DOCUMENT RESUME

ED 145 677 FL 008 907

AUTHOR Angelis, Paul J.
TITLE Language Testing and Intelligence Testing: Friends or Foes? Occasional Papers on Linguistics, No. 1.
INSTITUTION Southern Illinois Univ., Carbondale. Dept. of Linguistics.
PUB DATE Apr 77
NOTE 9p.; Proceedings of the International Conference on Frontiers in Language Proficiency and Dominance Testing (1st, Southern Illinois University, Carbondale, Illinois, April 21-23, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Adult Students; Aptitude Tests; English (Second Language); Higher Education; *Intelligence Tests; *Language Proficiency; *Language Tests; National Competency Tests; Native Speakers; Second Language Learning; *Testing; Testing Problems; *Test Interpretation; Test Results; *Test Validity
IDENTIFIERS Graduate Record Examinations; Test of English as a Foreign Language

ABSTRACT

This paper discusses the relationship between language proficiency and intelligence. In particular, the paper is concerned with the elements of intelligence testing which creep into tests designed to determine language proficiency, and the proliferation of testing of all types and the kinds of interpretations made of these tests. Particular reference is made to the non-native adult speaker of English who is frequently subject to these types of tests. Native-speaker data indicate that the reading section of the Test of English as a Foreign Language (TOEFL) presents difficulty for the native and non-native speaker of English alike, and that both groups find this section the most difficult of the test. A comparison of TOEFL and Graduate Record Examination (GRE) scores for foreign students applying for admission to Texas A&M University did not show a very high correlation. A Swedish study which attempted to examine the relationship between proficiency in English as a second language and various intelligence factors was also unable to find high correlation between the two types of test. Indications are that the GRE and similar tests are not appropriate for determining second language proficiency. More investigation is needed of the relationship between language and intelligence; in particular, joint research on this question by linguists, psychologists, and measurement specialists is needed. (Author/AM)

Language Testing and Intelligence Testing: Friends or Foes?

Paul J. Angelis

The purpose of this paper, as the title indicates, is to raise a question. And, as is the case in many instances where questions are raised, there is often no promise of a satisfactory answer. But such a situation, while often intellectually unsatisfying, seems appropriate in this case. Since the theme of this conference focuses on frontiers in the field of language proficiency testing and dominance, it would appear in order to raise issues which, at least until recently, have been on the borders of at least two complementary fields of inquiry: linguistics and psychology. More specifically, to focus on one of the principal points of emphasis at this conference, it would appear both logical and necessary to discuss the relationship between language proficiency and intelligence.

I do not propose, however, to present any suggested models for what this relationship might be. Nor do I pretend to evaluate any such models which may already have been developed. On the contrary, my remarks stem from two concerns, one theoretical and one very practical. The former refers to my own curiosity concerning the elements of intelligence evaluation which creep into tests designed to determine language proficiency. The latter refers to my concern over the proliferation of testing of all types and the eventual conflicting and sometimes careless interpretations which are made of test results by counselors, advisors and administrators in general. Perhaps some knowledge gained on the first point will help to solve some of the difficulties concerned with the second.

I will not be concerned in this paper with the relationship of intelligence to language in children. Although we are far from any clear answers as to what this relationship might be, enough has been written on this subject to at least point the way toward resolving some of the most immediate problems in this domain. What I shall discuss is the question of language testing and intelligence testing as they relate to adults in a second language context. In particular, I shall be concerned with the interplay of these factors for those non-native speakers of English who are most frequently subject to these types of testing: foreign students in English medium institutions of higher learning, our own colleges and universities.

At least one other paper presented at this conference has investigated the role of language proficiency in intelligence tests (Oller). The first of my two-part question raises somewhat the opposite point: What is the role of intelligence in language proficiency testing? Because of my acquaintance with the Test of English as a Foreign Language (TOEFL) and my role as a member of the TOEFL research committee, many of my remarks will stem from a knowledge of current studies relating to that test. Many of these studies are not yet complete and others are yet to be implemented but even looking at what questions are being investigated should raise some interesting points for discussion.

The first study, the only one for which we have some actual data, investigates the performance of native and non-native speakers on the TOEFL test. The assumption underlying most foreign language tests, including TOEFL, is that the items inlcuded should represent features within the competence of all native speakers of the language. In the case of TOEFL, which is used to measure the English proficiency of non-native speakers seeking admission to post-secondary academic institutions,

a further restriction, however, is that the level of language would be appropriate for all "educated" native speakers of English. The interpretation of the word "educated" is perhaps the source of some of the problems we shall discuss. But basically, what is intended in such a test is that it would present no difficulty to native speakers of the same age and educational level as those being tested.

