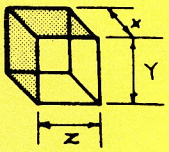
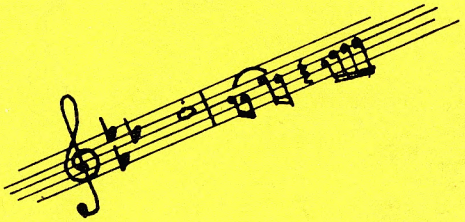
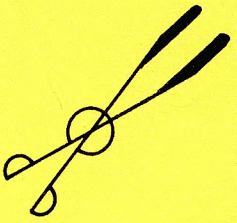
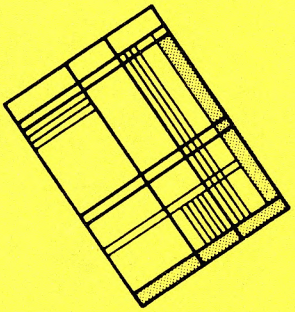
RELIABILITY OF ENROLMENT DATA BANK-FEASIBILITY STUDY IN KINGSTON EDUCATION DISTRICT 1969-70 TO 1973-74

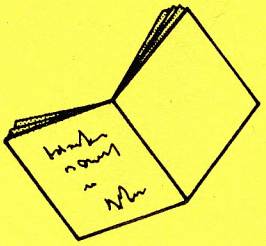


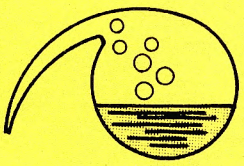






12 x 2 - 7 = 17





DEPARTMENTAL STATISTICS DIVISION
DEPARTMENT OF INDIAN AFFAIRS AND NORTHERN DEVELOPMENT

DIVISION DE LA STATISTIQUE DU MINISTERE
MINISTERE DES AFFAIRES INDIENNES ET DU NORD CANADIEN

RELIABILITY OF ENROLMENT DATA BANK-FEASIBILITY STUDY IN KINGSTON EDUCATION DISTRICT

1969-70 TO 1973-74

D.G. SAIGAONKAR

SENIOR STATISTICIAN

Departmental Statistics Division

Department of Indian Affairs & Northern Development

October 1975

ACKNOWLEDGEMENTS

The author wishes to gratefully acknowledge the constant co-operation and guidance received during the preparation of this report from Mr. D.K.F. Wattie of the Education Branch, and from Mr. W. Zayachkowski and Dr. S. Kumar of the Departmental Statistics Division.

He wishes to express his sincere thanks to the Support Staff for carrying out the collection and compilation of the data. Thanks are also due to Mr. T. Vella for the cover design and to Mrs. P. Murphy for the diligent typing.

September 18, 1975.

Datta Saigaonkar,
Senior Statistician.

INDEX

		Page
	Acknowledgements	i
1.	Introduction	1
2.	Sample and Sources of Information	2
3.	Items of Information and Their Importance	4
4.	Date and Duration	5
5.	Evaluation	6
	(a) Preliminary Analysis	6
	(b) Further Analysis	7
	(i) Record Reliability Index	8
	(ii) Aggregate Reliability Index	10
	(iii) Item Reliability Index	10
6.	Manpower & Cost	11
7.	Assessment of Objectives	11
8.	Responsibility	12
9.	Some Findings	12
10.	Recommendations	13

TABLES & CHARTS

1.	Percentage Distribution of Record Satisfying Criteria	15
2.	Distribution of Student Records Satisfying Criteria Groups	17
3.	Frequency Distribution of Student Records by Their Reliability Indices: 1969-1974	19
4.	Contributions of the Groups to the Aggregate Reliability Indices	21
5.	Percentage of Correct Records for Items Studied for Reliability 1969-70 to 1973-74	23

1. Introduction

The Eudcation Branch of the Indian and Eskimo Affairs Program has been collecting annually the educational statistics relating to the registered Indian Students. A survey known as the Student Registration Survey has been conducted annually to collect such data for registered Indian students enrolled in elementary and secondary schools since the school year 1969-1970. The items of information collected include student's name, band number, date of birth, grade, parental residence status, and name and type of school. The data about the individual students are updated annually by the school or the district officials and are stored in the Education Data Bank maintained by the Computer Information Systems Division of the Department. The data bank contains information on approximately 70,000 students. Various quality control checks, including the monitoring and editing of the errors, were carried out during the development and maintenance of the data bank.

Generally, data banks of this size are susceptible to errors due to the large volume of data and the diverse background of the individuals associated with their development and maintenance. In 1974, the Educational Management Information System working group recommended a reliability study of the data bank covering the period from the school year 1969-1970 to 1973-1974 inclusive, and the recommendation was accepted by the Education Branch.

Prior to undertaking this reliability study of the entire data, it was considered necessary to conduct a feasibility study on that part of data bank relating to the students enrolled in Kingston Education District. The

...2

objectives of this feasibility study were:

(a) To develop a suitable methodology to detect and estimate the extent of discrepancies between the data bank and the field records;

(b) To identify the field problems and to develop corrective measures;

(c) To have a choice of reference birth years; and

(d) To develop a method of correcting existing errors in the data bank. The Kingston District was chosen for the study because it appeared to have field problems likely to be encountered in other districts.

This report deals with the sample selection, methodology, and analysis of the feasibility study results.

2. Sample and Sources of Information

The choice of the sample and the retrieval of the information for the study were completed in three stages.

