Journal of Educational Planning and Administration
Volume XII No. 4, October 1998. pp. 417-427

Completion, Retention and Drop-out: Measuring Children's Prospects

P.G. Vijaya Sherry Chand
A.H. Kalro

Abstract

This article discusses the impact of selected factors on the relative chances of children dropping out of school or completing their schooling. The relative importance of age at the time of entry into school, sex, caste, father's education, family income and participation in the mid-day meal scheme in predicting the drop-out is analysed through a set of logistic regression models. The importance of the coefficients in predicting the odds ratios for different categories of each predictor are also evaluated. Implications for the management of primary education are then drawn.

The issue of children dropping out of school before completing their primary schooling cycle has been a persistent and much debated problem at the national level (Naik 1969; Government of India 1990; World Bank 1997). The issue is no less serious in Gujarat (Visaria *et al* 1993; Vijaya Sherry Chand and Shukla, 1993). How do children who have dropped out, differ from those who complete schooling within a specified time period? How important are caste and sex identities, and family background, in determining membership in these two groups? How critical is age at entry into school in determining the probabilities of being in either group? How do the factors believed to promote educational attainment (Lockheed, Verspoor and others 1991; Heneveld and Craig 1996), like attendance in pre-primary schools, homework support and regular mid-day meals offered in the schools, affect group membership? This article attempts to answer these questions in the context of Banaskantha district, which suffers from the highest rates of drop-out in the state (Government of Gujarat 1996).

Educational Status and Its Predictors

The educational status outcome considered in this article is position of a child, at the end of seven years after beginning school, either in a drop-out category or a non-drop-out category. The latter category includes two groups of children, those who have completed the primary cycle of five years and those who are still in school at various grade levels. We begin by looking at the chances of being in the drop out population, in relation to being in the in-school or passed group (Model A). We then ignore those children who are still in school and contrast the chances of dropping out versus completing schooling (Model B). The analytic approach is fitting a set of logistic regression models in which

the outcome variable is the log-odds (relative chances) of being in either of two groups, with the predictors being child-related socio-economic and demographic variables. Initially, the following predictors were proposed: age at entry into school, caste status, sex of the child, levels of parental education (father and mother), family income, attendance in a pre-primary school, support received in matters of homework, migration, and participation in the mid-day meal scheme. Some of these had to be dropped, as will be explained later.

The outcome variable, p, is the probability of obtaining either of the outcomes, derived from a non-linear function of the best linear combination of predictors: $p_i = e^u / (1+ e^u)$, where p_i is the estimated probability that the i^{th} case is in one of the two groups and u is the linear regression equation $c + \Sigma b_j X_{ij}$, with constant c, parameters b_j, and predictors Xj for k predictors (j = 1...k). The model may also be expressed as the log of the odds, (p/(1-p)), (Kmenta 1986: 550-53; Tabachnick and Fidell 1996: 575-76),

$$\log (p/(1-p)) = c + \Sigma b_j X_{ij}.$$

In Model A, the probability of the outcome of being in the drop-out group in relation to the in-school and passed group will be considered, and in Model B, the probability of the outcome of being in the drop-out group in relation to only the passed group will be considered.

Sample and Data

For purposes of this study, the 1988-89 batch of students who entered grade one in Banaskantha district was selected.[1] A sample of four schools was selected from each of the eleven talukas, from among all the district-panchayat run upper primary schools; thus 44 schools were selected in all from out of the 960 such schools in the district. All the schools in each taluka were ranked by the educational inspectors in charge of the taluka on a performance scale of zero to 100, ranging from extremely poor to very good. The ranking drew upon the reports prepared by the inspectors and their assessments of the enrolment and retention performance of the schools. The ranked schools in each block were then divided into four equal, or almost equal, groups: good, high-medium, low-medium and poor. One school from each group was then selected at random. All the children who entered grade one in 1988-89 in these 44 schools constituted the sample of children. In all, there were 1595 children.

