The use of *post hoc* time-lag analyses to compare the effects of schooling on groups of pupils, controlling for age-related factors

Peter W. Hill and William A. McAtee
Research Branch, Education Department of W.A.

ABSTRACT

The use of sequential designs to investigate generational change and to isolate the three main components of development - age, cohort and time-of-measurement - have been advocated in recent years. While such designs may not be feasible or essential in many situations, it may nevertheless be possible to make use of data obtained in longitudinal studies to conduct retrospective time-lag comparisons through the *post hoc* blocking of subjects. Used in this way, the time-lag method offers an interesting alternative to traditional methods of investigating differential change, which typically yield results that confound amount of schooling with other age-related factors. This paper reports on the use of *post hoc* time-lag comparisons in analysing data obtained in the Australian Open Area Schools Project.

Though it is a perfectly natural question to inquire whether gains made by two or more groups over the same time are significantly different, it is an extremely difficult question to answer properly. As Linn and Slinde (1977, p.174) note, "Problems in measuring change abound and the virtues in doing so are hard to find". On the other hand, the ever-pressing need to obtain information concerning the relative effectiveness of educational programmes means that researchers are constantly tempted by the notion of gain scores, in spite of their inherent difficulties.

The Australian Open Area Schools Project

These temptations were met and not entirely resisted in the Australian Open Area Schools Project, a national study which investigated relationships between school design and various teaching practices and pupil outcomes in 45 conventional design and 70 open area primary schools throughout Australia.

In this study a two-stage sampling design was employed and data were collected from 115 schools and from random samples of 18 Year 5 pupils from within each of the schools sampled. In analysing the results, special problems were posed by the need to compare pre-formed groups of pupils on the basis of measures obtained several years after the 'treatment effect' of school design had been in operation. The data available for the analyses of the pupil outcomes were:

1. two measures for each pupil on 5 outcome variables (Mathematics, Reading, Written Expression, Attitude to School and Self Concept), obtained through a test-retest in April and November of 1974, using the same instruments on both occasions, and

2. measures on certain uncontrolled, 'extraneous' variables thought likely to influence scores on the outcome variables.

By testing pupils on two occasions, more stable estimates of the cumulative effects of schooling in the different categories of school design were obtained. In addition, the April-November test-retest provided an opportunity to estimate and compare short-term, within-grade effects of schooling over the seven-month interval separating testing occasions, holding constant sources of variation associated with age.

In testing for differences in cumulative effects of schooling between the various design categories, the two measures for each outcome variable were regarded as intercorrelated responses from the same subjects and, after aggregation at the school level, were analysed conjointly according to a bivariate MANOVA model, using covariance adjustments to control statistically for age, general intellectual ability and social status.

In testing whether differential change between the two testing occasions had occurred, independently of age-related factors, *post hoc* time-lag comparisons were employed.

The authors are indebted to Larry Goulet who first aroused their interest in time-lag comparisons during his stay as Visiting Professor with the School of Education, Murdoch University, and to Professor Barry McGaw who inspired and assisted in the application of the technique as reported in this paper.

The time-lag method

As Goulet (1975) has noted, studies which incorporate a longitudinal dimension in order to assess the amount of change that has occurred over a given time interval, invariably confound amount of schooling with other age-related factors. In an attempt to overcome this problem, Goulet has advocated the use of sampling designs which take into account various components of change so that differences associated with amount of schooling can be separated from differences associated with chronological age through the use of time-lag comparisons.

The time-lag method involves collecting data to enable three critical components of change - age, cohort and time of measurement - to be considered simultaneously. The data are organised according to the general model for assessing time-related behaviour change of Schaie (1965), as shown in Figure 1. The comparisons made are indicated by arrows. By comparing pupils of different birth dates at the same chronological age, it is possible to estimate the effects of school experience over a specified period of time independently of the other age-related factors.

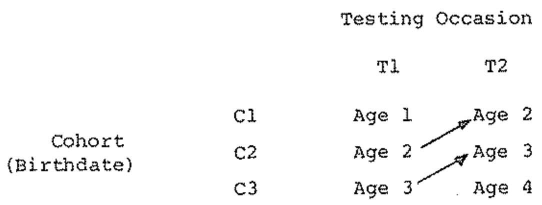


Figure 1: General model for assessing time-related behaviour change (Schaie, 1965)

In the single group situation, a two-factor analysis of variance design may be employed to test the nullity of effects associated with amount of schooling, with age, and with their interaction. Subject's scores are assigned to cells as shown in Figure 2, which illustrates the simplest case of a time-lag design.