As the first stage, a list of all individuals was prepared, using Band Membership lists as of June 30, 1974 with birth dates between 1958 and 1963 inclusive. These individuals were at least six years old during the school year 1969-1970 and at most sixteen years old during the school year 1973-1974.

Ideally speaking, this group will contain children attending schools during the life of the data bank under study. There were 998 persons on this list and they constituted the population for the selection of the sample. A sample of 168 persons, using a systematic sampling, was

chosen from this population and is referred to as the "Initial Sample" in the report. The band number and date of birth for each individual in the initial sample were retrieved from the Band Membership list.

At the second stage, the individuals in the initial sample having records in the data bank were identified for each of the years under study. This group is referred to as the "Data Bank Sample" in the report. Their number ranged between 95 and 102 and the individual data were retrieved from the data bank for each year of the study. The lack of computer records for about 70 sampled individuals with off-reserve status included in the initial sample can be attributed partly to the lack of Departmental responsibility for these students.

At the third stage, information about the same items as on the data bank records was collected independently from the district and school records for each of the individuals in the data bank sample for each of the school years.

It is reasonable to assume that the field records collected from the district and school offices are correct. Discrepancies in the computer records were analysed for different items of information, separately for each year.

Possible areas of bias

It is important to examine the introduction of potential bias inherent in these data sources in any analysis undertaken.

The Band Membership list is a list of all registered Indians, updated in June and December of every year. These updates include additions and deletions of members for natural and legal reasons, and changes in the band numbers. The data bank contains the names of all students who

...4

are a Departmental responsibility but some of whom are not on the Band Membership list. This could create some bias in the findings. In a number of cases, the information on certain items was not available in the school or district records. It was provided by the school principal, or the education or social counsellor on the basis of memory. Use of memory instead of official records could be another source of bias. Memory played an important role in the information for the school year 1969-1970. Generally, memory's role decreased with each succeeding year.

Records of two students who were in the data bank sample could not be obtained in the field. This could possibly be due to the use of legal names in the Band Membership list and in the Data Bank but the use of nicknames in the school records.

3. Items of Information and Their Importance

The statistical information generated by the data bank is used for various purposes. These range from budgetary planning and forecasts to projections of enrolments. For the purposes of this study, each of the eighteen items of information was assigned to one of three groups: Essential, Acceptable and Negligible. The guiding factor in this assignment was the importance of the item of information in relation to the objectives of the Student Registration Survey and the use to which the resulting statistical information is put. For example, the parent's residence status, namely on reserve or off reserve, is an important item of information for the Department in discharging its financial responsibility effectively. This was assigned to the Essential Group. Items of lesser importance are day and month of birth and type of course in which student is enrolled. These items play a significant role in

studies relating to progress and to the quality of education. These were assigned to the Acceptable Group.

Information about the family name cannot be used for any policy or research purposes. This item was assigned to the Negligible Group. The complete grouping is as follows:

Essential Group: School Jurisdiction (Management); Home district; Band Code; Family number; Child Position; Year of Birth; Grade; On/Off Reserve Status; Accommodation; School district number.

Acceptable Group: School Name; Day of Birth; Month of Birth; Type of Course; Allowance; Language(s) spoken at first entry.

Negligible Group: Name of the child; family name.

Information about sex, religion and transfer of students, also collected from the data bank and field records, was not analyzed in the study. Moreover, a few other items of information stored in the data bank were excluded from the study because their importance was judged to be insignificant.

4. Date and Duration

The data retrieved from the enrolment data bank and collected from the field records is as of September 30 for each of the school years under review.

The Project was initiated in early December of 1974. The retrieval of data bank records and visits to the Kingston Education District office

were completed in January of 1975. Visits to the schools were made in April of 1975. The compilation, tabulation of data was done during June-July of 1975 while the analysis and reporting was carried out during August-September of 1975.

5. Evaluation

As mentioned earlier, it is assumed that the field records collected from the district and school offices are correct. The agreement or disagreement between the item-information in the data bank and the field records of the students in the sample is observed. Agreement is considered to represent the correctness of the item-information in the data bank and disagreement is treated as an error in the data bank.

(a) Preliminary Analysis:

Before proceeding with the detailed analysis of the data, a preliminary analysis was carried out. For the purpose of preliminary analysis, we examined the number of student records that have correct information for a majority of items in each group. To formalize this examination, we introduced a criterion concept for each student record.

A record was said to satisfy the essential criterion if at least six items of the essential group had correct information in the data bank. It should be pointed out that the essential group and essential criterion are different concepts. Similarly, a record was said to satisfy the acceptable criterion if at least four items of the acceptable group had correct information in the data bank. The record satisfied the negligible criterion only if both the

...7

items of that group had correct information in the data bank. Table I on page 15 shows the percentage distribution of records satisfying the various criteria.

We observe from this table that at least 86 percent of the records satisfy the essential criterion for each of the last four years. For the same years, the percentage of records satisfying the essential and acceptable criteria is at least 75 percent. The percentages of records satisfying all 3 criteria are also fairly high. However, the corresponding numbers are considerably lower for the school year 1969-1970. An examination of Table II on page 17, shows that only 1 to 2 percent of the records failed to satisfy any of these criteria in the last four years.

(b) Further analysis:

Two numerical measures, called the Record Reliability Index and the Aggregate Reliability Index, are used to analyse the data in greater detail. We also investigate later on the frequency of correctness for each item. As their names suggest, the Record Reliability Index assesses the reliability of the individual student records, whereas the Aggregate Reliability Index assesses the collective aspect of the reliability of the data for each school year.

