Information systems in education: An interactive model for projecting primary school enrolments

E Gould
University of Wollongong
P Casperson
formerly NSW Department of Education

Sophisticated information systems are poised to play an important part in educational administration. Fundamental to the success of these systems is a means of accurately predicting school enrolments. In this paper an enrolment projection methodology is described for predicting numbers of primary school children in NSW five years into the future. Beginning with preschool age children, the expected number of kindergarten enrolments are forecast and these cohorts are progressed through the first six years of primary school. The model described is implemented on a microcomputer and uses an interactive technique which enables human intervention in order to take full account of local knowledge in predicting the numbers in each year group. Control on totals is maintained by allowing groups of schools to be amalgamated and by applying to these groups similar routines to those which are applied in obtaining projections for individual schools. This counteracts the effects of student mobility across wider areas and overcomes the problems associated with simple aggregation of individual school year group numbers. The technique provides a valuable planning tool when enrolment figures are needed for future decision making.

Education occupies a dominant position in most of the western worlds economies. In spite of this, the means by which the taxpaying shareholders can judge the effectiveness of their largest industry is made well nigh impossible by the lack of any recognisable balance sheet (OECD, 1983).

Predictions that accountability in the form of "performance indicators" will be a central issue of twenty-first century education (Birch & Smart, 1989) are likely to put increasing pressure on educational administrators to

produce the type of information which is capable of ensuring that taxpayers recognise value for money (Wholebren, 1986). This, combined with the other dilemma facing administrators, namely, responsiveness to political imperative (McPherson, 1986) point the way towards a more judicial use of computerised information systems in support of the decision making process.

The use of computers is firmly entrenched in educational administration mainly in the areas of transaction analysis and electronic data processing (EDP). Less use is made of this powerful tool in decision making in spite of the growing awareness within the administrative community of such computerised aids as Decision Support Systems (DSS) or Management Information Systems (MIS).

Educational administration can benefit from developments in the provision of information by considering ways of incorporating existing transaction processing and other information systems into the decision making activities of administrators. Assistance in pursuing the tasks they face is a primary focus of any MIS/DSS. Neither should be seen however as a mechanisation of the proper decision making function of human administrators but rather as a means of improving their performance. This can only be achieved by careful integration of the information systems along structured lines.

A formal information network that links each task and individual to a common (information) data pool must be developed as it has long been recognised that information handling cannot be left to the vagaries of an informal information system (Mellor, 1976).

Clearly a key element in this increased demand for better, more efficient management is access to accurate information at all levels in the decision making process. A fundamental requirement in any educational information system is an accurate prediction of student population not only in global (macro) terms but at the individual school (micro) level.

Predictions based only at the macro level (such as population growth, redistribution and mobility) are of limited use to planners faced with making decisions at the individual school level (Rowland, 1983). Accommodation needs for example, require planning based on the availability of continuously updated demographic information for relatively small geographic areas. This information requirement cannot be met from macro level sources such as population census data which is available only at five yearly intervals.

A micro level approach to predicting school enrolments has been adopted by the NSW State Department of School Education in Australia which has

responsibility for approximately three quarters of a million students in more than 2000 government schools. These schools are grouped into clusters within the State's ten education regions and account for approximately 75 per cent of the total school age population with the remainder attending a variety of non government schools. A system consisting of interactive computer programs has been developed to predict enrolments for each Primary (elementary) school in the State.

In overview the system can be seen to consist of two major parts.

The first part is a five year prediction of prospective kindergarten enrolments at each government school. Medicare (a national health scheme) data is used as input as it indicates the distribution of preschool age children in each school's feeder area by relating the number and age of children to their home address (ie. postcode/ zipcode).