Out of these 1595 children, six had died during the seven year period of the study, and another 191, from 37 of the 44 schools, had left their schools after obtaining leaving certificates. The records of these 197 children were not considered. Information on the examination outcomes of 37 students, for one or more years, was missing. These 37 records were deleted. Thus, 1361 cases, 418 drop-outs and 943 in-school or passed children, were retained for analysis.

Age was calculated from the date of birth of the child. Homework support received by the child during the last five years of the seven-year period was initially gathered according to the source of such support. Since there were very few cases under most of the sources of support, homework support was recoded as a dichotomous variable. Annual family income (mean of annual incomes of the last five years of the seven year period)

was recoded into four categories: 'high' of more than Rs. 36,000, 'middle-high' of Rs. 18,000 up to Rs. 36,000, 'low-middle' of Rs. 10,500 up to Rs. 18,000, and a 'low' category of less than Rs. 10,500. Participation in the mid-day meal scheme was worked out from the attendance notes kept in the schools. Different levels of education reached by the mothers were also obtained, but the very low frequencies of the non-illiterate levels necessitated a dichotomous variable with the categories of illiterate and educated. The number of days the family migrated during a year (mean of the last five years of the seven year period under study) were considered for the migration variable.

The predictors initially considered are listed in Table 1.

TABLE 1

List of Predictors Initially Considered

No.	*Predictor*	*Description*
1	Age at entry	Completed age at time of entry into school
2	Pre-primary experience	Dichotomous: attended or not attended
3	Caste status	Discrete: General, Scheduled Caste or Other Backward Class
4	Father's education	Discrete: Secondary/higher education; elementary education; illiterate
5	Sex	Dichotomous: boy or girl
6	Homework support	Dichotomous: regularly supported or not
7	Income	Discrete: low; low-middle; middle-high; high
8	Mid-day meal	Discrete: availed 70%; partly availed (70%); not availed
9	Mother's education	Dichotomous: illiterate or not
10	Migration	Number of days in a year

Data Screening

The analytic procedure proposed — logistic regression — is relatively more flexible as far as assumptions of linearity, normality and homoscedasticity are concerned. Thus, screening was limited to checking for accuracy of data entry, missing values, frequencies (especially splits for dichotomous variables) and univariate outliers. Screening was done on the grouped data, drop-out (DO) and in-school/pass (IP) populations. The dichotomous variables were screened for their splits, group-wise. Pre-primary experience (attended to not attended) had extreme splits of 10.4 per cent to 89.6 per cent in the IP group and 6.9 to 93.1 per cent in the DO group. Homework support (received to not received) also had extreme splits of 15.3 per cent to 84.7 per cent in the IP group and 5.1 to 94.9 per cent in the DO group. Thus, these two variables had to be dropped, since the scores in the smaller category would have been much more influential than the scores in the larger category and because the correlation coefficients of these variables with other variables would be truncated (Rummel 1970). Mother's education (educated or illiterate) also had to be dropped because of extreme splits of 13.2 per cent to 86.8 per cent in the IP group and 3.3 to 96.7 per cent in the DO group. The splits in this variable reflect the very low levels of female literacy in the district. The migration variable also had an extremely high

proportion of children who had not migrated. Thus, four variables in all were dropped from the analysis.

For the analysis of Model A, 31 cases with missing values on father's education, income and participation in the mid-day meal scheme (24 in the IP group and seven in the DO group) were excluded. Thus, finally, for the analysis of Model A 919 in-school/ pass children and 411 drop out children were considered. For Model B, 28 cases with missing values (21 in the passed group and seven in the drop-out group) were excluded, leaving 750 passed children and 411 drop-out children to be considered in the analysis.

Since a goodness-of-fit test to compare observed and expected frequencies was proposed to be used, expected cell frequencies for all pairs of discrete predictors including the outcome variable were checked. The condition of less than 20 per cent of the expected frequencies of less than five was satisfied. Subsequent analysis did not show very large standard errors for the parameters, nor were there convergence problems, indicating the absence of multicollinearity; the continuous variables were, however, checked for multicollinearity. The criterion of all conditioning indices less than or equal to 30 and not more than one variance proportion greater than 0.50 was satisfied.