The reliability indices are defined later. However, it is instructive to note that a reliability index always takes on a value between zero and one. A value of one indicates a complete agreement between the data bank and the field records; a value of zero indicates that every item of information in the data bank is wrong. Moreover, a higher value of the index indicates a higher accuracy of the data bank.

The development of the Record Reliability Index is based on the numerical weights assigned to each item in the student record. The values of these weights, although determined arbitrarily, are motivated by the relative importance of the various items and are required to have one as their sum for analytical purposes. The items with a relatively higher importance are assigned a higher weight. Each of the ten items belonging to the essential group is assigned a weight of 0.08, each of the six items in the acceptable group is assigned a weight of 0.03, and each of the items in the negligible group is assigned a weight of 0.01.

(i) Record Reliability Index:

Each student data bank record contains information about eighteen items pertaining to that student for the school year. A Record Reliability Index has been defined for each student data bank record as the sum of the weights of those items of the record that have identical information about them in the data bank and the field records. These are the items that have no errors in the data bank.

Mathematically, if R_{xy} is the record reliability index for the student x in the school year y, then

$$R_{xy} = 0.08E + 0.03A + 0.01N,$$

where

E = number of items having no errors in the essential group for the record x in year y.

A and N have similar definitions where the word essential is replaced by acceptable and negligible, respectively. For example, if 7 items in the essential group, 4 in the acceptable group, and 2 in the negligible group have no errors in the data bank for a particular student record x in the school year 1970-1971, say, then the record's reliability index $R_{x,\ 1970\text{-}1971}$ is given by

$$R_{x,\ 1970\text{-}1971} = \frac{0.08\ (7) + 0.03\ (4) + 0.01\ (2)}{} = 0.70.$$

A record reliability index of at least 0.9 means that, as a minimum, either

(a) Nine items in the essential group and all items in the acceptable and negligible groups are correct; or

(b) All the ten items in the essential group and at least four items in the acceptable group are correct; or

(c) All the ten items in the essential group, three items in the acceptable group and one item in the negligible are correct. An examination of Table III on page 19 shows that more than 61 percent of the sample records have a record reliability index value of at least 0.9 for each of the last four years of study.

A record reliability index of at least 0.8 means that, as a minimum, either

(a) Eight items in the essential group and six in the acceptable group are correct; or

(b) Eight items in the essential group, five in the acceptable group and one in the negligible group are correct; or

...10

(c) Nine items in the essential group and three in the acceptable group are correct; or

(d) Nine items in the essential group, and two items in each of the acceptable and negligible groups are correct; or

(e) All the ten items in the essential group are correct.

More than 75 percent of the records have a record reliability index value of at least 0.80 for the last four years. No more than 15 percent of the sample records for the last four years have a record reliability index of less than 0.60. The examination of the record reliability index distribution for the 1969-1970 school year shows that 49 percent of the records have a reliability index of less than .60. It is reasonable to conclude that the data bank is reliable for the years 1970-1971 to 1973-1974 but is less satisfactory for the school year 1969-1970.

(ii) Aggregate Reliability Index:

For a number of policy purposes, the interest lies in the reliability of the data as a whole as compared to the individual records. The Aggregate Reliability Index of a sample is defined as the average of the Record Reliability Indices for the records in that sample. The higher aggregate reliability index indicates a higher reliability-correctness of the data in the data bank of the sample. Table IV on page 21 shows the aggregate reliability indices. For the last four years, this index is at least 0.80, indicating a high reliability of the data as a whole. The contributions of the various groups to the aggregate reliability index are also exhibited in this table.

...11

(iii) Item Reliability:

Now we examine the various individual items for each year under study. Table V on page 23 gives the percentages of the data records having no errors for each of the items. It shows these percentages are quite high for items in the essential group, with the exception of item number 1, for the last four years. In particular, the percentage is greater than 90 in the school years 1971-1972 and 1973-1974. These percentages also indicate the reliability of the information relating to these items.

6. Manpower and Cost

The study was designed, developed and executed by a senior statistician in about thirty days. A data research officer and a statistical clerk worked on data collection and compilation, utilizing approximately 40 man-days of time.

The travel costs incurred while collecting the field information from the district office, the pertinent schools and the school counsellors amounted to $1,100.

7. Assessment of Objectives

Three of the four objectives set forth for conducting this study, as mentioned in Section 1, were mostly achieved. A suitable methodology for correcting the existing data in the data-bank could be developed later on, if necessary, after the results of the sample survey of all Canada records in the data bank become available.

The findings of this study and its recommendations will be listed separately later on.

...12

8. Responsibility

The Education Branch of the Indian and Eskimo Affairs Program was administratively responsible for this study and arranged for the visits of the study team to obtain the field records from the Kingston Education District.

The Departmental Statistics Division developed and organized the project, and received the data-bank information on computer printouts from the Computer Information Systems Division.

The Kingston Education District Office supplied approximately 25% of the field records while the school principals and education counsellors at Cornwall, Desoronto, Belleville, Golden Lake, Curve Lake, Lakefield and Alderville supplied the remaining information from their field records.

9. Some Findings

(a) Most of the children whose parents stay off a reserve are not covered in the data-bank and are also not recorded in the field reports, since the responsibility for their education is not that of the Department. They are, however, recorded on the band-membership list of registered Indians. This causes major variations in the three sources under reference.

