DOCUMENT RESUME

ED 064 351 TM 001 518

AUTHOR Thorndike, Robert L.
TITLE Reading Comprehension Across National Boundaries.
NOTE 7p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests; Comparative Analysis; *Cross Cultural Studies; Elementary School Students; *International Programs; *Predictive Validity; *Reading Tests; Secondary School Students; Test Construction; *Test Results

ABSTRACT

Some aspects of the results from IEA studies of reading are discussed. The instruments used in the studies were a reading comprehension test of the conventional type and a short reading speed test. The studies were conducted with 10- and 14-year-olds and individuals in the last year of secondary school. A supplementary measure, Same-Opposite vocabulary test of 40 words, was also used at each level. Among the 15 countries that participated, 12 different languages were represented. Difficulties encountered, results for the United States, between-country differences, and between-schools differences are discussed. Predictors of reading achievement are discussed from the standpoint of their effectiveness. (DB)

Reading Comprehension across National Boundaries

Robert L. Thorndike

Teachers College, Columbia University

For my part in today's proceedings, I plan to talk mostly about some aspects of the results from our IEA studies of reading. In the course of talking about these results, I shall probably have occasion to consider with you some of the methodological problems that arise in the sort of study in which we have been engaged, but I will discuss these in the context of the materials and the findings on reading comprehension and reading speed.

Dr. Postlethwaite has taken responsibility for telling you something about the development of the instruments for the latest of the IEA surveys. In the case of reading, the instruments of primary concern were two. One of these was a reading comprehension test of the conventional type, in which the examinees read a passage and answered multiple choice questions upon it. Each level of the comprehension test was made up of two sections composed of four passages and 20 to 30 test items depending upon the level of pupil involved. As you have already heard, we have worked with 10-year-olds, 14-year-olds and individuals in the last year of secondary school. We speak of these as Populations 1, 2 and 4, and in what follows some of the materials and discussion will be couched around these three populations: Population 1 - the 10-year-olds, Population 2 - the 14-year-olds, and Population 4 - the group at the end of secondary education.

In addition to the reading comprehension test, the examinees in Populations 1 and 2 were also given a short reading speed test. This consisted of very short paragraphs each one made up of three sentences, the third of which was a question. The question was to be answered by underlining one of three words. The successive paragraphs dealt with a continuous story about a boy and a dog who walked out into the field, what they saw and what they did. The test was intended to be very easy reading, and the questions were designed to be answered correctly by 90% or better of students. The little test consisted of 40 of these paragraphs and was given with a four-minute time limit.

We had hoped to include some study-type reading in which an extended passage is studied and removed before testing, and to get evidence on this type of performance, but practical limitations of time, both to produce materials and to apply them to the students, led us eventually to abandon this and restrict ourselves to the comprehension test and the speed test.

As you have also heard, we had as a supplementary measure a Same-Opposite vocabulary test of 40 words at each level. This served as a simple descriptor of word knowledge and vocabulary size, but will play relatively little part in what I would like to talk about today.

Dr. Postlethwaite has talked in general about the test development, but perhaps I should amplify a little bit some of the special problems in connection with reading tests. The most obvious one is the problem of language. Among the 15 countries that participated in the administration of the reading tests, 12 different languages were represented. Operationally, this meant that when test materials were initially supplied in some other language they had to be translated into English as a common medium of communication among the countries. They were

reviewed and criticized by each national committee with regard to their appropriateness to that country and with regard to the technical aspects of the passages and such items as were initially available on them. Items were written in English and were reviewed by individuals from the different countries. Each passage was translated into at least two languages other than English for preliminary try-out, as part of the development of the test materials. The final passages and items were selected on a basis of the item statistics from generally three and sometimes four different countries, and final versions of the passages and items were prepared--but still in English. At this point, all materials had to be translated into the language of each of the participating countries. We were encouraged that this could be done without seriously disrupting the character of the reading materials on the basis of our previous experience in the earliest study that we did back in 1962, in which it appeared that not only general level of difficulty but even the difficulty of specific items showed good stability from language to language.

Each national center was responsible for handling the translation into its own language, and each national center was encouraged to provide an independent translation back into English by a separate translator to see to what extent the character of the original material had been maintained through the process of translation. However, both time and resources have been limited, and we received at most a back-translation of a sampling of the passages and items. Most of the participating countries felt that their limited resources were better spent on doing dual translations from English into their own language and trying to maximize the precision and quality of the forward translation, and consequently any evidence from a re-translation as to the precision with which the original passage had been maintained in translation is quite limited.

We also have data for the present tests concerning the consistency of item difficulty and other item statistics from country to country, and we have data on the popularity of the error choices in different countries for each item, but this material is voluminous and has not yet been very thoroughly examined. I can say in general that the correlations of item difficulties across languages are quite substantial, with the possible exception of those countries in which students had a very great deal of difficulty with the material and were responding to substantial numbers of the items at a chance or near chance level.