The second part uses these kindergarten enrolment predictions along with data related to the children who are already at school in order to predict student numbers in each grade as children move through the school system over a period of five years. A simplified block diagram of the system is presented in Figure 1.
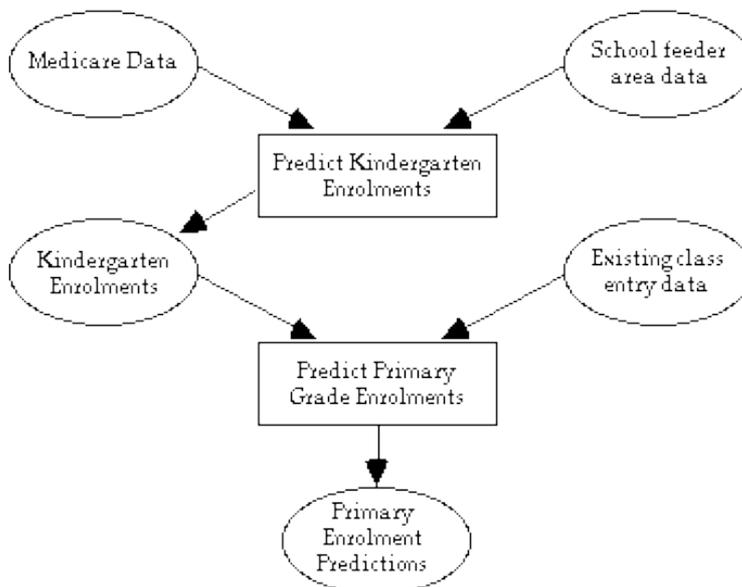
Figure 1: Simplified Grade School Enrolments System

Prediction of kindergarten enrolments

The past six years historical Medicare data obtained from the Department of Community Services and Health provides the age group populations for each postcode and is the initial input for this section. For example Table 1A shows the historical age group populations for postcode 9999.

Table 1A: Historical Age Group Populations 0-5 Years for Postcode 9999

Year	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5
1986	502	428	445	466	419	452
1987	407	488	447	432	459	444
1988	510	468	496	440	515	466
1989	490	513	455	481	515	506
1990	489	515	506	453	478	429
1991	482	522	525	529	465	469

In order to predict the number of five year olds in this postcode for the next five years (ie. 1992-1996) it is necessary to "progress" the 0-4 year olds from one year to the next. This is accomplished by calculating the proportion of children aged A in year Y who will progress to age A + 1 in year Y + 1. This proportion is called the retention rate (RR in the following tables) and is calculated by dividing the population of age group A + 1 in year Y + 1 by the population of age group A in year Y. For example the population of age 1 in 1987 divided by age 0 in 1986 gives a retention rate of 488/502 = 0.97 indicating that 0.97 of those aged 0 in 1986 "progressed" to age 1 in 1987. These retention rates are calculated for all the data in Table 1A and are shown in Table 1B below.

Table 1B: Retention Rates for Postcode 9999

	Cohort Age Group				
Year	0-1	1-2	2-3	3-4	4-5
1986-1987	0.97	1.04	0.97	0.98	1.06
1987-1988	1.15	1.02	0.98	0.96	1.02
1988-1989	1.01	0.97	0.97	0.99	1.02
1989-1990	1.05	0.99	1.00	0.99	0.99
1990-1991	1.07	1.02	1.05	1.03	0.98

A number of methods are available for calculating an average retention rate for each age group to be used subsequently in making the projections. One obvious method is to average the five retention rates for each age group and apply this to give the next year's projection. Alternatively, the last two or three years retention rates could be averaged to emphasise more recent trends. Yet another method which makes use of smoothing techniques applied to seasonally varying data is a weighted average where larger weights are applied to most recent years. A useful three year weighted average is one where the most recent year is weighted three, the next most recent weighted two and the third most recent weighted one. The respective retention rates are added together after being multiplied by these weights and the sum divided by six. Theoretically any of the above averages could be the most suitable for use depending on the perceived change to age profiles prevailing in a particular postcode. For this reason the system displays a selection of these average retention rates as shown in Table 1C.

Table 1C: Average Retention Rates by Various Methods for Postcode 9999

Retention Rate Method	0-1	1-2	2-3	3-4	4-5
2 Year Av.	1.06	1.00	1.02	1.01	0.98
3 Year Av.	1.04	0.99	1.00	1.00	1.00
3 Yr. Wtd. Av.	1.05	1.00	1.02	1.01	0.99
5 Year Av.	1.05	1.01	0.99	0.99	1.01

One advantage of having a choice of techniques is that confidence in the predictions is increased particularly if different methods give similar results. Marked differences in the figures are a signal that care needs to be exercised in selecting the most plausible average retention rate.

To make the system as flexible as possible the design is such that the initial calculation of future 5 year olds for all the postcodes is done in batch mode (usually over night) using the simple average of the last two years retention rates. Experience has shown the two year average retention rate to be the most appropriate for providing the initial set of figures for the majority of postcodes with relatively stable populations. Modifications of course, can be made to those postcodes with abnormalities and this is explained further on. Table 1D shows these predictions for 1992-1996.