Coding Procedures

Data management and analysis were carried out with SPSS for Windows statistical package. A consistent coding strategy for the dichotomous and discrete variables was adopted. The drop-out group was treated as the response group since the focus of the study was on the chances of falling into the drop-out category. The reference group was the non-drop-out group. That is, the logistic regression equation was solved for drop-out coded 1. Therefore, for all dichotomous and discrete variables, categories expected to be more closely associated with drop-out were coded with higher values, so that the category coded 0 was the reference group. This was done to facilitate interpretation of the coefficients and the odds ratios.

Model A: Drop-out and In-school/Passed Groups

A test of the full model with all the six predictors against a constant-only model was statistically significant (Table 2), indicating that the predictors taken together reliably distinguished between drop outs and in-school/pass children. Prediction success was impressive for in-school/pass children (91.40), and for drop-out children it was 52.80, giving a good overall prediction success rate of 79.47 per cent.

TABLE 2

Model A, Log-Likelihood and Goodness of Fit

Constant only model:	-2 Log Likelihood = 1644.722		
Model with all predictors:	-2 Log Likelihood = 1223.808		
	Goodness of Fit 1451.026		
	Chi-Square	df	Significance
Model Chi-Square	420.914	12	.0000
Improvement	420.914	12	.0000

Classification Table for Model A[2]

Observed	*Predicted*		*Percent Correct*
	In-school/Pass	*Drop-out*	
In-school/pass	840	79	91.40
Drop-out	194	217	52.80
Overall			79.47

Given that caste appeared to be not associated with the outcome (Table 3), a model without caste was also evaluated. The log-likelihood for this model was 1230.887. This [Chi square = 2 (log-likelihood of 6 variable model log-likelihood of 5 variable model)] works out to 14.16, at df 3. This is significant (p.005), indicating a significant decrease in drop-out prediction as a result of dropping caste.

The variables, along with the coefficients and their standard errors, the Wald statistic[3] and the significance based on it, and the odds ratios, are presented in Table 3.

TABLE 3

Model A: Educational Status: Drop-Out Versus In-School/Pass

Variable	*Coefficient*	*S.E.*	*Wald*	*df*	*Sig*	*R*	*Exp (b) (Odds ratio)*
Age	.6226*	.0822	57.4025	1	.0000	.1835	1.8637
Caste							
General			6.9349	3	.0740	.0238	
SC	-.1687	.2957	.3254	1	.5684	.0000	.8448
OBC	.3555	.1904	3.4854	1	.0619	.0301	1.4268
ST	.0756	.3611	.0438	1	.8342	.0000	1.0785
Father's education							
Secondary/high			27.1946	2	.0000	.1188	
Elementary	1.0589*	.3847	7.5761	1	.0059	.0582	2.8832
Illiterate	1.6436*	.3830	18.4170	1	.0000	.0999	5.1739
Sex (Girl)	1.2476*	.1489	70.1597	1	.0000	.2036	3.4820
Income							
High			31.2410	3	.0000	.1239	
Middle-high	.8385*	.4192	4.0002	1	.0455	.0349	2.3129
Low-middle	1.2533*	.4091	9.3835	1	.0022	.0670	3.5018
Low	1.7367*	.4125	17.7221	1	.0000	.0978	5.6787
Mid-day Meal							
Availed		2	69.4162	2	.0000	.1994	
Partially	1.6872*	.2038	68.5140	1	.0000	.2011	5.4043
Not availed	.0964	.2091	.2123	1	.6450	.0000	1.1012
Constant	-8.2077	.7138	132.2113	1	.0000		

Notes: 1. R is the partial correlation between the outcome and each predictor variable.
2. Significance based on Wald test (which is a function of coefficient divided by standard error) is indicated by *.