Cross-national studies have two unique contributions to make to our understanding of educational phenomena. On the one hand, one is interested in parallel analyses within each country, in which case the countries serve as replications of an experiment and provide an opportunity to test in a broader context relationships that have been previously observed in a single national system. It is also possible to make comparisons across countries, seeing in what respects countries differ in their performance and attempting to understand these differences in terms of the characteristics of the countries involved. We have repeatedly stated that we are not interested in an intellectual Olympic games, but in using national differences in educational practices and procedures as a quasi-experiment in different educational treatments. Hence, the kinds of analysis which I shall present today will be at three levels. On the one hand, there will be analyses at the level of the individual pupil. We have a number of types of data which characterize the pupil as an individual, and it is possible to study within each country individual correlates of reading achievement. At the second level, we will be interested in analyses in which the school is the unit; we will be trying to identify those characteristics of the school as an institution that are associated with the reading achievements of its students. Here, it soon becomes apparent that the most

important determiners of the reading achievement of the pupils in a school are the characteristics of the pupils themselves in terms of their backgrounds and out-of-school resources. One of the sticky problems to which we will need to give some attention is the problem of identifying school effects, as distinct from the input characteristics of the students that represent the clientele of a particular school. Finally, as I said earlier, we are interested in analyses that use the country as a unit and compare characteristics of performance with other variables from country to country. In what follows, I want first to talk about between-country differences, then to talk about between-pupil differences and finally to talk about between-school differences.

As I talk about the test performance of the pupils in different countries, I shall identify only the results for the United States. When the full results appear, some analyses by country will be made in which individual countries will be identified, but it is not our purpose either to applaud or point the finger at any country in relation to its achievements. We are interested only in understanding, and in this context it is the nature of the country and its economic, social and educational system that we are interested in trying to relate to the achievements of school pupils.

As a very first question, we might ask how big the differences in reading test performance were from country to country at the three educational levels at which we were working. In order to provide a kind of frame of reference for looking at the size of these differences and at their direction in relation to performance in the United States, I have scaled each country in relation to the mean and standard deviation of the United States. That is, the United States defines the baseline, and deviations above or below that baseline are expressed in standard deviations of the USA distribution. (Parenthetically, I might say that the variability of performance in the United States is either uniquely the largest or tied for largest of any of the countries involved. The heterogeneity of reading performance is very great here.)

I call your attention to Chart 1 in the handout that has been prepared, which shows the amount and direction of the differences of the various countries from the USA baseline or USA par. You will notice that in Populations 1 and 2, the means for a large number of the countries cluster rather closely around the United States mean and the United States seems to be fairly representative of this large cluster of primarily European or European-oriented countries. These are typically economically and industrially developed countries with a practice of universal education for individuals up through the age of 14.

Included in our study this time were three relatively underdeveloped countries, and their lack of economic development and of a background of universal education shows up very dramatically in their performance on the tests. These are the three countries that fall very clearly at the bottom of the distribution of average reading scores in all three of the populations that we studied. You will notice that the discrepancy is a full standard deviation or more in terms of the United States distribution and that these countries deviate much more dramatically from the baseline than do any of the other countries.

It is clear from this chart that the United States shows up very much less well at the end of secondary school than at the two earlier levels. This is a familiar phenomenon and we all recognize that it results in substantial measure from the much higher proportion retained in school through the end of secondary education in the United States than in the other countries with which we had to deal. This

greater retentivity can be documented in many ways, and there are many kinds of evidence that relate to and explicate this difference between the educational system in the United States and in the European countries.

Another way of looking at the reading performance in the various countries with which we worked is to ask what proportion of the children were scoring at or not significantly above a chance level on the test that we gave. This is, in a way, saying what proportion were "reading incompetents" relative to a test that was designed to be appropriate for the average youngsters in most of the countries on which the test development was carried out. These percentages are shown in Chart 2, and document even more dramatically the plight of the underdeveloped countries in terms of the substantial proportion of their children in school who appear illiterate or nearly so. As you can see, the percentage at each level goes up to a high of nearly 50%.

One further indication in this same direction comes from a study of the Reading Speed Test. This test was designed to be at a very simple level so that almost everybody could get a very high proportion of the items right. In order to check upon the extent to which this was actually the case, each country was asked to score the first page of the Reading Speed Test for errors and to report an error score for that first page as well as reporting a speed score represented by the last item attempted. We reproduce for you as Chart 3 in the hand-out the first page of the Reading Speed Test, so that you can s e concretely the level of reading with which we are dealing here.