(b) Students living along the border have a considerable mobility between the United States and Canada; consequently, their records are inconsistent because of irregular reports.

...13

(c) Records prior to the school year 1972-1973 are not readily available in the field; those for the school year 1969-1970 are very sketchy and hence do not permit a meaningful assessment of the reliability of the data-bank.

(d) Late reporting of births, kinship, affiliations of non-Indians with registered Indian families, adoptions and differences in local and registered names are some of the problems encountered in correctly updating the band-membership lists. Developing separate codes for provisional listings of doubtful and problem cases seems to be necessary.

10. Recommendations

The findings of this study have lead to the following recommendations:

(a) The proposed study for all Canada may be carried out on a sample basis, using a similar evaluation methodology.

(b) The reference period for the all-Canada study should include the period from the school year 1970-71 to the school year 1974-75 inclusive. Certain items of information for the school year 1975-76 may also be included in the study to enable the respondents to more easily identify the students covered in the sample through the use of current information rather than from past records alone.

(c) A separate listing of the children with off-reserve status should be done for the sample study since most of them are not a Departmental responsibility.

(d) The data-bank on student records for the school year 1969-70 should be excluded from extensive use because of its low reliability. This school year should also be dropped from the all-Canada study since adequate records are not available.

...14

(e) Children born between the years 1960 and 1964 inclusive should be included in the population for the all-Canada study for the purpose of selecting the sample units. This will ensure the coverage of proper school-going children in the age group 6-10 for the school year 1970-71 and in the age group 10-15 for the school year 1974-75.

(f) A separate study could be undertaken for the students commuting between the United States and Canada to consider such items as their frequency of moving, financial liability to the Department, and methods of reporting.

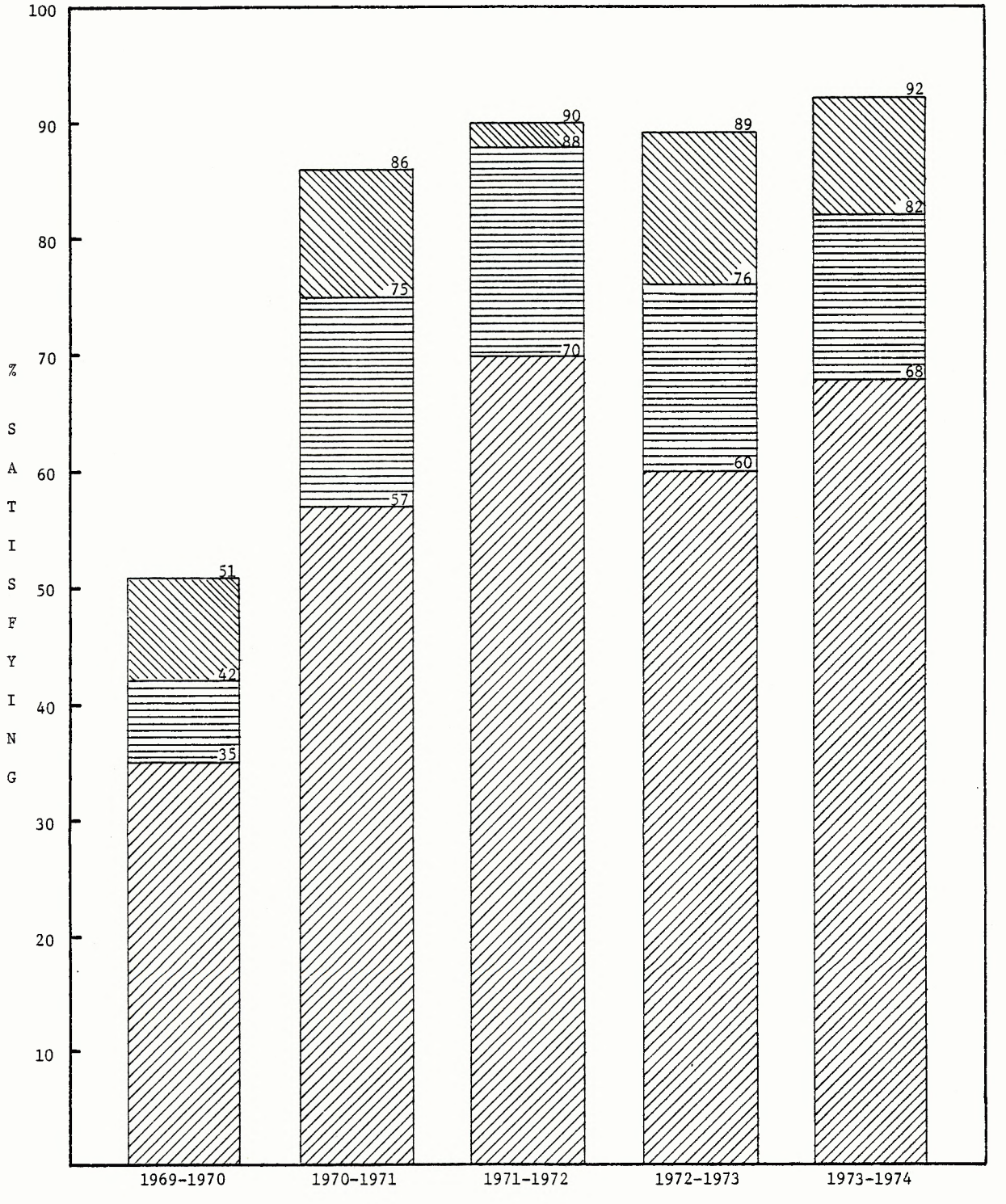
(g) The all-Canada study should be conducted by Departmental Statistics Division staff at Headquarters, supported by administrative and clerical assistance from the regional and district offices.