From Table 3, it appears that higher age at entry, poor education and income levels of the father, and partial attendance in the mid-day meal scheme, are significantly associated with greater chances of being in the drop-out group than in the in-school/ passed group. A discussion of the coefficients and the odds ratios of these variables follows. In general, if a coefficient is more than twice its standard error, it may be said that it is significantly different from zero. As explained earlier, the coding procedure followed enables comparison of a response group with the reference group. For instance, if the educational status of a particular caste group, say OBC, was associated with the other predictors in the same way that it is for the reference group (General category children), then the OBC coefficient would not be significantly different from zero. A negative coefficient would mean that the particular group had higher chances of avoiding drop-out status. The odds ratios are easier to interpret since they are the exponentiated values of the log-odds ratios (the coefficients). Thus, the odds ratio indicates the relative changes (decrease if the ratio is less than one) in the odds of being in one outcome category when the value of the predictor increases by one unit.

Caste status is not significantly associated with the outcome. This indicates that, controlling for other predictors, children belonging to the SC, OBC or ST categories do not show significantly higher chances of being in the drop-out category as compared to the General category children. Interestingly, in comparison with General category children, SC children have higher chances (though not significantly higher) of avoiding the drop-out outcome.[4] Whether this is translated into completion of schooling, is an issue which is not addressed in this study.[5]

Age at entry into school is strongly associated with the outcome with the odds of getting into the drop-out category almost doubling for every year increase in the age at entry. Sex is even more strongly associated, with girls, controlling for other predictors, 3.5 times as likely to join the drop-out group as boys. This is an indication of the severe disparities that girls in general face during their progression through school.

Father's education is very strongly associated with the outcome, with children of fathers with elementary education almost three times as likely to be in the drop-out category as children of fathers with higher educational levels, and children of fathers with no education almost five times as likely to be in the drop-out category as children of fathers with more than elementary education. This indicates an inter-generational continuity in matters of poor educational performance and reinforces the need for evolving alternative home support mechanisms for learning. A pattern similar to that seen in the influence of father's education is evident in income, with the lower three quarters of the income distribution disadvantaged in relation to the uppermost quarter.

An interesting pattern emerges when we consider the role of participation in the mid-day meal scheme. Whether a child avails of the scheme or does not, does not really matter. The problem is with children who are irregular in their participation in the mid-day meal scheme. These children are almost 5.4 times as likely to fall into the drop out category as those who avail of the benefits of the scheme.

Model B: Drop-out and Passed Groups

The focus of analysis in this section is a comparison of the drop-out versus non-drop-out outcomes with the drop-out versus schooling completed outcomes. First, a test of the full model with all the six predictors against a constant-only model was fitted and found statistically significant (Table 4), indicating that the predictors taken together reliably distinguished between drop outs and passed children. Prediction success was impressive for passed children (88.80) and drop-out children (59.37), giving an overall prediction success rate of 78.38 per cent.

TABLE 4

Model B: Log-Likelihood and Goodness of Fit

Constant only model:	-2 Log Likelihood = 1509.046		
Model with all predictors:	-2 Log Likelihood = 1086.107		
	Goodness of Fit 1271.892		
	Chi-Square	df	Significance
Model Chi-Square	422.939	12	.0000
Improvement	422.939	12	.0000

Classification Table for Model B

Observed	*Predicted*		*Percent Correct*
	Passed	*Drop-out*	
Passed	666	84	88.80
Drop out	167	244	59.37
Overall			78.38

As with Model A, a model without caste was also evaluated, but found to be significantly different from the full model (Chi square of 10.84, at df 3, p). The significant variables, along with the coefficients and their standard errors, and the odds ratios, for Models A and B, are presented in Table 5.

From Table 5, we find that the coefficients in the two models are of similar magnitude. Age at entry and sex are the only variables which have a lower coefficient in Model B, implying that the disadvantages imposed by a higher age at entry and being a girl are more severe in the case of children who are still in school at the end of seven years. Otherwise, the expected pattern of greater odds of being in the drop-out group for the significant response categories in relation to the reference categories holds. The changes in the odds are particularly high in the case of children with illiterate fathers and whose utilisation of the mid-day meal is partial.