Table 1D: Projected Number of Children 1991-1996 for Postcode 9999

Year	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5
1991	482	522	525	529	465	469
2 yr ARR	1.06	1.00	1.02	1.01	0.98	
1992		511	524	536	534	458
1993			512	534	541	526
1994				523	540	533
1995					528	531
1996						520

The first row of figures in this table are the 1991 actual age group populations from Table 1A and are used as the starting point for the predictions.

Although retention rates are displayed to two decimal places the computer holds many more significant figures and uses these more accurate values to perform the calculations. Only whole numbers of prospective students are displayed in the table which explains the slight differences if the retention rates shown are used to progress the age groups.

Each age group is "progressed" by multiplying its population size by the corresponding retention rate. For example the 1991 age 0 population of 482 is multiplied by the two year average retention rate for the 0-1 age group giving a projected population of 511 one year olds (482 x 1.06) in 1992. Similarly this age 1 population in 1992 is progressed to 2 year olds in 1993 and so on until the final column is complete and shows the age 5 population for the years 1992 to 1996.

Regional demographers inspect these figures and using local area knowledge may either accept them as accurate predictions, select another calculation formula from Table 1C better suited to the postcode or enter their own manual figures. This interactive process is fully recursive in the sense that continuous modifications can be made to a postcode population. This ability to perform on screen fine tuning is an important design feature of the system and allows the full application of local knowledge. Experience has shown that many postcodes do not need modification as they have fairly stable populations but the ability to modify the few that are not stable is critical.

When changes are made to a postcode group either by the selection of an alternative average retention rate or by manually entered figures, the effect on macro totals has to be carefully watched so that the sum of the parts

never exceeds the projections for the system as a whole. For this reason the new regional totals resulting from the recalculation of the postcode figures are displayed on screen so that demographers can see the effects of their changes before committing them to the file. The calculation process therefore, can be a series of on-screen fine tunings done interactively until the demographer is satisfied with the results. Judgement and local knowledge play a vital part in this process, controlled always by the system reporting the results of user changes for the individual postcode and any aggregation required. This approach of using the whole as a control over individual parts is an important feature of the system and applies to all tables of projections.

When all the fine tuning is complete and demographers are satisfied with the predictions, the projected population figures for age 5 are apportioned to each school in the postcode. To do this postcodes first need to be matched with schools using interactive submodules which accommodate problems such as children attending schools out of their local zone, children moving out of the state system to non-government schools, or changes made to postcode boundaries by Australia Post. This depends on good local knowledge by the regional demographer particularly in cases such as the dezoning of schools which was implemented recently and for which good historical data is not yet available.

Once the schools have been assigned to postcodes and other checks on input data are complete the significant correlation between postcode population trends for five year olds and kindergarten enrolments is used to calculate the proportion of five year olds expected to enrol in each school for the following five years. Past experience has shown that the three year weighted average technique as outlined earlier produces the most consistently accurate results for kindergarten projection. For this reason it is used here to predict the next five years kindergarten enrolments for each school using the three previous years actual enrolments. Five years previous data is supplied for use if manual intervention is required. Table 2 shows a typical postcode containing three schools (X, Y and Z in this example).

The first row of figures shown boxed in the following table are the actual postcode totals from Table 1A and the projected figures from the last column in Table 1D. The left hand side of the table shows the actual kindergarten enrolments for the last five years (only the last three are actually used) with the proportion each constitutes of the total (eg. 124 is 0.292 of the 425 total for School X in 1989). The three year weighted average proportion of each school shown in the seventh column is used to give the predictions for the next five years. For example for School X, 0.287 is successively multiplied by each of the postcode totals to give its projected kindergarten enrolments (ie. 0.287 x 458 = 131). This is done for each school with the results shown in the right hand side of Table 2. Not

all students in each postcode attend state schools hence the total proportion is rarely 1.0.

Table 2: Projected Kindergarten Enrolments for Each School

Postcode: 9999	Actual Kindergarten Enrolment						3 Yr Wtd Av Prop	Kindergarten Projection Using Three Year Weighted Average of the Proportions				
Schools Included	1986	1987	1988	1989	1990	1991		1992	1993	1994	1995	1996
Postcode Total From Tables 1A, D	452	444	466	425	429	469		458	526	533	531	520
School X Proportion	110 0.243	122 0.275	116 0.249	124 0.292	120 0.280	136 0.290	0.287	131	151	153	152	149
School Y Proportion	92 0.204	75 0.169	84 0.180	85 0.200	73 0.170	101 0.215	0.198	91	104	106	105	103
School Z Proportion	77 0.170	70 0.158	75 0.161	70 0.165	82 0.191	77 0.164	0.173	79	91	92	92	90
Total Pupils Total prop	279 0.617	267 0.601	275 0.590	279 0.656	275 0.641	314 0.670	0.657	301	346	350	349	342

The compilation of this table completes the kindergarten prediction process. These figures can now be used as input to the primary enrolment projections described below.