Age (at testing date)	April	Nov.
9-9	A	B
10-4	C	D

Testing Occasion

Figure 2: Simple single-group time-lag design

In addition to enabling tests of main effects and their interaction, the above design may be used to obtain an estimate of the influence of schooling over the time interval separating testing occasions, holding constant effects associated with age. This is given by summing the differences between age-matched sample means and dividing by the number of comparisons. For the situation depicted in Figure 2, this would be

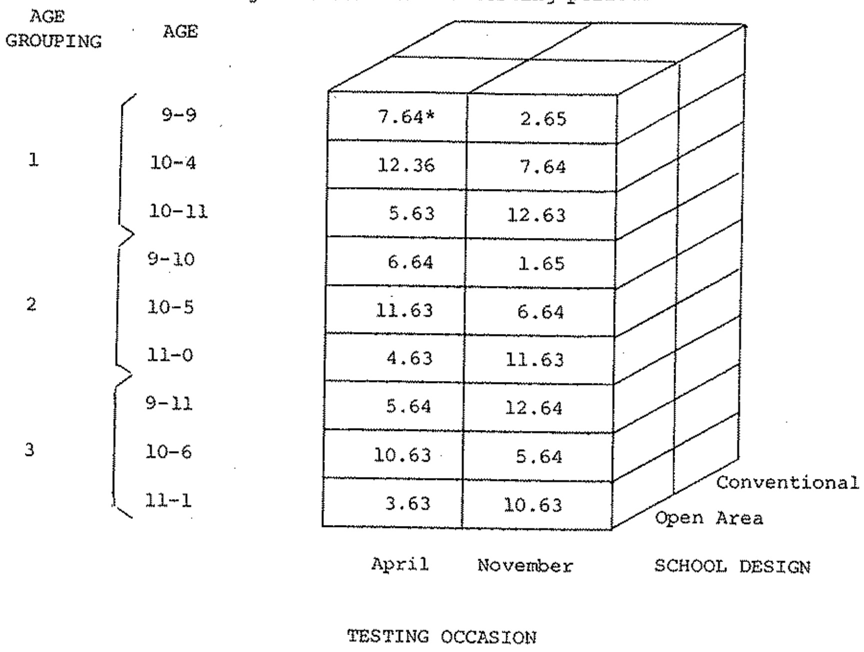
$$\frac{(\bar{X}_B - \bar{X}_A) + (\bar{X}_D - \bar{X}_C)}{2}$$

When subjects have been sampled from or assigned to distinct populations in addition to those represented by the crossing of testing occasion with age, the analysis model is a 3-way design (testing occasion x age x group).

Post hoc time-lag comparisons

The adoption of the sampling strategy depicted in Figure 1 would seem to offer advantages not shared by other designs commonly encountered in educational research. On the other hand, there are clearly many situations in which sampling on age is neither feasible nor essential. In such cases, however, it may nevertheless be possible to make use of longitudinal data to conduct time-lag comparisons through the *post hoc* blocking of subjects.

This method was applied to data obtained in the Australian Open Area Schools Project. Pupils' scores were arranged in blocks as shown in Figure 3 to create a series of 36-cell, 4-way designs (age grouping x age x testing occasion x school design). By introducing the 'age grouping' factor, the number of cases which could be incorporated in the analyses was maximised. The tesing occasion x school design interaction was then used to test the hypothesis of no differential change between the two testing periods.



* Month and year of birth (cohort)

Figure 3: Arrangement of subjects in block for time-lag analyses

The analyses were carried out using the Multivariance computer routines of Finn (1976), which employ a general non-orthogonal solution in which the ANOVA model is reparameterised to a basis with orthogonal columns, thus permitting independent tests of effects.

A full summary of the analysis of variance for pupil outcomes on the two administrations of the mathematics tests if given in Table 1. It can be seen that the 4-way and all of the 3-way interactions are non-significant. In addition H_0 is accepted for the testing occasion x school design interaction used to test the hypothesis of no differential change between testing occasions, holding constant other effects associated with age.

Table 1: Analysis of variance for time-lag comparisons of mathematics scores of pupils in conventional and open area design schools

Source	d.f.	M.S.	F
Group	2	93.1713	3.6739*
Age	2	131.0843	5.1689*
Occasion	1	1087.5514	42.8842*
School Design	1	1474.598	58.1462*
Group x Age	4	4.9096	0.1936
Group x Occasion	2	2.3443	0.0924
Group x School Design	2	82.6878	3.2605*
Age x Occasion	2	174.9088	6.8970*
Age x School Design	2	8.1377	0.3209
Occasion x School Design	1	1.3021	0.0513
Group x Age x Occasion	4	1.0208	0.0403
Group x Age x School Design	4	23.8928	0.9421
Group x Occasion x School Design	2	51.0836	2.0143
Age x Occasion x School Design	2	14.5685	0.5745
Group x Age x Occasion x School Design	2	23.1661	0.9135
Residual	1236	25.3602	

* $p<.05$

The estimated combined means for the testing occasion x school design interaction are given in Table 2. When these are graphed, as in Figure 4, it is evident that pupils in both the conventional and open area schools gained in approximately the same way. This indicates that over the seven month interval separating testing occasions it was not possible to detect any differential effects of schooling among pupils in conventional and open area schools, while controlling for age-related factors. This finding was repeated in tests involving the other pupil outcome measures (Table 3).

