Ceylon Journal of Science : Physical Sciences, 9(1), 102-109 (2002)

CLASSIFICATION OF STATE SCHOOLS IN SRI LANKA: A MULTIVARIATE APPROACH

S.S.K.B.M.DORABAWILA[a*], S. SAMITA[b] AND R.O. THATTIL[c]

[a]*Department of Economics, Faculty of Arts, University of Peradeniya.**
[b] *Department of Crop Science, Faculty of Agriculture, University of Peradeniya*
[c] *Post Graduate Institute of Agriculture, University of Peradeniya*

ABSTRACT

It is imperative that the key-personnel directly engaged in the transformation process of the Sri Lankan education system, especially in respect of the state schools are possessed with a comprehensive knowledge of the compositional aspects of all schools in the country. This analysis is carried out to evaluate the present status of schools and introduce a new classification for schools. Three multivariate statistical techniques, cluster analysis, multivariate analysis of variance (MANOVA) and discriminant analysis are used in the analysis. The identified six broad groups were such that they were placed in a ranking order, showing a gradual decline in standards. In addition each group was distinguishable by its markedly different levels of available facilities and performance. Furthermore, six discriminant functions were formulated to assign new schools into groups. The techniques applied and the outcome generated used in this study can be extended to all 10,358 schools in Sri Lanka and thus to evaluate their status.

1. INTRODUCTION

Sri Lankan education system is going through a phase of drastic transformation as a result of educational reforms introduced by the government with the assistance of the World Bank. The on going changes are addressed to several spheres in the system, namely, administration, financing, recruitment, teaching methods, curriculum development, and system updating. These transformations that target the vital components of the system require that they be based on proper planning, and, then zealously implemented for the betterment of the ultimate receivers – children, the future generation. In this context it is imperative that both the policy makers and the key personnel responsible for are possessed with a comprehensive knowledge of all dimensions of the school system in Sri Lanka.

It is generally observed that the present day decision-making with respect to allocation of resources or provision of facilities is mainly made in considering one single aspect of the school, namely, to which stream the school belongs to. These are primarily based on whether classes are conducted at the advanced level in Science, Commerce, Arts, or a combination of them. Thus, this type of decision making can lead to ill-distribution of resources, neglect of schools in the "lower" hierarchy, and thus further widening the disparities that exist by urban/rural location or by district.

* dorab@slt.com

Table 1.1 Description of the Types of Schools

Type of School	Description of Schools
1. National schools	Type 1AB and/or Type 1C schools
2. Type 1AB	Science A/L only, or with non-science A/L
3. Type 1C	Non-science A\L only
4. Type 2	Year 1 to 11
5. Type 3	Year 1 to 8 or Year 1 to 5

Table 1.1 (Ministry of Education and Higher Education. 1998) provides a description of the types of schools identified by the Ministry. There, some of the schools found in types 1 (National Schools) to 3 contain schools with classes starting from grade 1 or grade 10 only.

The School Census data reveal that 5% of the schools fall into Type 1AB, which contains, among others, Advance Level Science stream. It is generally considered to be the best-equipped type of schools. 76% of the schools fall into Type 2 and Type 3, which are the small schools with classes up to Grade 5, 8, or 11; the rest, 9% fall into type 1C -- schools with classes up to advanced level, but no science stream.

Ministry of Education has designated some schools as National Schools, but without a rational basis. These schools comprise of schools with both the science stream and the non-science stream, or one of them. National schools comprise of Type 1AB as well as 1C. Majority of them are Type 1AB. And also, there are two separate types under the headings of Type 1AB and 1C, in addition to the national schools. Evidently, a clear criterion is lacking. That is, one cannot understand the difference between National Schools and Type 1AB and 1C.

Table 1.2 Distribution of Schools in Galle District

Type of School	**No. of Schools**
National schools	21
Type 1AB	14
Type 1C	113
Type 2	169
Type 3	175
Total	492

Table 1.2 (Ministry of Education and Higher Education. 1998) provides details of types of schools in Galle district. Percentage distribution of schools by type in Galle district shows those 2.84% and 22.97% are Type 1AB and 1C respectively; and, 4.27% are National Schools.