Only limited research has been done to validate this statement. One study by Angoff and Sharon (1971) did show that native speakers did not seem to have any difficulty with the items on the test. But because no analysis of items was done as part of this study and also because a new form of TOEFL had been designed which included certain more functionally oriented items types in both the listening comprehension and reading section, a new study was proposed by the TOEFL research committee early in 1976. The subjects for this study were 88 college-bound high school seniors, all of whom were native speakers of English and who had little or no exposure to foreign languages beyond some limited high school courses. The results of that limited study raise some questions which relate to our topic of language and intelligence.

The assumption once again was that the college bound native speakers would have no difficulty with the test. One area of uncertainty, however, was the listening comprehension section where pre-testing with non-native speakers seemed to show that the new "mini-talk" items were more difficult than expected. While the total scores achieved in the study indicated that the native speakers did well on the test as a whole with mean raw scores of 134.4 and 134.9 on the 150 item test compared to the non-native speakers' mean raw scores of 89.6 and 88.5 respectively, the listening comprehension section posed almost no difficulty for the native speakers (Clark). Only 3 of 100 items on the two forms of this section of the test had a pass rate below 80%.

In the Structure and Written Expression section, however, 22 items of the 80 on the two forms of the test had below 80% pass rates for the native speakers. Most of these were on the error recognition items and dealt with points such as noun-verb agreement, informational redundancy and parallelism. To find so many cases of native speaker errors is probably surprising. The pessimists (or perhaps we should say realists) among critics of the current status of the American educational system would no doubt simply cite this as confirmation of the sad state of our pupils' ability to use standard written English.

The Reading and Vocabulary section, however, raises more serious problems. Here, 22 out of 120 items on the two forms of the test had below 80% correct rates for the native speakers. Eight of these were vocabulary items and although some such as "broach" and "raze" might be explained on the basis of low frequency, others such as "scope" and "remuneration" do not. It is in the reading section where we begin to find an apparent conflict between factors of language and intelligence. Six of the items in question dealt with factual information but the particular passages were later found to deal partly with abstract concepts. Three others were sentence restatement type items. But the remaining five items with less than 80% correct rates were found to be based on passage summarization or interpretation. This is even more significant when we note that none of the five had pass rates near 80% (the range was from 67.44% to as low as 35.56%). Also these five items represented one half of the number of this type on the two sections of

the test. Clearly, the reading items of this type on the test presented an unusual amount of difficulty for the native speakers who took the test.

A closer look at the reading items which proved to be the most difficult showed that they contained statements such as "it can be inferred, the author assumed," or "the author would be most likely to agree". The performance of mental processes which focus on inference and evaluation may indeed be part of what researchers view as "reading" but the fact remains that language tests can be constructed without including such items. The choice of doing so should probably best be made after considering native-speaker data such as those we have just discussed. But a large part of such decisions would require a clearer understanding of the relationship between language and intelligence than we now have.

One other feature of the native-speaker study which bears comment at this point is the judgment of difficulty by the candidates themselves. At the time of the test, the subjects were asked to check those items which seemed unusually difficult or confusing. Also, they were asked to indicate their judgment of the difficulty of each section of the test on a 3 point scale with 1 being the most difficult. For the reading items which were the most difficult, the percent of candidates who indicated difficulty ranged from a low of 11.6 to as high as 22.2. On the test as a whole, the reading section was also considered to be the most difficult by the native speakers. Both from the point of view of the native speakers' judgments of difficulty and their actual performance on the test, reading was shown to be a problem area. At least in the Chomskyan sense, then, there seem to be factors beyond simple language proficiency which are playing a role here. We can only conclude that these factors, however they may eventually be described, form part of what we know as intelligence.

One final observation on the results of the native speaker study seems worthy of mention. The strong disparity between the high performance of the native speakers on the listening section of the test and the lower than expected performance on other sections, particularly reading, seems to support the distinction made by a number of researchers, most recently Krashen (1976), between acquired and learned skills. Since all normal persons acquire the ability to speak and understand their native language but must learn to read and write it, it is not an unusual occurence to find that they will perform almost to perfection in listening tests but with some difficulty in at least certain aspects or reading and writing. An interesting extension of this distinction would apply to teaching. It may be reasonable to assume that students can be taught those skills in the target language which even native speakers much learn viz., reading and writing. But the same degree of success may not be attainable for those skills which natives regularly acquire: speaking and understanding.

