufficient information on most children with the result that the number of specialist assessments, that had shown an exponential growth before the introduction of FIPS, has been reduced to an absolute minimum.

We have collected data from the project and have been able to feed that back into a number of situations and are able to show that it works as well in German as it does in English.

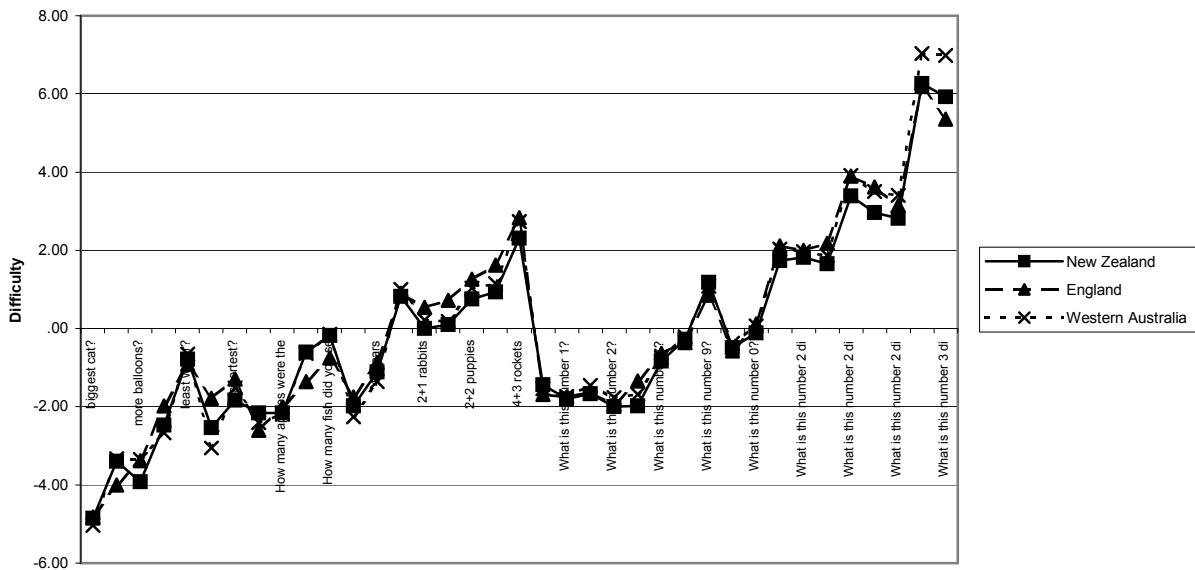
Anecdote

One key feature of this is that it can sometimes show teachers things that they previously had not known. In one German nursery there was a girl who had been showing behaviour problems for some time. The teacher had not really got to the bottom of the problem. However, following the FIPS assessment it became clear that the child was exceptionally bright and academically well ahead of her group for her age. The problem had been development and resultant boredom. Once that was understood the situation was remedied.

International comparisons and possibilities

Having collected a large amount of data in New Zealand, Australia, Holland, Scotland and England we are able to look at the relative progress of children or the levels of attainment of children at the ages of 4 and 5 in the different countries and find remarkably similar profiles in those different situations. We also find the same pattern of difficulty and ease of items in different situations and the chart below shows the relative difficulty using Rasch scaling of the items on the mathematics part of the PIPS baseline assessment from a number of different countries. One can see a remarkable correlation between them suggesting a common developmental pathway through which children go in early mathematics development in all countries.

Maths (first part)



This leads to the possibility that one could actually have a project across many nations looking at the progress of children which would help to inform the interpretation of other data such as TIMSS and PISA. It would also provide data which could have important implications for countries examining their pre-school provision for young children.

References

- Fitz-Gibbon, C. T. (1995). Monitoring School Effectiveness: simplicity and complexity. *Merging traditions: the future of research on school effectiveness and school improvement*. J. Gray, D. Reynolds, C. T. Fitz-Gibbon and D. Jesson. London, Cassell.
- Fitz-Gibbon, C. T. (1995b). "Measuring Value Added in Schools: Distributed Research." *Education and Training Statistics*: 184-207
- Fitz-Gibbon, C. (1998). "Ofsted: time to go?" *Managing Schools Today* 7(6): 22-25.
- Fitz-Gibbon, C. T. and P. B. Tymms (2002). "Technical and Ethical Issues in Indicator Systems: Doing things right and doing wrong things." *Education Policy Analysis Archives* 10(6).
- Glass, G. V. (1979). "Policy for the unpredictable." *Educational Researcher* 8(9): 12-14.
- Howie, S. (2001). Mathematics and Science Performance in Grade 8 in South Africa 1998/1999. Pretoria, Human Sciences Research Council.
- Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: Their impact on academic achievement and progress. *British Journal of Educational Psychology*, 71, 43-56.

- Shaw, I., D. P. Newton, M. Aitkin and R. Darnell (2003). "Do OFSTED Inspections of Secondary Schools Make a Difference to GCSE Results?" *British Educational Research Journal* **29**(1): 63-75.
- Tymms, P. (1996). Theories, Models and Simulations: School Effectiveness at an Impasse. *Merging Traditions: The Future of Research on School Effectiveness and School Improvement*. J. Gray, D. Reynolds, C. Fitz-Gibbon and D. Jesson. London & New York, Cassell: 121-135.
- Tymms, P., Merrell, C., & Henderson, B. (1997). The First Year at School: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3(2), 101-118
- Tymms, P. B., C. Merrell, Henderson, B.. (1999). "Baseline assessment and progress during the first three years at school." *Educational Research and Evaluation*.
- Tymms, P. (1999a). *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-added Indicators*. London: David Fulton Publishers.
- Tymms, P. B. (1999b). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading*, 22(1), 27-36
- Tymms, P. (2001). "The development of a computer-adaptive assessment in the early years." *Educational and Child Psychology* **18**(3): 20-30
- Tymms, P. and S. Albone (2002). Performance Indicators in Primary Schools in. *School Improvement Through Performance Feedback*. A. J. Visscher and R. Coe.(eds) Lisse/Abingdon/Exton PA/Tokyo, Swetz & Zeitlinger: 191-218.
- Visscher, A. J. and R. Coe, Eds. (2002). *School Improvement Through Performance Feedback*. Contexts of Learning: Classrooms, Schools and Society. Lisse/Abingdon/Exton PA/Tokyo, Swetz & Zeitlinger.
- Wolfendale, S. (1993). *Baseline Assessment: A Review of current practice, issues and strategies for effective implementation*, OMEP.