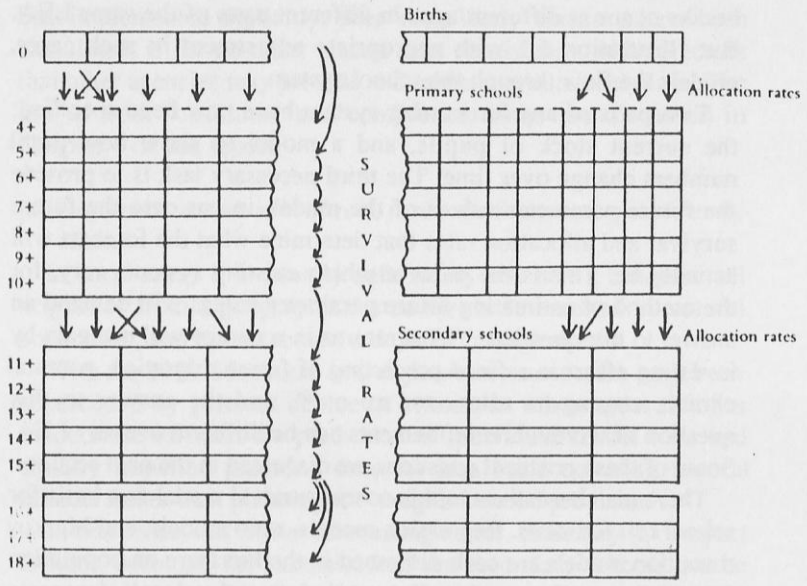
A common computer system for school roll projections for educational planning

This document outlines a core and sufficient system for school roll projections. This introduction introduces the model. Then Section A outlines the core elements in a projection, and some guidance for system development. Section B outlines the datasets that a user would need to specify.

The diagram below, taken from the DES-funded report of 1987[1], remains a good summary of the requirements for school roll projections.

A representation of the school system as it is modelled by LEA school roll forecasts



Key: *Survival rates* represent gains and losses to the LEA school population. Between birth and primary school they are sometimes known as *arrival rates*. For children beyond compulsory school leaving age, they are equivalent to school *staying-on-rates*.

Allocation rates represent the distribution of pupils between schools, including the effects of parental choice of school.

[1] Forecasting pupil numbers for educational planning: a guide to school roll forecasting in the planning processes of local authorities in Britain, Stephen Simpson with David Lancaster, report to the Department of Education and Science, Sheffield Papers in Educational Management no. 71, Sheffield City Polytechnic, Centre for Education Management and Administration, 165pp, 1987. Out of print, but available in local authorities, libraries, or second hand. The diagram is from p.83.

This proposal is underpinned by other experience and appraisals of school roll projection systems, including the DfE guidance issued up until 2010[2], reviews from ScotXed[3] and Wales[4] in recent years, and advice to a variety of local authorities in the past three decades. These confirm that Education Authorities face very similar demands for projections of school rolls, that the data manipulation required to achieve projections are also similar, but that the variety of school structures and available data sources have so far prevented the development of a common system for use by many authorities.

The 1987 report discusses how different local datasets and approaches to projections can be incorporated into a common computer system for use in different local authorities. There are some approaches which are so distinctively different that they could not be accommodated in a common approach: in particular these include approaches which develop an all-age population projection for individual households (micro-simulation, eg West Lothian) or for very small areas, before deriving school roll projections as one by-product. These occasionally-used population-based models involve a level of complexity unsuited to many authorities.

In contrast, the great majority of Education Authorities use the annual statistics of past school rolls as the main basis for school roll projections. Together with data for infants or births in areas relevant to primary schools, the annual series of past school rolls provides most of the information needed for robust school roll projections, and additionally allows an annual check on the accuracy of projections which helps to improve subsequent projections and gives important context for users. Nonetheless, some links to population projections can and should be made as part of a projection based on past school rolls.

Following the advice of most appraisals, the strategy is to design a simple and robust system with premises that can be easily understood, but with flexibility to allow further complexities if local conditions demand them. While it is essential to respect local knowledge and to build in the facility to easily bring local knowledge to bear on specific schools, it is also important not to waste a lot of time on considering individual adjustments to assumptions that are never going to be accurate. The aim is to identify what can be predicted in an automated fashion, and to distinguish assumptions that can be improved by local knowledge.