The question that we raise is: For what proportion of children does this test function genuinely as a speed test and for what proportion does it become substantially a power test in which errors are of frequent occurrence? Setting a frequency of error at which the test becomes a power test is obviously somewhat arbitrary, but for purposes of illustration, I have chosen three or more errors as an indication that the individual was having genuine difficulty in understanding this simple material as he read it and tried to mark the answers. Table 1, which follows the speed test, indicates for both the 10-year-olds and the 14-year-olds in the participating countries the percent making three or more errors on this little set of nine quite simple and straightforward items. This, to me, dramatized very sharply the reading problem that the developing nations face.

With the extreme differences that are found within this set of 15 national groups, it seems clear that any one of a variety of economic or educational indicators will give a substantial prediction of the between-country differences in educational achievement. I have selected a few from the questionnaire that was completed by the 14-year-old children in the study and show them in Table 2. Clearly, the level of education of parents is a very potent indicator of reading level across countries (and we will presently see that it is one of the better indicators within a country). However, indicators such as the availability of magazines, books and newspapers; TV viewing; and other resources and amenities are also substantially related to this striking difference in level of performance. The correlation with national TV viewing is particularly dramatic.

Enough for between-country differences. Let us now turn our attention to between-pupil differences. What are some of the factors that are associated with the reading achievement of single pupils and to what extent are these consistent across the range of countries with which we have to deal?

Naturally, the first place one looks for information that will predict the performance of the individual is in the social and cultural background of the child in his home and family. We had several indicators of home and family status. The information on these was undoubtedly much less than perfectly accurate, since it was furnished by the children themselves in response to a questionnaire dealing with a multitude of facts about their background and schooling, and at least in some countries furnished by children who would certainly have had very great difficulty in reading the questionnaire, in view of their obviously limited reading ability. In Table 3, I have listed several of the home background predictors that were most effective, and have shown the range of correlations for these across countries. As before, at this time I identify the relationship only within the United States.

The general impression that one gets from this table is that to a very considerable extent the relationships are consistent from country to country, and that father's occupation, parents' education, and size of family mean much the same thing in each of the countries studied. In those countries in which the relationships break down, we do have the very serious question of whether the questionnaires were adequately completed by the rather limited readers that had to work with them. Thus, a good deal of consistency of pattern emerges across countries and continents as far as the order of magnitude of the relationships that are involved for the home indicators of reading competence.

One issue of special interest is the relative reading performance of boys and girls in different national cultures. We have a long series of studies and a strong tradition that girls develop more rapidly in and are better at verbal skills than boys are, but a large part of this research as it is known to us in the United States is based upon testing in this country. In our present data, we have results on boys and girls in 15 different countries at three different levels of maturity, and it is of some interest to inquire about the size and direction of the sex differences country by country. Table 4 shows the amount by which girls surpass boys or fall behind them, expressed in units of the total group standard deviation, for each level and country.

The typical situation is for girls to do slightly better. This occurs in 11 of 14 countries at age 10, in 8 of 14 countries at age 14, and in 8 of 13 countries at the end of secondary school. But the variability in this relationship is considerable, and at each age there are some countries where the boys did better. The variability is especially pronounced at the end of secondary school, where national policies and expectations with respect to continuing education seem likely to play their largest role. However, even at ages and in countries where the proportion of the two sexes is about equal, there is still a considerable variation. One has a feeling that cultural factors are playing a considerable role.

Finally, we turn to differences in performance between schools. Here, we are looking for aspects of the school situation that facilitate or hamper the reading performance of the individuals in that school. It is at this point that we ran into some of our most serious methodological difficulties. The difficulties stemmed in part from the fact that the information upon which we had to rely was information provided in questionnaires by a school administrator, teachers in the school, and pupils in the school. Thus, we were dependent upon second-hand reports of conditions and procedures rather than any direct evidence of what went on within the school. We could get reports on the expenditure level of the school, the size of classes in the school, the characteristics of the teachers in terms of their age, sex, and training background, the availability of auxiliary resources such as reading specialists, guidance counselors and school psychologists, and other

auxiliary resources such as libraries and librarians, books in the classroom and so forth. However, we were not able to get any detailed picture of the classroom materials or instructional procedures excepting as these were reported in general terms by the teachers.

A second difficulty arises from the very large number of specific bits of information that are generated by a questionnaire that seeks thrcugh many indicators to characterize a school program. We found that the several questionnaires generated literally hundreds of items of information about each school, each one being a weak little indicator of something about the school situation. We ended up with substantially more predictors than we had schools in any given country, and we faced a very serious problem of degrees of freedom in an analysis of this sort. We needed in some way to reduce the number of variables to a manageable size.