Another way in which we can approach the subject of language and intelligence testing is to look at actual tests of each type. Among the tests used with non-native speakers in an academic setting, two of the most known are the TOEFL test and the Aptitude test of the Graduate Record Examinations Board. The TOEFL test is designed as a measure of English proficiency and the GRE aptitude test is designed as a measure of general abilities necessary for graduate study. But because language

is the medium through which aptitude is evaluated on at least half of the GRE test (the test is composed of a verbal section and a quantitative section), students are often confused about this distinction. Moreover, admissions and other personnel frequently use scores on tests such as the GRE as indicators of a prospective student's language proficiency or scores on a test such as TOEFL as measures of future academic success.

In an attempt to get a clearer idea of how these two tests relate to each other, I have recently conducted a study at Texas A&M University in which I compared the results of scores which foreign students applying for admission made on both the TOEFL and the GRE Aptitude tests. As indicated in Fig. 1, the two tests, while not completely unrelated, do not show a very high correlation. On the basis of this study using scores of 91 students, the correlation coefficient for the total scores on both tests was .53. As might be expected, the correlation between the TOEFL scores and the GRE verbal portion of the test (.55) was higher than that between the TOEFL and the GRE quantitative scores (.31). Validation of this fact can be found among the many foreign students who are extremely proficient in mathematical skills despite a lack of proficiency in English.

Figure 1

Performance on TOEFL and GRE Aptitude

	TOEFL	GRE TOTAL	GRE VERBAL	GRE QUANTITATIVE
TOEFL	1.00	.532	.550	.311*
GRE TOTAL	.532	1.00	.629	.854
GRE VERBAL	.550	.629	1.00	.134*
GRE QUANTITATIVE	.311*	.854	.134*	1.00

The mean score for the 91 subjects on the TOEFL test was 544 (SD = 68.6), on the GRE verbal 338 (SD = 81), GRE quantitative 604 (SD = 121.4) and GRE total 942 (SD = 154.8).

*All correlation coefficients are significant at the .0001 level except those indicated by an asterisk. The .311 coefficient is significant at the .01 level but the .134 coefficient is not significant.

A scatter plot of these 91 pairs of scores gives a similar representation of how the two tests relate. But by dividing the graphs into quadrants to indicate high and low scores for each of the two tests an even clearer picture emerges (Fig. 2).

Figure 2

Comparison of High and Low Scores

A. TOEFL -- GRE Verbal

High TOEFL Low GRE 41	High TOEFL High GRE 4
Low TOEFL Low GRE 46	Low TOEFL High GRE 0

B. TOEFL -- GRE Total

High TOEFL Low GRE 24	High TOEFL High GRE 20
Low TOEFL Low GRE 31	Low TOEFL High GRE 16

When only the GRE verbal scores are compared with the TOEFL results there are no cases of persons with low langauge and high aptitude scores. And all but 4 of the 91 pairs of scores are in the low GRE range, almost evenly divided between those with high and low TOEFL scores. Increased language proficiency seems to have little effect on the GRE aptitude scores, at least for non-native speakers. When the quantitative scores are added a more linear pattern appears. Sixteen of the ninety-one pairs of scores appear in the low TOEFL-high GRE range. And twenty with high TOEFL scores also fall within the high GRE range. Again the low GRE scores are divided almost evenly between the two TOEFL groups. Clearly, the addition of the quantitative factor produces a much more balanced array of GRE scores. This is further confirmed by the .[illegible] correlation between the scores on the quantitative portion and the total GRE scores while only a .63 correlation was found between the scores on the verbal portion and the total GRE scores (Fig. 1).

If we look at tests such as the TOEFL and the GRE aptitude from the point of view of their content we find a number of differences which explain to a large extent the variation in the scores found in my study. These differences appear in three main areas: item types, quantity of material on which test items are based, and the time allowed for the candidate to take the test. As far as item types are concerned both tests are of the multiple choice variety. And some item such as those related to listening comprehension on the TOEFL do not apply to the GRE. But where they are more closely related i.e., in reading and vocabulary,

the tasks presented to the student are decidedly divergent. Vocabulary is tested on the TOEFL by asking the student to choose a synonym for a word in context. Reading is tested by asking the student to answer questions referring to passages ranging from 2 or 3 sentences to no more than about 150 words. Vocabulary is used in those items on the GRE which attempt to test relationships (related pairs of words) including opposites. Reading is tested on the GRE also by asking the candidate to answer questions referring to given passages. But the passages are much longer and much more complex.