The platform for a projection is not detailed here. It would benefit from being a software which each education authority is likely to be familiar with, such as Microsoft Office or Open Office. It could be based entirely within a database application such as Access, or a spreadsheet application such as Excel, or a combination of these. When such decisions are made, they should aim to allow maximum ease of use such that the software is flexible enough to produce all functions and scenarios that most would find useful, while minimising the work involved.

---

[2] http://webarchive.nationalarchives.gov.uk/20100608061208/http://www.teachernet.gov.uk/docbank/index.cfm?id=14699

[3] Pupil Forecasting Seminar, Thursday 3rd November, Summary of workshop discussions, ScotXed, 2005.

[4] Investigation of best practice in forecasting primary pupil numbers, Anne Evans, Pembrokeshire County Council, project report for University of Glamorgan, undated, c2004/5.
http://www.virtualstaffcollege.co.uk/wp-content/uploads/VSCEMDPWalesMDPAnneEvans.doc

A. The core elements in a projection

Referring to the diagram above, the following two sets of 'counts' and five sets of 'rates' are the core elements essential and common to all local authorities in the UK:

Two core sets of counts:

> *Potential primary pupils, in relevant areas*. These are figures for the past and future population of children in areas relevant to each school. 'Births' in the diagram may be replaced by information on infants from the health service (not only at age 0). The areas for statistics of births (or infants) should be relevant to the schools, but availability will differ between authorities: school catchment areas, or postal sectors, or statistical areas (OAs or LSOAs in England and Wales, Data Zones in Scotland), or individual grid-referenced or post-coded records that have been aggregated to areas relevant to each school.
>
> *School rolls*. Past and current counts of pupils in each school, disaggregated by age group.

Five core sets of rates:

> *Allocation rates to primary intake*. These allocation rates define the relationship between potential primary pupils and actual numbers of pupils arriving in the first year of school. There may be one-to-one or many-to-one or many-to-many relationship between the areas for which potential primary pupils are counted and the schools. The rates will be calculated from past statistics of potential primary pupils and school intake. Postcoded pupil records can also provide a relationship between areas and schools. They will usually reflect 'parental choice' and 'placement requests'.
>
> *Cohort survival rates to primary intake*. As indicated in the diagram, survival rates apply not only from each school year to the next, but also from potential primary pupils to school intake, when they might be called 'arrival rates' or 'admission rates'. Sometimes the allocation rate to primary intake (above) has already included this survival rate, reflecting children migrating from birth, or not going to any state school. The reception year of primary may not be complete, and should be handled separately, perhaps as a proportion of the following year's full primary intake, or with a separate survival rate from potential primary pupils.
>
> *Cohort survival rates in school*. Once in school, a year-group tends to stay the same size. However, it can change due to migration to and from the area, including as a result of housing developments, and due to transfers to and from other schools. The experience of past years is sufficient to accurately estimate each school's cohort survival rates in school, which are ratios that are 1.0 if there is no net impact on a year group from year to year, or more or less than 1.0 to reflect net increase or net decrease.
>
> *Allocation rates to secondary intake*. These allocation rates will reflect feeder systems, as well as 'parental choice' and 'placement requests'. They will normally be calculated from past annual records of the origins of each secondary's intake, or could be calculated by

default from the past relationship between a secondary's intake and the size of the final year of its closest or 'feeder' primaries.

Staying on rates. Survival rates beyond compulsory age schooling are easily calculated from past years' school rolls.

A system suitable for many authorities would need to allow a user to easily identify:

- Labelling of school years (but a default of P1 etc in Scotland and Y1 etc in England would satisfy most education authorities)
- The tier structure of schools, such that in some sub-areas three tiers were present (but two tiers with fixed ages between primary and secondary would satisfy most authorities)
- Categories of schools for reporting purposes: groups of schools for subtotals (often geographical areas); types of schools (denominational of various types; other types).
- The optional delineation of a catchment area, to explicitly count the impact of out-of-area placements ('Placement Requests', 'parental choice').

A system could optionally extend to:

- The projection of a single school separately from others.
- The inclusion of capacity figures to report on implied numbers of classrooms, and mismatches with current capacity.
- Nursery places, children's centres
- Integer rounding, such that all results were whole numbers of pupils in each year group, which sum exactly to school totals, without biasing results through 'losing' the cumulative effect of small changes. This rounding might be for each calculation, or of final results.
- The projection of a range around best estimates of future rolls, with justifications.
- The inclusion of special schools (currently it is normal to exclude these from the projection of school rolls).
- The further use of capacity figures to reallocate children when capacity is exceeded.
- Test scenarios of schools opening, closing, merging.