TABLE I

PERCENTAGE DISTRIBUTION OF RECORDS SATISFYING SPECIFIED CRITERIA

School Year	No. of Records In The Sample	Percentage Satisfying		
		Essential Criterion	Essential and Acceptable Criteria	All Criteria
1969-1970	96	51	42	35
1970-1971	99	86	75	57
1971-1972	95	90	88	70
1972-1973	101	89	76	60
1973-1974	102	92	82	68

CHART 1
PERCENTAGE DISTRIBUTION OF RECORDS SATISFYING SPECIFIED CRITERIA
1969-1970 TO 1973-1974



LEGEND

All (EAN) Lower Column

Essential and Acceptable (EA) Lower and Middle Column

Essential (E) Whole Column

TABLE II

DISTRIBUTION OF STUDENT RECORDS SATISFYING CRITERIA GROUPS

Criteria Groups	School Years				
	1969-1970	1970-1971	1971-1972	1972-1973	1973-1974
	number of records				
EAN.....	34	57	66	61	69
EAn.....	6	17	17	16	15
EaN.....	8	11	2	13	9
Ean.....	1	-	-	-	1
Total 'E'	49	85	85	90	94
eAN.....	-	-	-	1	1
eAn.....	-	-	-	-	-
eaN.....	38	13	9	8	6
ean.....	9	1	1	2	1
Total Records	96	99	95	101	102

Symbol: - Nil or Zero

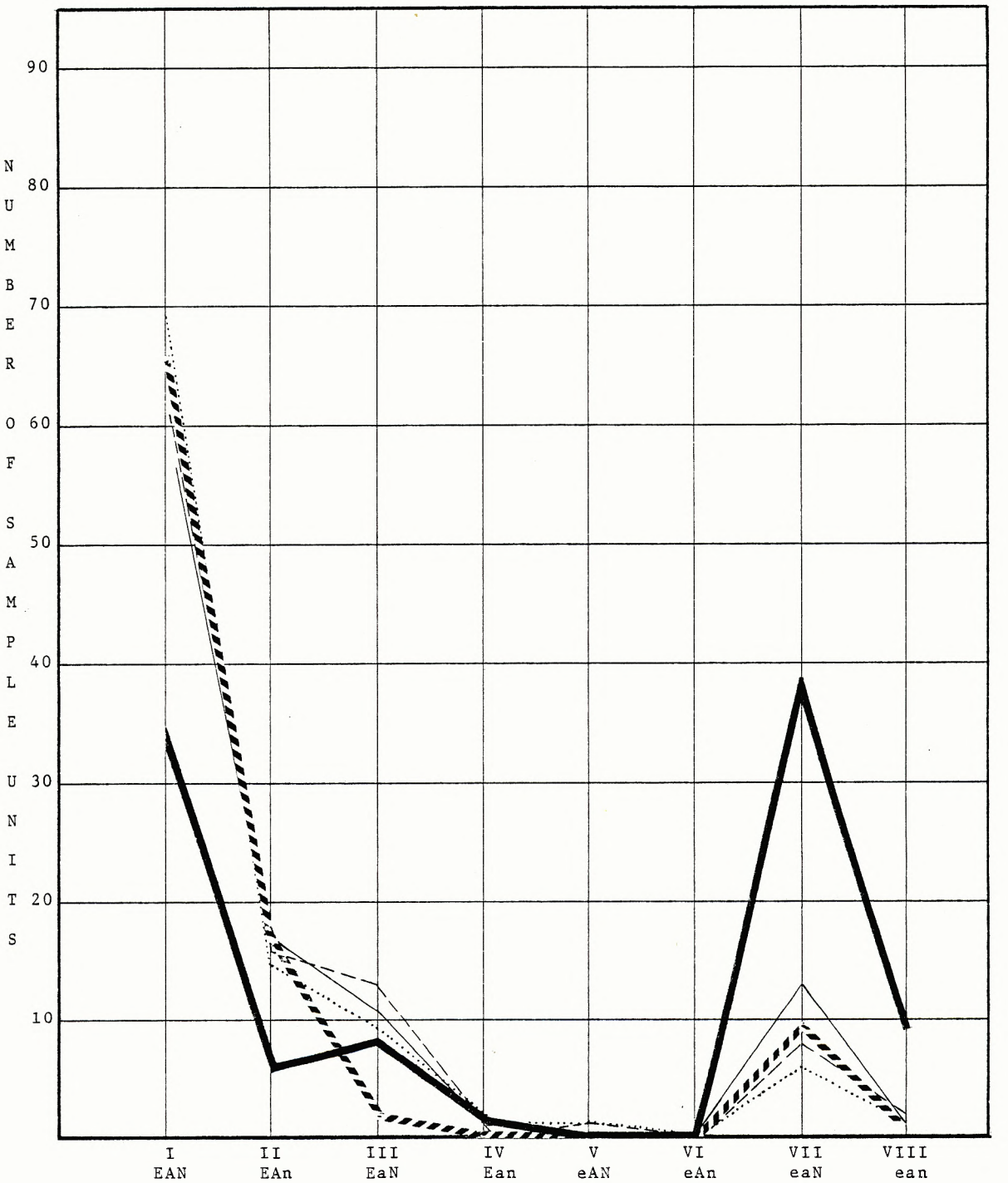
LEGEND

E:6 to 10 items from 'Essential' group tally.
e:0 to 5 items from 'Essential' group tally.