We encountered a third and perhaps most serious difficulty in that the clearly potent predictors of reading achievement for a given school related to the average quality of the pupil input in terms of the kinds of indicators that were effective for predicting achievement of single individuals. Thus, in Table 5 are shown the correlation of scaled father's occupational level, average father's education, average mother's education, and average number of books in the home with average reading achievement for the various countries at level 2. The median size of these correlations is shown in the last row of the table and it can easily be seen from these correlations that a very substantial proportion of the variance in reading achievement for pupils in any given school is accounted for by factors that lie outside of the school's control and that represent the life background from which the individual pupil has come. Under the above circumstances, it is only too possible that any school variables that turn up with appreciable correlations have these correlations because of the fact that they are also related to the type of pupils attending that school. For the full and formal analysis of school variables as predictors of achievement, then, it is necessary to look not only at the initial relationship between the variable and the achievement measures but also at the partial correlation when a complex of these individual background variables is partialled out. As a way of partialling out differences in input, we developed what one of my colleagues has called a "school handicap score" to represent the expected achievement based upon a composite of these individual background factors, and have looked at the residual correlations with achievement after these background factors had been statistically partialled out.

There are many too many variables characterizing schools to make it either possible or meaningful to present them fully at this time. I have picked out a few that some of you might have anticipated a priori to be effective predictors of school achievement and have prepared in Table 6 a representation of their original zero-order correlations with reading achievement. There is too much detail in the table for you to apprehend at the present time, but it will pay sobering study. Let me just comment on one or two of the characteristics that you may find there.

In general, educators argue for the need for a higher level of funding of education, and the implication is that more expenditure per pupil should yield higher levels of pupil achievement. You will find in the table a variable on per pupil expenditure, based upon the reported expenditure level supplied by the administrator of the school. The accuracy of these reports is somewhat suspect. However, as you examine the column of correlations for this variable, I think that you will agree that the relationship is puzzling and somewhat disillusioning.

A second point that is frequently argued is that smaller classes are highly desirable for educational achievement. The implication is that children in smaller classes should achieve more than children in larger classes. If you look at the column labeled mother tongue class size, you will find that the direction of the zero-order correlations is as often the reverse of what this doctrine would propose as it is consistent with it. At the superficial level, the children in the larger classes often make out rather better than the children in the smaller ones. You might quite reasonably argue that the relationship that we are concerned with here is not a linear one and that the linear correlation that we display is inadequate to represent it. You could also argue that the children who are assigned to small classes are assigned to small classes because they are poor performers and that the direction of causation is not from class size to achievement but from achievement to class size. You could also point out that small classes are likely to occur in small rural schools where other handicaps exist that are not reflected adequately in our school handicap score. This is also a reasonable proposal. However, be that as it may, the evidence that we have gives little encouragement to expecting substantial improvements in achievement by reducing the size of the classes.

Other elements that would typically be thought of as supporting reading achievement are a staff of reading specialists, a school library, books in the classroom and other supporting auxiliary school personnel. We show evidence on a number of these factors in the table. Again, there is no consistent evidence of a positive effect from these supporting services, and in some instances the relationships are fairly consistently negative. Once again, we see only relationship and not causation. Once again, we can argue with some cogency that remedial teachers and other kinds of special services are concentrated in those schools where problems are known to exist. To this extent, a negative relationship may represent an adaptation of a school system to the problems that it faces. However, we find a great dearth of evidence that institutional arrangements designed to help youngsters in their efforts to read well and effectively do in fact do so.

Generally speaking, our whole effort to identify school causes of achievement, at least in the reading area, has seemed a rather unproductive one--at least to me. At this point, I am not sure that I know fully why we have done no better than we have. Of course, our findings are not without precedent. The Coleman Report did not tend to attribute great potency to school influences upon achievement in the United States, and perhaps we should not have expected to find them in the countries that we studied either. Certainly, the indicators that we have are limited and rather superficial ones and no one of them could be expected to carry a great deal of predictive weight. That is why we had hoped to merge them into compounds and composites that would be more effective as predictors. However, there seems to me to be little to compound if one looks at these variables individually. The results from country to country are modest and inconsistent, and one questions whether putting the little bits together in either an a priori fashion or an empirically determined one could add up to a very great deal.

These last findings have been somewhat discouraging to me. I'm not sure quite what I'm discouraged about more profoundly--the kind of international survey in which we have engaged or the educational enterprise itself. Of course, we are not making a comparison between schooling and no schooling. We are trying to differentiate the effects of different types and qualities of schools within a given cultural setting. The variability among schools may be small relative to the impact of schooling taken in toto. From this point of view, variability may not be terribly important or terribly productive to study. However, we do seem to have drawn a rather poor hand as far as being able to provide strong cues as to what it is about a school environment that results in better reading achievement in the pupils thereof.