TABLE 5

Significant Coefficients, and Odds Ratios: Models A and B

Variables	*Model B*			*Model A*		
	Coefficient	*S.E*	*Odds Ratio*	*Coefficient*	*S.E*	*Odds Ratio*
Age at entry	0.5942	0.0859	1.8117	0.6226	0.0822	1.8637
Father's education						
Secondary/high	0	0		0	0	
Elementary	1.2342	0.3832	3.4355	1.0589	0.3847	2.8832
Illiterate	1.9635	0.3822	7.1241	1.6436	0.3830	5.1739
Sex						
Boys	0	0		0	0	
Girls	1.2026	0.1575	3.3286	1.2476	0.1489	3.4820
Income						
High	0	0		0	0	
Middle-high	0.8981	0.4257	2.4551	0.8385	0.4192	2.3129
Low-middle	1.3184	0.4148	3.7374	1.2533	0.4091	3.5018
Low	1.7707	0.4183	5.8752	1.7367	0.4125	5.6787
Mid-day meal						
Availed				0	0	
Partial	1.8766	0.2327	6.5315	1.6872	0.2038	5.4043

Conclusions

This study has modelled data on children's age at the time of entry into school, sex, caste, father's education, family income and participation in the mid-day meal scheme in order to predict the outcome of dropping out of school. A set of logistic regression models was fitted for this purpose. The variables used, taken together, were significantly related to the outcome, with each one of the six variables contributing significantly to the prediction power of the full model. The relative importance of the predictors and the importance of their coefficients in predicting the odds ratios for different categories of each predictor were also evaluated.

We find that after controlling for other factors, caste status is not significantly associated with drop-out outcome. That is SC, Other Backward Classes and ST children do not show significantly higher chances of being in the drop-out group as compared to the General category children. We may be observing a narrowing of educational performance differentials among the children of these four social groups, as far as staying in school is concerned. Achievement differentials in the district is of course a different matter.

Higher age at entry into school is significantly associated with the chances of dropping out of school. The problem posed by a wide age band in a particular cohort may ease once issues related to the demand for education are sorted out in the district. In the meanwhile, teachers have to be better prepared to deal with multiple levels within a classroom. Teachers in the district are quite aware of multiple levels of ability within a classroom; they often use some criteria of 'intelligence' (*buddhi*, ranging from *hoshiar*

or clever to *mand* or dull) to segregate children into different ability groups. But often these groupings tend to be associated with age groups, higher ages falling into the 'dull' category. Ways of dealing with multiple levels in the classroom need to be built into in-service teaching training programmes.

Girls, as expected, are significantly disadvantaged in relation to boys as far as their progression through school is concerned. Reducing their chances of getting into the drop-out category implies an examination of the structuring of schools and schooling from the point of view of evolving gender-sensitive schooling practices. Another study carried out along with the present study indicates that how girls learn in their first two grades, and providing learning support to girls once they reach the upper primary level (grade five in Gujarat), are important areas for action in this regard (Vijaya Sherry Chand, Kalro and Shukla 1998).

An inter-generational continuity in matters of poor educational performance is indicated, with children whose fathers have less than secondary education being at risk. In addition, children in the lower three-quarters of the income distribution are disadvantaged. There are clear implications for providing homework support or alternative learning support mechanisms to children from such homes.

The significant influence of partial attendance in the mid-day meal scheme on the chances of dropping out may appear paradoxical at first glance. The finding only indicates the varied impact of the mid-day meal as an incentive for children to remain in school. Commonly heard remarks in the district are that the scheme is a 'backward class' scheme, and that the better-off do not participate in it. Perhaps the educational performance of those who do not need it is not influenced by the scheme. On the other hand, the scheme may have a positive impact on those who need it, and utilise it fully. Extending the latter argument, those who need it, but are unable to utilise it regularly for a variety of reasons, are at risk. This means that monitoring the attendance in the mid-day meal scheme, which is not a significant concern at the moment, is an important aspect of school management. Many teachers are known to be antagonistic to the scheme since they feel that the time they spend on supervising the cooking of the meals reduces the time available for strictly educational duties; they would prefer distribution of food grain. However, given that mid-day meals may have a significant impact on retention of children in school, provided children avail themselves of it regularly, teachers may need to be more serious about ensuring regularity in the utilisation of the scheme by those who need it.