Denominational schools will normally be dealt with in the same way as other schools. Their relationship with potential primary pupils will usually reflect a broader geographical intake.

Schools outside the state system will normally be excluded (private schools, English academies etc), but their inclusion is likely to improve the forecasts if those schools take many of their pupils from the education authority area.

B. The datasets and computer system

The user would have to provide data relevant to each school, and the two core sets of 'counts':

1. School name and identifier(s), school type(s), relevant primary area(s) for primary schools, relevant primary schools for secondary schools.
2. School roll for each combination of school identifier, age group, and minimum of five past school years.
3. For each relevant area, the potential primary pupils corresponding to the past school years' intakes, and as many further years as have been born.

An automated projection

From these three datasets, a robust school roll projection can be automatically calculated:

a. All five core sets of 'rates' required for a school roll projection (section A above), calculated by default from the experience of the past three years.

b. A school roll projection, with clear and justifiable assumptions based on recent past experience, that can give confidence with a relatively low input of resources.

c. Summaries of results with a variety of standard tables,

d. Diagnostic analyses to highlight where elements of the projection are based on past experience which has been unstable and therefore contain uncertainty (such as migration rates fluctuating from year to year),

e. Summaries of accuracy created for projections run from a school roll earlier than the latest,

f. For some local authorities, these projections will be sufficient for all their regular needs.

For example, the *allocation rates to primary intake* would be calculated automatically by dividing the five most recent years of intake, by the five years of potential primary pupils in those areas that the user has identified as relevant for the school. In cases of rural catchment areas this may be very suitable. In urban areas it will result in one overall proportion of a school's relevant areas being projected to attend that school, which will provide a reasonable projection.

Amending the automated projection

Users would be able to import their own evidence for each of the five sets of rates (section A above).

For example, if other information allows the *allocation rates to primary intake* to be calculated directly, so that a school may take all the children from some areas and a proportion of other areas, then this information could be inserted by the user and may improve the projection.

Adding further datasets

Further datasets would be optional, but may be used routinely by some education authorities. The two datasets envisaged relate to housing and parental choice:

Housing. The 'pupil yield' from housing developments is the most vexed question for those who project school rolls. School head teachers and parents are concerned that there be sufficient places for children living in new housing, and councillors want developers to be aware of the implications of their building, sometimes to insist on a contribution to school building should it be needed. However, repeated research evidence shows that the pupil yield is lower than expected for at least two reasons: to some extent families from existing housing spread out to new housing without there being a net increase in pupils; additionally, the allowance for 'survival rates' in the projection already accounts for the impact of the level of new housing experienced in the recent past. It is only the *net* effect of *raised levels* of house-building that needs to be taken into account.

Even then, the level of the yield from new housing is debated and in reality does vary according to the type of development, which cannot always be foreseen when projecting new housing.

With this in mind, the automated projection defined above may suffice, together with details of past and future housing in the school area, to be used by education planners as contextual information. However, some authorities will have sufficient political pressure to require that housing information be incorporated directly within school roll projections. The common school roll projection system should therefore allow for this, by helping the user specify past and future housing in each school's area, together with a pupil yield that will be used to amend the projection.

Parental choice. Parental choice can be coped with in an automated fashion, assuming that the experience of the recent years will continue. This experience is included in the automated calculation of the *allocation rates* to primary and to secondary intakes.

However, in some authorities the impact of parental choice will be politically necessary to identify even if it is not technically necessary for a good projection. This may be especially likely where there are delineated catchment areas and requests are made explicitly by parents of children to be placed across these boundaries, as in all of Scotland. In those cases, the common school roll projection system should allow for the specification of past levels of net parental choice or placement requests for each school, and for a projection of these to be identified in the projection results. The projection would be automated by continuing recent experience, but easily over-ridden by the user.

Finally, it is worth emphasising that the use of local datasets brings with it knotty issues of their quality and completeness in all areas of the education authority. A common computer system for school projections can help to identify such issues, by identifying unusual changes from year to year or between schools, but cannot improve the datasets themselves. There will always be a significant outlay of resource to understand and compensate for poor quality datasets. The system would require guidance notes on how to use alternative and additional datasets. On the other hand, the advantage of a common computer system proposed here, is the use of core datasets that are generally of good quality and will be sufficient for a defensible school roll projection.