This study attempts to introduce a system that provides a basis to ascertain a number of distinct groups or clusters (of schools) with similar characteristics. Once such clusters are identified, it provides guidelines to assign a given school to a particular group. Undoubtedly, this would facilitate the process of allocating resources to the school according to the requirements. On the other hand, when a policy decision is to be made with respect to upgrading of a given school, it can be rationally arrive at, thus not being dictated by political considerations.

The broad objective of this study is to obtain an objective oriented classification for Sri Lankan government schools based on empirical evidence largely related to the existing resources and current level of performance. Initially this study attempted to partition schools in a selected district (Galle) into relatively homogeneous groups based on similar characteristics. Secondly, using the formed groups

functions are formulated to decide to which group each school used for this study as well as any new one fell to. Finally, using statistical techniques the most influential variables for this study are identified.

2. MATERIALS AND METHODS

The data used in this study were obtained mainly from a survey of schools, under the title 'A Baseline Survey of Schools 1999', carried out by the Finance Commission. The questionnaire in the main survey was structured to elicit information required for the purpose of developing a programme to ensure that all schools receive the necessary funds, support and other forms of assistance in consistent with their needs. Therefore, it contained the information on various essential dimensions (characteristics) of schools. However, it was deficient on the aspects that capture the level of productivity (student performance) of a school. To fulfill this shortcoming, the relevant data used in this study were gathered from the 1998 Statistics on schools performances at examinations of the Examinations Department.

Population of study covers all the state schools located within the Galle District, numbering 492. A sample of 252 schools, approximately 50 percent of the population, is randomly selected.

2.1 Variables

The questionnaire of the Base-line Survey of Schools, 1999, contained 38 questions under 10 sections addressed to different characteristics of schools. Of them, 16 variables were identified as influential and decisive for this analysis and another seven variables were obtained from the statistics of the Examinations Department records.

They are as follows:

1. Location of the school (residence sector)- V2
2. Gender of pupils - V3
3. Classification by convenience - V4
4. General infrastructure of the school(land) - V5
5. Availability of hostel facilities - V6
6. Availability of water - V7
7. Availability of electricity - V8
8. Availability of telephones - V9
9. Total floor space - V10
10. Student enrolment - V11
11. Demand for enrolment - V12
12. Dropouts from enrolment - V13
13. Particulars of Principals and Teachers(status by designation of staff) - V14
14. Educational and professional attainment of the staff - V15
15. School furniture and other supplies - V16
16. Stream-wise classification - V17
17. Total number of student sittings at the Ordinary Level Examination. - V18
18. Percentage of students passing the Ordinary Level Examination. - V19
19. Index obtained for passes in English at the Ordinary Level Examination. - V20
20. Total number of student sittings for non-science Advance Level Examination. - V21
21. Total number of student sittings for science Advance Level Examination. - V22
22. Index obtained for passes for non-science Advance Level Examination. - V23
23. Index obtained for passes for science Advance Level Examination. - V24

2.2 Analysis

This study attempts to elicit information on a variety of characteristics and functions of schools. It is accepted that most appropriate method of carrying out analysis with many variables is using multivariate techniques. According to the objectives, multivariate techniques are applied to analyse the data in this study.

2.3 Determination of Groups

Cluster analysis is used to combine schools with similar characteristics into groups since the possible membership is unknown in advance, i.e., prior to the analysis. In this analysis, agglomerative, centroid hierarchical clustering technique is used to combine the two closest clusters initially, and then, successively, until all observations are grouped into one (Thattil,1993). Generally, with more than 2 variables, a device called dendrogram is utilized to

summarize the clustering at successive steps. But, in this analysis, since 23 variables are incorporated with 252 observations, depicting the clusters with a dendrogram was found to be a difficult and an obscure task. Thus, as an alternative, the clusters are listed by following the sequence in which each cluster was formed at successive steps. Finally, based on this sequence, the most appropriate number of clusters (groups) is identified. It should be emphasized that the choice of the necessary number of groups is a subjective decision.