To conclude, this study has estimated the importance of age at entry into school, caste, sex, father's education and income, and participation in the mid-day meal scheme, in helping children avoid the drop-out status. The relative chances of being in the drop-out group for the various categories of each factor have been presented in this article. An issue that needs further study is how well children of the Scheduled Castes, in relation to children of the General category, translate their avoidance of dropping out into completion of the schooling cycle. The relationship between attendance in school and educational performance among those who do not avail of the mid-day meal scheme in relation to those who utilise it satisfactorily (and thus by implication also attend school regularly) needs to be explored.

Notes

1. The data on which this article is based were collected in 1996. This first phase of data collection tracked the progress of the sampled children over the seven year period 1988-89 to 1994-95.
2. See Appendix for an exploration of the outliers in the solution.
3. The Wald test is used to evaluate the statistical significance of each of the coefficients. It is a function of the coefficient divided by its standard error, which in turn is a z statistic.
4. Separate models were fitted to examine whether there were any interactions between caste and sex, father's education or income. No significant interaction effects ($p < 0.05$) were found.
5. The models derived from the design of this study do not reliably distinguish between the in-school population and the passed population. A separate model was fitted for the outcome of being in school or passed. While the group of predictors as a whole was significantly related to the outcome, the ability of the model to discriminate between in-school and passed children was found to be very poor. Hence this analysis was dropped. However, since the coefficient for SC children in relation to General category children found in this model was significant with a positive sign, SC children may not be translating their avoidance of drop out status into school completion. This issue needs further study.

References

Government of Gujarat (1996): *State Plan for DPEP*, Gandhinagar: Government of Gujarat, Mimeo.

Government of India (1990): *Towards an Enlightened and Humane Society, NPE 1996: A Review*, New Delhi: Government of India.

Heneveld, W. and Craig, H. (1996): *Schools Count*, Washington, D. C.: The World Bank.

Lockheed, M. E., Verspoor, A. and others (1991): *Improving Primary Education in Developing Countries*, New York: Oxford University Press (for the World Bank).

Kmenta, J. (1986): *Elements of Econometrics*, New York: Macmillan Publishing Co.

Naik, J. P. (1969): *The Main Recommendations of the Education Commission*, New Delhi: National Council of Educational Research and Training.

Rummel, R. J. (1970): *Applied Factor Analysis*, Evanston: Northwestern University Press.

Tabachnick, B. G. and Fidell, L.S. (1996): *Using Multivariate Statistics*, 3rd ed. , New York: Harper Collins College Publishers.

Vijaya Sherry Chand, P.G., and Shukla, S.R. (1993): *Drop-out, Enrolment and Poverty: Policy Implications*, Ahmedabad: Ravi J. Matthai Centre for Educational Innovation, Indian Institute of Management, Mimeo.

Vijaya Sherry Chand, P.G., Kalro, A.H. and Shukla, S.R. (1998): *Progress through School: Repetition and Grade Completion*, Ahmedabad: Ravi J. Matthai Centre for Educational Innovation, Indian Institute of Management, Mimeo.

Visaria, Pravin, Gumber, A. and Visaria, Leela (1993): *Literacy and School Attendance in Gujarat: A Status Report*, Ahmedabad: Gujarat Institute of Development Research.

World Bank (1997): *Primary Education in India*, Washington, D. C.: The World Bank.

Appendix

Model A: Exploration of Outliers

A listing of the outliers with standardised residuals outside 2 standard deviations produced 31 cases. These are examples of the cases not predicted well by the model. Two cases were predicted to be drop-outs, but the majority (29 cases) were predicted to be in school or passed while they were observed to have dropped out. The two cases of in-school/pass wrongly predicted as drop outs were girls with illiterate fathers, belonging to the OBC, low family income and partial utilisation of the mid-day meal scheme. One of them was nine years old at entry. They can only be treated as exceptional cases. Out of the 29 other cases, seven were boys from two villages, one in Radhanpur and one in Santalpur. Six entered school at the age of six and four had fathers with elementary education. Six out of the seven were OBCs. Village factors may be responsible for their outlier status. Overall, there were 23 boys out of the 29 and 26 below the age of seven at entry. Most of the boys are clustered at the middle of the income distribution and have fathers with elementary education or who are illiterate.