A:4 to 6 items from 'Acceptable' group tally.
a:0 to 3 items from 'Acceptable' group tally.

N:Both items from 'Negligible' group tally.
n:0 or 1 item from 'Negligible' group tally.

CHART II
DISTRIBUTION OF STUDENT RECORDS SATISFYING CRITERIA GROUPS
1969-1970 TO 1973-1974



L E G E N D

E: 6 to 10 items from 'Essential' group tally.
e: 0 to 5 items from 'Essential' group tally.

A: 4 to 6 items from 'Acceptable' group tally.
a: 0 to 3 items from 'Acceptable' group tally.

N: Both items from 'Negligible' group tally.
n: 0 to 1 items from 'Negligible' group tally.

1969 - 1970 ━━━━━
1970 - 1971 ———
1971 - 1972 ⟋⟋⟋⟋⟋⟋
1972 - 1973 – – – – –
1973 - 1974 .....

TABLE III

FREQUENCY DISTRIBUTION OF STUDENT RECORDS BY THEIR RELIABILITY INDICES: 1969-1974

Groups[1]	School Years				
	1969-1970	1970-1971	1971-1972	1972-1973	1973-1974
	number of records				
0 -.04....	47	14	7	11	7
.50-.54....	-	-	-	1	-
.60-.64....	-	-	-		1
.65-.69....	-	1	-	1	4
.70-.74....	1		-	2	3
.75-.79....	1	1	1	10	3
.80-.84....	3	4	1	8	4
.85-.89....	3	18	11	5	12
.90-.94....	24	26	16	14	17
.95-.99....	17	35	35	24	20
1.00....	-	-	24	25	31
Total Records	96	99	95	101	102