2.4 Testing the Difference between Groups

Initially, Multivariate Analysis of Variance (MANOVA) is used to verify whether the centroids (vectors of means of the variables) of various groups differ significantly (Tacq, 1997). The objective of using MANOVA in this study was to validate the number of clusters identified by the cluster analysis. In addition the same technique is utilized to determine the important variables in distinguishing the clusters. The statistic Wilks's lambda was used for this purpose.

2.5 Derivation of Functions for Groups

Discriminant Analysis is used to derive classification functions for clusters identified. The use of these functions is that when an unidentified observation is given the appropriate group can be found by substituting relevant variable values to these functions (Afifi & Clark,1996). The discriminant or classification function is given by the form:

$$t = k X$$

where

$t = t_1, t_2, \ldots t_g$ (Discriminant score Vector)

k = column vector for discriminant function coefficient values

X = matrix of scores of the discriminating variables

The classification is done by evaluating each classification function for each school. The unknown school is classified into groups of which classification function is maximum.

3. RESULTS AND DISCUSSION

3.1 Identification of Groups

The number of groups (types of schools) should be decided based on several considerations, such as purpose of having classifications, the variability among groups and validity convenience in analysis. A tentative, subjective decision on the possible number of groups is made based on the sequence of clustering of schools. Accordingly, six groups were identified from the analysis.

A visual examination of the groups thus identified show that each group distinguished from one another, being markedly different on facilities available, and student performance. These groups are given in Table 1.3 Group 1 consisted of seven schools accounting for 2.77% of all schools in the sample. This group is the highest ranked and corresponds, at least partly, with the National School category in the existing system. These seven schools are the most distinctive among the schools with respect to every aspect or variables taken for consideration. Of the 20 National schools in the sample, only seven fell into the highest ranked group.

Table 1.3 The Distribution of Schools Derived by the Cluster Analysis Technique.

Group	No. of schools	Percentage of the sample
1	7	2.77
2	20	7.94
3	50	19.84
4	25	9.92
5	52	26.63
6	98	38.88
All	252	100

Table 1.3 gives the distribution of schools by the six groups generated by the cluster analysis. Ranking of schools in the Table, from 1 to 6, showed a gradual decrease in the standards of the schools, successively. However, the ordering of schools (shown in Table 1.3) is largely influenced by the availability of facilities and the size of operation, both curricular and non-curricular. Evidently, schools in Group 6, which has the highest number of schools (98), are the ones least provided with facilities and also lowest in

scales of operations. It is of interest to note that the number of schools in Type 3(the lowest rank according to the Ministry's existing classification system) was only 26.

3.2 Validation of Existence of Identified Groups

The presence of considerable difference among the identified groups or the centroids of the groups was seen from the low P value (0.0001) associated with the Wilks' Lambda statistic. This is further authenticated by the pair-wise generalized squared distances between groups shown in Figure 1.1. Furthermore, the analysis of variance (ANOVA) showed a probability value less than 0.01 for all 23 variables. This confirms variation among the identified groups.

Discriminant Analysis

Pairwise Generalized Squared Distances Between Groups

Generalized Squared Distance to V1(Identified Groups)

From Groups ...	1	2	3	4	5	6
1	0	270.50064	420.59786	516.15358	499.12274	... 562.22710
2	270.50064	0	47.72834	115.84969	109.51475	152.13616
3	420.59786	47.72834	0	59.22899	..58.51989	94.90231
4	516.15358	115.84969	59.22899	0	...1.33494	26.90583
5	499.12274	109.51475	58.51989	1.33494	0	21.07043
6	562.22710	152.13616	94.90231	26.90583 ...	21.07043	 0

Fig. 1.1 Pair – wise Generalized Squared Distances between Groups

3.3 Classification of Schools

Discriminant analysis is performed using PROC DISCRIM of SAS for the identified 6 groups and following linear discriminant functions (classification functions) is obtained. (Table 1.4). Above discriminant functions can be used to identify the group of an unclassified school. The procedure is as follows. Each discriminant function is evaluated by substituting each variable value for an observation (school). Then the school is classified to the group for which score of discriminant function was maximum.