Table 2: Estimated combined means for time-lag comparisons of mathematics scores of pupils in conventional and open area design schools

School Design	Testing Occasion	
	April	Nov.
Conv.	17.56	20.08
Open	15.62	17.83

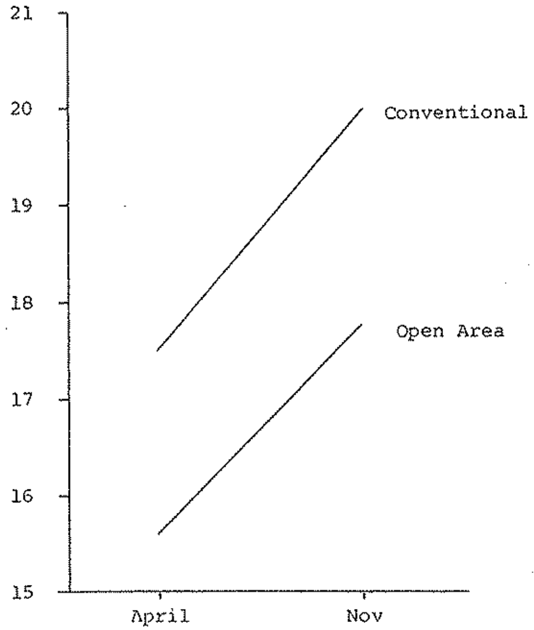


Figure 4: Time-lag comparisons of mathematics scores of pupils in conventional and open area design schools

Table 3: Analyses of variance of pupil outcomes to investigate the testing occasion x school design interation, using time-lag comparisons

Source	Variable	d.f.	Mean Squares	F
Testing occasion x School design	Mathematics	1	1.30	0.05
	Reading	1	1.66	0.02
	Written Expression 1	1	1.19	0.39
	Written Expression 2	1	0.00	0.00
	Attitude to School	1	0.54	0.59
	Self Concept	1	13.61	1.05
Residual	Mathematics	1236	25.36	
	Reading	1248	70.30	
	Written Expression 1	1041	0.48	
	Written Expression 2	1138	0.46	
	Attitude to School	1183	0.91	
	Self Concept	1221	13.00	

* p<.05

Conclusion

At a time when it is fashionable to ask whether schools, teachers and educational programmes really do make a difference, the time-lag method appears to have certain advantages in many situations confronted by educational researchers. In particular, it leads to more direct estimates of the effects of schooling than do designs traditionally employed to explicate both short- and long-term influences of educational experiences.

In the case of the adaptation of the method reported above, any advantages accrued by the use of time-lag comparisions were offset by the fact that the technique was used in a situation in which random assignment of pupils to schools of different design types was not possible. For this reason, the particular application of the time-lag method described here was exploratory rather than exemplary. Hopefully, as the method becomes more widely known, researchers will be more inclined to adopt designs which allow more valid time-lag comparisons of groups of subjects.

References

Finn, J.D., MULTIVARIANCE: Univariate and multivariate analysis of variance, covariance and regression. National Educational Resources, Inc.: Chicago, Illinois, Version V, Release 3, June 1976.

Goulet, L.R. Longitudinal and time-lag designs in educational research: an alternate sampling model, Review of Educational Research, 45(4), Fall 1975, 505-523.

Linn, R.L. and Slinde, J.A., The determination of the significance of change between pre- and post-testing periods, Review of Educational Research, 47(1), Winter 1977, 121-150.

Schaie, K.W., A general model for the study of developmental problems, Psychological Bulletin, 64, 1965, 92-107.

SIGNIFICANCE LEVELS FOR TEST RELIABILITY (KR20)

Robert M.H. Hind
Gippsland Institute of Advanced Education
Switchback Road
Churchill, Victoria 3842

(Paper presented at the Annual Conference of the Australian Association for Research in Education, Perth, November, 1978.)

A B S T R A C T

Significance levels for test reliability, i.e. the values for the reliability above which one can say that the reliability is significantly greater than zero, are not available.

This paper reports the result of an investigation of significance levels for test reliability, calculated by the KR20 formula, in which the number of test items, the number of examinees and the mean score on the test were varied.

The investigation showed that only the number of examinees has any significant effect on the sampling distribution of KR20. Significance levels for KR20 are given.

Payne and Anderson (1968) reported that they had found the significance levels for KR20 using an automated sampling experiment approach. They investigated the sampling distributions of KR20 as a function of three variables: number of items in the test, number of examinees and distribution of true scores.