Prediction of primary enrolments

The techniques for obtaining primary projections are similar to those used to give kindergarten enrolments, but this time the fundamental unit of data is the size of year cohort group. The past six years historical enrolments are obtained from each school as shown in Table 3A.

Table 3A: Actual Enrolments K to Year 6

School: X

	Year Group						
Year	K	1	2	3	4	5	6
1986	110	108	107	106	100	103	109
1987	122	116	105	94	100	101	97
1988	116	135	104	97	90	100	103
1989	124	131	119	99	96	85	103
1990	120	141	129	111	93	96	84
1991	136	115	130	114	108	87	99

Retention rates are calculated as each year group advances from one year to the next (eg. K to year 1 from 1986 to 1987). This is shown in Table 3B and is calculated in exactly the same manner described previously (see Table 1B).

Table 3B: Retention Rates for School X

| Year | Cohort Age Group | | | | | |
	0-1	1-2	2-3	3-4	4-5	5-6
1986-1987	1.05	0.97	0.88	0.94	1.01	0.94
1987-1988	1.11	0.90	0.92	0.96	1.00	1.02
1988-1989	1.13	0.88	0.95	0.99	0 94	1.03
1989-1990	1.14	0.98	0.93	0.94	1.00	0.99
1990-1991	0.96	0.92	0.88	0.97	0.94	1.03

Again the average retention rate is calculated by a number of methods for each of the transition groups (ie. K-1, 1-2, etc) as shown in Table 3C.

Table 3C: Retention Rates by Various Methods

Retention Rate:	K-1	1-2	2-3	3-4	4-5	5-6
One Year	0.96	0.92	0.88	0.97	0.94	1.03
2 Year Av.	1.05	0.95	0.91	0.96	0.97	1.01
3 Year Av.	1.07	0.93	0.92	0.97	0.96	1.02
3 Year Wtd. Av.	1.05	0.94	0.91	0.96	0.96	1.02
5 Year Av.	1.08	0.93	0.91	0.96	0.98	1.00
5 Year Absolute (see Note)	9.40	-8.80	-9.80	-4.00	-2.00	0.20

Note: The five year absolute retention rate shown here in the last row of Table 3C is to cater for the case of small schools where the population in one or more grades may be zero or near zero and the use of percentage changes is meaningless. Under these circumstances the average of the actual numbers of students rather than the percentage change must be used. For example the -8.80 figure in the 1-2 column of this table indicates that over five years the average number progressing from year 1 to year 2 has declined by 8.8 students.

The principal method used in the calculation of the projections is the two year average due to its high correlation with actual figures in schools not affected by abnormalities.

This two year average is now applied to the last years figures (1991 in this example) to project enrolments for the next five years (see Table 3D).

Table 3D: Projected Enrolments Using 2 Year Average Retention Rates

Year	K	1	2	3	4	5	6	Total
1991 (actual)	136	115	130	114	108	87	99	789
ARR		1.05	0.95	0.91	0.96	0.97	1.01	
1992	131	142	110	118	109	105	88	805
1993	151	137	136	100	113	105	106	851
1994	153	159	130	123	95	109	107	877
1995	152	161	151	118	118	92	110	900
1996	149	160	153	137	113	114	93	910

The kindergarten figures shown boxed are the projections from Table 2 for school X in years 1992-1996. Each of these are "progressed" using the K-1 average retention rate (ARR) to give the year 1 figures (eg. 136 x 1.05 = 142, 131 x 1.05 = 137, etc). Again population figures are displayed to the nearest whole number. Continuing with this process the year 1 figures are progressed to year 2 and so on to complete the table.

Regional demographers again check these projections and arc given a number of choices. They can alter either the kindergarten figure or the retention rates. Alternatively they are able to manually amend the enrolment projections for any existing school or enter figures for new schools to be opened during the planning period.

Finally when projections for all regions have been completed the data is sent to the State Central Head Office for amalgamation into cluster groups, regional and state totals.