3.4 Validation of Classification

Error count estimates given by the discriminant analysis for each group is as shown in Figure 1.2. According to the error count estimates, Group 1 and Group 2 did not provide any misclassifications. For Group 3 and Group 6, the misclassifications of observations were low with 2% and 1% respectively. But, for Group 4 and 5, misclassifications of observations were considerably high with 28% and 48% respectively. Overall, this study gave a considerably low total misclassification of 13 %. This observed misclassification of observations in certain groups can be attributed to inconspicuous values seen in certain observations in several variables. These obscure values could have been resulted from shortcomings in reporting the data.

Error Count Estimates for Groups

	1	2	3	4	5	6	Total
Rate	0.0000	0.0000	0.0200	0.2800	0.4808	0.0102	0.1318
Priors	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	

Fig. 1.2 Error Count Estimates for Groups

3.5 Identifying Important Variables for Classification

The Wilks' Lambda Statistic in MANOVA is employed to investigate the importance of variables in classification. The resulting Wilks' Lambda values for MANOVA in each case (i.e. omitting the variable under consideration) are given in Table 1.4.

Table 1.4 Linear Discriminant Functions' Coefficient Values for each Group

Variable	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Constant	**-355.93875**	**-154.1249**	**-130.05671**	**-100.53475**	**-97.54287**	**-89.35606**
V2	19.12599	14.2543	11.11259	11.59640	12.50731	11.77123
V3	2.01209	1.9715	1.77165	1.44025	1.56212	2.81036
V4	-12.27351	10.7156	16.51684	19.25176	15.65221	13.74898
V5	7.03712	4.9756	3.40520	3.38546	3.84764	3.53174
V6	74.38228	66.6702	60.70990	62.81858	62.17020	68.48260
V7	-1.13921	-1.1866	-1.03133	-0.76519	-0.85310	-0.72430
V8	27.69783	8.2570	6.32409	8.07990	7.56711	7.72022
V9	3.42990	1.5757	0.80251	0.12562	0.43736	0.66916
V10	-8.66647	-4.6439	0.05609	-0.70862	-1.57050	-3.01263
V11	13.60665	7.8091	7.15596	5.69771	5.42210	4.98902
V12	0.10728	0.1447	0.15168	0.12224	0.13514	0.26025
V13	0.09117	0.0970	0.09165	-0.03378	-0.00896	0.03677
V14	-3.19290	-24.2261	-21.60119	-25.79580	-23.51366	-21.10919
V15	30.31940	14.5349	28.53313	32.83913	34.39234	29.44853
V16	4.29008	6.2601	6.18424	6.30288	6.28682	6.28883
V17	0.11395	0.0152	-0.04260	0.04153	0.02681	0.07693
V18	0.64842	0.7475	0.85559	-0.35804	-0.35035	-0.34846
V19	0.53931	-0.0613	-0.05081	-0.03123	-0.02920	-0.02184
V20	-0.9425	-0.0568	-0.23087	0.09438	0.09436	0.21847
V21	0.56644	0.2823	0.03580	-0.04380	-0.03055	-0.12932
V22	0.62293	0.4340	0.37677	0.27008	0.28057	0.15089
V23	0.36521	0.5006	0.63514	0.50551	0.40407	-0.05807
V24	-0.83124	-6.52459	-9.11315	-8.42729	-7.21055	-6.21914