Symbol: - Nil or Zero

Note: [1]Series with Reference to Record Reliability Indices

CHART III
FREQUENCY DISTRIBUTION OF STUDENT RECORDS BY THEIR RELIABILITY INDICES
1969-1970 To 1973-1974

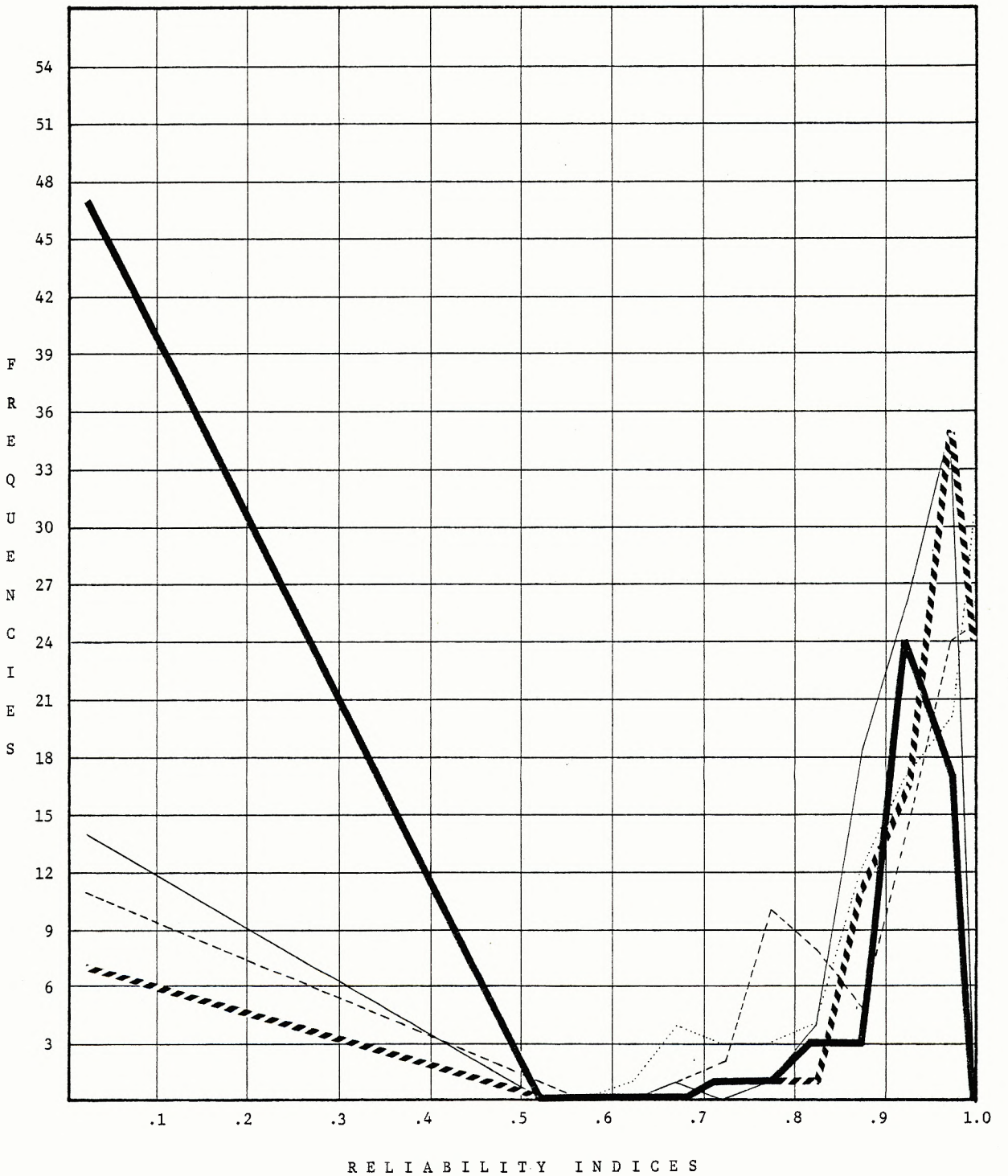


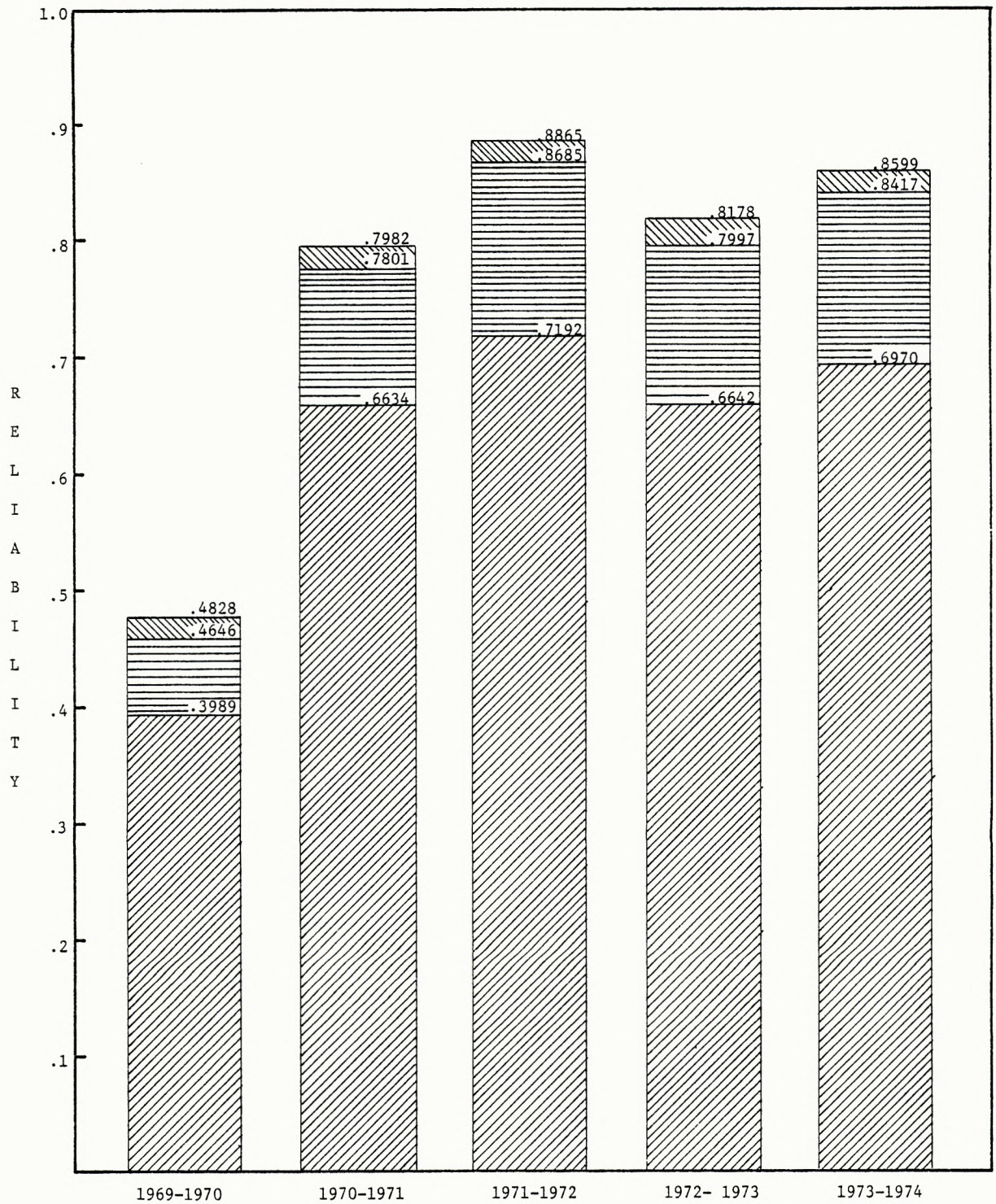
TABLE IV

CONTRIBUTIONS OF THE GROUPS TO THE

AGGREGATE RELIABILITY INDICES

Year	Essential Group	Acceptable Group	Negligible Group	Aggregate Reliability Index
1969-70	.3989	.0657	.0182	.4828
1970-71	.6634	.1167	.0181	.7982
1971-72	.7192	.1493	.0180	.8865
1972-73	.6642	.1355	.0181	.8178
1973-74	.6970	.1447	.0182	.8599

CHART IV
CONTRIBUTIONS OF THE GROUPS TO THE AGGREGATE RELIABILITY INDICES
1969-1970 TO 1973-1974



LEGEND

Essential (E)		Lower Column
Acceptable (A)		Lower and Middle Column
Negligible (N)		Whole Column

TABLE V

PERCENTAGE OF CORRECT RECORDS FOR ITEMS

STUDIED FOR RELIABILITY

1969-70 TO 1973-74

Year	No. of Records	Essential										Acceptable						Negligible	
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	1	2
1969-1970	96	50	51	51	51	51	51	45	51	50	48	48	50	51	49	0	21	86	96
1970-1971	99	80	86	86	86	86	86	72	86	85	78	83	85	86	81	0	56	85	96
1971-1972	95	81	93	93	92	93	93	80	93	92	92	92	89	91	89	75	62	84	96
1972-1973	101	56	89	89	88	88	86	79	89	97	72	74	86	86	80	64	62	85	96
1973-1974	102	65	93	91	93	92	91	81	91	90	83	83	89	83	79	71	67	86	96