There are inherent weaknesses in obtaining predictions for cluster groups and regions by simple aggregation of individual school enrolments. Compounding of errors and population mobility between different geographical areas must be considered. To overcome these problems two sets of results are prepared for the cluster group, region and the state. The first set is the sum of the individual schools projections. The second is produced by treating each group of schools as if it were one big "school" and applying the same procedures to amalgamated year group totals. Because larger numbers are included this second grouping is statistically more accurate than the first and can be used as a control comparison to allow modifications to individual schools as desired. As well, these aggregated figures can be compared with independently derived results of demographic projections for the region and state calculated using macro

techniques involving the use of state and national population figures as well as immigration and population mobility between the states.

Final observations

Early versions of this system used a centralised mainframe computer to produce tables which were then mailed to regional demographers for checking and amendment. This proved both time consuming and cumbersome and so a change to an interactive microcomputer based system has been effected with a notable improvement in efficiency. Copies of the software and data are made available to regional demographers so that projections can be produced locally and subsequently forwarded to the Head Office for compilation. Since the time-consuming drudgery of performing the prediction calculations manually has been eliminated it is possible to offer demographers the choice of using one of several prediction formulae depending on the characteristics of each school or group of schools. If none of the formulae is suitable the option exists to use a manually derived set of figures.

Data relating to the number of students in any year for all schools is readily obtainable from school census forms. As this data is collected annually by the Department of School Education no special problems exist with its format or integrity. Predicting kindergarten enrolments on the other hand is dependent on the availability of data relating to the age distribution of preschool children in the feeder areas of each school. Nationwide census data is obviously the most accurate source but is only collected every five years and takes a further two years to be collated and published. Birth statistics, while giving good indications of the number of babies born in a particular area, suffer from problems relating to the time-lag mobility of families. Medicare data has proved to be the most stable and accurate source of this data as it is a reflection of this mobility, is readily available and contains the age and location (by postcode) of all children in the 0-19 age bracket.

The built in flexibility of the system has proved a bonus for the system due to recent upheavals in the State Education system under the "Schools Renewal" program. The system has been able to cope with such changes as the dezoning of schools and the devolution of responsibility to schools so that it is still as useful today as it was when it was designed in 1983.

The projections of enrolments are not an end product in themselves but because of the role they play in determining staff and accommodation needs are an important component of an educational information system.

Clearly such projections could also play a vital role in decision support modelling where the consequences of change to, say the staffing formulae for schools can be evaluated for a number of different options and

appropriate decisions taken with some knowledge of the effects. Any system used to model education costs would need enrolment projections in order to calculate and validate the cost drivers.

Careful monitoring of the system has taken place since its mainframe implementation in 1983 and numerous minor modifications have been implemented in order to improve its efficiency.

References

Birch, I. & Smart, D. (1989). Economic Rationalism and the Politics of Education in Australia. *Journal of Education Policy*, 4(5).

Kemmis, S., McTaggart, R., Smyth, J. & Ross, K. (1988). Monitoring School Performance: Testing and Alternatives to it. *The Australian Administrator*, 9(2), April.

McPherson, R. B., Crowson, R. L. & Pitner, N. J. (1986). *Managing Uncertainty, Administrative Theory and Practice in Education*. Merrill.

Mellor, W. L. (1976). Structure and Rationality in Educational Information Decision Systems. *Journal of Educational Administration*, May.

NSW Department of Education, Division of Management Information Services (1987). *Primary School Enrolment Projections System on Open Access 11: Users Manual.*

OECD (1983). *Educational Planning - A Reappraisal*. OECD, Paris.

Pollard, A. H., Yusuf, F. & Pollard, G. (1974). *Demographic Techniques*. Pergamon.

Pressat, R. (1978). *Statistical Demography*. New York: St Martins Press.

Rowland, D. T. (1983). Population and Educational Planning: The Demographic Context of Changing School Enrolments in Australian Cities. *Educational Research and Development Committee Report No 36*. Canberra: Australian Government Publishing Service.

Sprague, R. H. (1986). *Decision Support Systems*. Prentice Hall.

Taeuber, K., Bumpase, L. & Sweet, J. (1978) *Social Demography*. Academic Press.

Walberg, H. (1979). *Educational Environment and Effects: Evaluation, Policy and Productivity*. Berkeley: McCutchan Publishing Co.

Wholebren, B. E. (1986). Strategic Planning for the Design and Development of Management Information Systems in Education. *AFIPS Conference Proceedings*. Virginia: AFIPS Press.

Authors: E. Gould, Department of Business Systems, University of Wollongong, PO Box 1144, Wollongong NSW 2500, and P. Casperson, formerly Senior Education Officer (Demography), NSW Department of Education, Sydney NSW 2000.