Table 1.5 Wilks' Lambda Values to Test the Importance of Each Variable

Variable/s	Wilks'Lambda Value(Λ)	Pr>F
All 23 variables	0.29087336	0.0001
V2(without)	0.28579070	0.0001
V3(without)	0.18790505	0.0001
V4(without)	0.24107217	0.0001
V5(without)	0.10349375	0.0001
V6(without)	0.24297245	0.0001
V7(without)	0.24905842	0.0001
V8(without)	0.11976547	0.0001
V9(without)	0.06778920	0.0001
V10(without)	0.07817711	0.0001
V11(without)	0.10092575	0.0001
V12(without)	0.09333406	0.0001
V13(without)	0.09002836	0.0001
V14(without)	0.04417252	0.0001
V15(without)	0.02938166	0.0001
V16(without)	0.02150461	0.0001
V17(without)	0.00434070	0.0001
V18(without)	0.01235890	0.0001
V19(without)	0.00307217	0.0001
V20(without)	0.00156712	0.0001
V21(without)	0.00141656	0.0001
V22(without)	0.00050693	0.0001
V23(without)	0.00085813	0.0001
V24(without)	0.00046682	0.0001

In each case, when the variable under consideration is omitted, value of the Wilks' Lambda statistic decreased from the overall value.(Wilks' Lambda value with all 23 variables). This indicated that all the variables were important for this study. But, the decreases in the thirteen variables from V9 to V24 were considerably higher than the other variables. These changes show that the thirteen variables were relatively more important than the others.

4. CONCLUSION

Six groups of schools identified, here, are illustrative models of possible levels of standards of government schools in Sri Lanka. Having such a systematic groupings of schools in the education system would motivate the schools authorities to achieve overall targets typically expected of such levels rather than existing in buoyancy. Furthermore, it would inspire the schools not only to attain the overall targets, but also to improve on the level of their available facilities and performances, and thus to raise the quality.

One of the most benefited groups of persons from this methodical classification of schools would be the policy-maker. This is because it enhances the proper and rational decision making on the part of the policy-maker. The next most benefited is the services provider or the implementer. The availability of groups in sequential order of standards allows the responsible personnel to allocate the level of requirement for a given school. And this also avoids the necessity to take ad-hoc decisions, and, at the same time provides a basis to restrict the possibility of arbitrary political interference.

When carrying out this study, due to certain limitations, the variables used for classification were restricted. If this study is carried out with a comprehensive data set that had a coverage of all government schools in Sri Lanka, this classification could be more complete and accurate. Therefore, it is suggested that a study, which takes a larger sample, having wider coverage of districts and more variables that further describe the aspects that is not included in this study would facilitate the formation of groups more appropriate for the Sri Lankan education system. And the derived functions for the groups would allow the classification of the total 10,358 government schools into distinctive groups. In the absence of such groups, assigning of schools to appropriate groups would be, perennially, an ad-hoc task of the personnel of the Ministry of Education. Furthermore, the presence of such a classification would enable the authorities to annually evaluate the standard of a given school and assign the appropriate status. It tantamount to ranking of schools based on a continuous process of performance evaluation.

5. ACKNOWLEDGEMENT

The authors are grateful to Mr. Nishantha Palihawadena and officials of the Finance Commission, and Mrs. G. Batagoda and officials of the Department of Examinations for providing

assistance in collecting the data-set, without which this would not have materialized.

REFERENCES

1. A. A. Afifi,. and Clark, V., *Computer – Aided Multivariate Analysis.* (1996).
2. Chapman & Hall, London.
3. Everitt, B. S.,. *Cluster Analysis.* J. W. Arrowsmith Ltd.,Bristol. (1993)
4. Johnson, R. A. and Wichern, D. W., *Applied Multivariate Statistical Analysis.* Prentice –Hall, Inc., New Jersey. (1988).
5. Ministry of Education and Higher Education. (1998). *School Census* Colombo (1997).
6. Tacq, J., *Multivariate Analysis Techniques in Social Research.* SAGE Publications Ltd., London. (1997).
7. Thattil, R. O., *Multivariate Application in Agriculture.* Faculty of Agriculture, University of Peradeniya (1993).