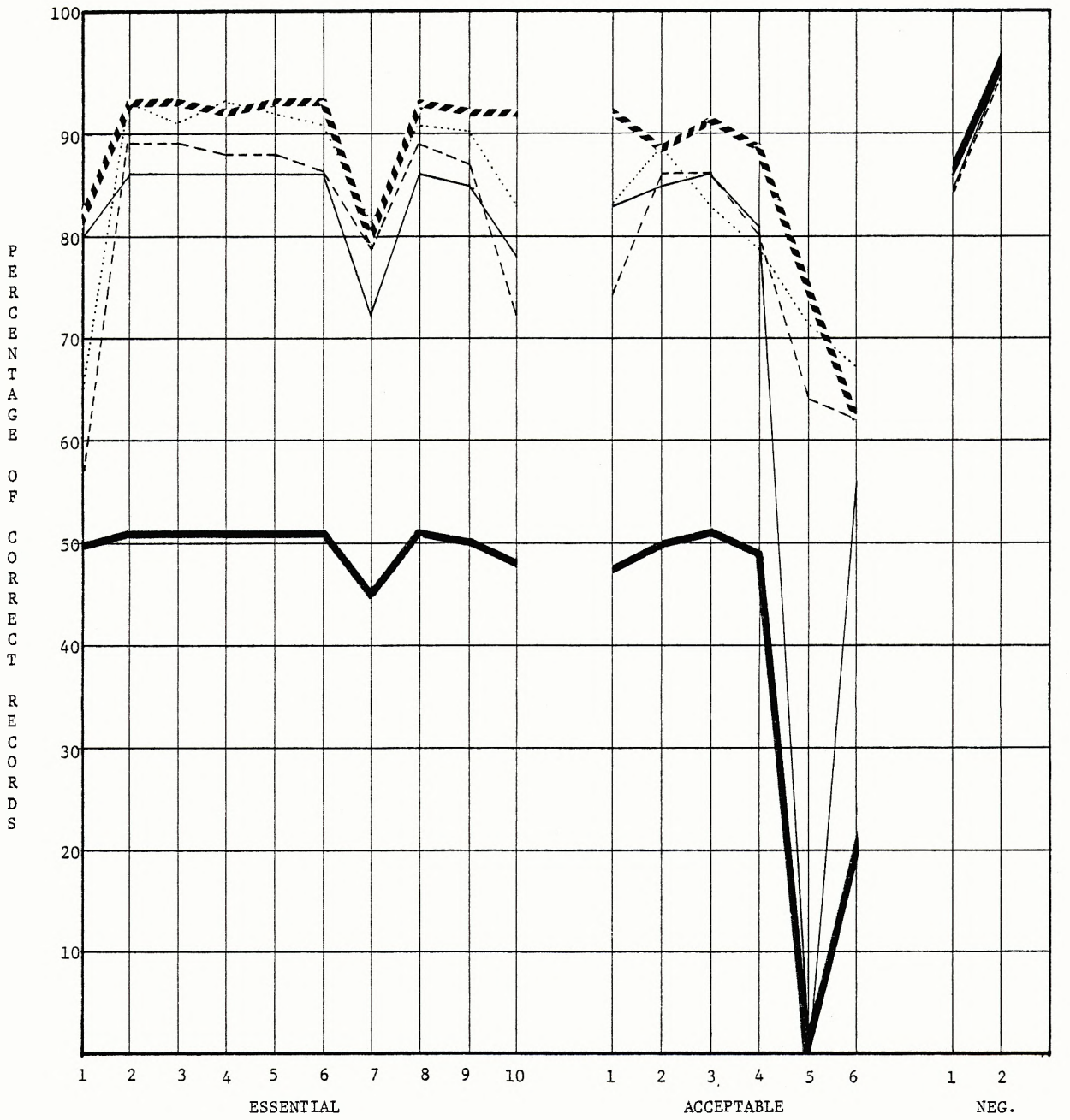
LEGEND

Essential
1. School Jurisdiction
2. Home District
3. Band Number
4. Family Number
5. Child Position
6. Year of Birth
7. Grade
8. On-Off Reserve
9. Accommodation
10. School District No.

Acceptable
1. School Name
2. Day of Birth
3. Month of Birth
4. Type of Course
5. Allowance
6. Language Spoken

Negligible
1. Name of the Child
2. Family Name

CHART V
PERCENTAGE OF CORRECT RECORDS FOR ITEMS STUDIED FOR RELIABILITY
1969-1970 to 1973-1974



LEGEND

Essential (E)	Acceptable (A)	Negligible (N)
1. School Jurisdiction	1. School Name	1. Name of the child
2. Home District	2. Day of Birth	2. Family Name
3. Band Number	3. Month of Birth	
4. Family Number	4. Type of Course	
5. Child Position	5. Allowance	
6. Year of Birth	6. Language Spoken	
7. Grade		
8. On-Off Reserve		
9. Accomodation		
10. School District No.		

1969-1970 ━━━━━
1970-1971 ———
1971-1972 ▰▰▰▰▰▰▰
1972-1973 - - - - - - -
1973-1974 ·····