My examination of one sample GRE aptitude test showed that it contained 6 consecutive reading passages averaging 420 words in length and 2 additional passages later in the test averaging 463 words in length. Also the type of questions asked in reference to the passages were primarily of the inferential type. A sample are the following:

1. The author is most concerned with answering which of the following questions?
2. The author's approach to his subject can best be described as . . .
3. The style and content of the passage suggest that it was excerpted from . . .
4. Which of the following words best expresses the underlying theme of the passage?
5. Which of the following is least consistent with a belief in determinism as the author defines the term?
6. The tone of the author's conclusions can best be described as one of . . .
7. With which of the following statements would the author most likely agree?

The time alloted to complete the items on the two tests is almost the same if we consider only the verbal portion of the GRE. That section of the test contains 130 items to be done in 100 minutes or 44 seconds per item. The Reading and Vocabulary section of the TOEFL contains 60 items to be done in 45 minutes or 45 seconds per item. But given the much longer passages and the increased complexity of the vocabulary and reading items on the GRE, time can easily be listed as a problem for the non-native speaker when taking the GRE aptitude test.

One possible problem with the Texas A&M study is the question of the time internal between the administration of the two tests. The maximum time span used in the study was 2 months. All of the 91 subjects had taken both the TOEFL and the GRE aptitude test within a period of no more than 2 months. Most likely this factor did not affect the results a great deal since 26 of the subjects had taken both tests in the same month. Of the remaining 65 subjects 39 had taken the tests one month apart (of these 24 had taken TOEFL first and 15 the GRE first) and only 26 had actually taken the tests as much as 2 months apart (18 had taken TOEFL first and 8 the GRE first).

The TOEFL research committee has recently approved a study similar to my own at Texas A&M but which will be more comprehensive both in terms of the number of subjects and the test used. Moreover, the possible problem of time should be eliminated by having subjects take the tests in question within a very short period. Specifically, the study will have as subjects 300 foreign graduate students who will take both the TOEFL and the GRE aptitude test and 300 foreign undergraduates who

will take the TOEFL and the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board. Also, the undergraduates will take the Test of Standard Written English (TSWE), a fairly new test designed to measure basic writing skills. Since this latter test is not designed for non-native speakers, the results achieved by such subjects will provide information about the usefulness of the TSWE for measuring proficiency in English as a second language. An additional feature of this study will be the attempt to give an evaluation of the appropriateness of tests such as the GRE and the SAT based on examination of the tests by selected ESL specialists.

This study as described will be carried out during the current spring and summer with a report of the findings available by late 1977. In general, the results of my study relating the TOEFL and GRE tests as well as the hypothesis behind the larger scale study to be carried out by ETS would seem to be that tests such as the GRE are inappropriate for non-native speakers. Furthermore, the nature of such tests as aptitude measures render them as inadequate devices for determining language proficiency, particularly in a second language context.

But what about the question of intelligence factors as measured by tests using the students' own language or tests which do not depend strictly on language? In a paper entitled "The Relationship Between Foreign Language Proficiency and Various Psychological Variables" Mats Oskarsson (1975) has reported on a project in Sweden which attempted to examine the relationship between second language proficiency (English) and various intelligence factors such as general verbal comprehension, ability to draw logical conclusions and ability to perceive spatial relations. Although the intelligence tests were conducted in Swedish, the native language of the students, their relationship with the language tests was not very great. The highest correlation coefficient achieved (.37) was between the sentence-combining portion of the English language test and the section of the intelligence test using opposites. All other coefficients were lower, some as low as .13.

In another portion of the same study reading comprehension skills were compared with the same battery of intelligence tests. Although the overall correlation coefficients are once again fairly low, it is interesting to note that one test, that which tested induction and logic, seemed to relate much more to reading proficiency than any of the others (.67). Oskarsson's conclusion, that "success on reading comprehension tests consisting of texts followed by multiple-choice questions is to some, if not a large, extent dependent on inductive and logical reasoning abilities" would seem to support some of our earlier comments with regard to the results of the TOEFL native speaker study.

The Oskarsson study also compared the English test results with those from a number of tests evaluating certain personality traits. Once again the correlation coefficients were low but the one interesting relationship is shown in the negative correlation (-.33) between the English language test scores and those on the test of deliberateness. This would seem to indicate that the person who is more deliberate or careful about using a second language will be least successful in doing so. In testing terms Oskarsson cites this as evidence that the person "who tends to ponder for too long over which option to pick in multiple choice items is more likely to go wrong than the person who acts on impulse."

In summary, then, we have seen how intelligence factors and language factors overlap in tests designed to measure language or intelligence. This is no doubt a reflection of the actual interplay of such factors within the human brain. But the problem comes in when we make judgments about language on the basis of test data which are primarily directed toward evaluating intelligence or vice-versa. Despite the overlap, the two are not the same. As Glucksberg and Danks (1975) have wisely stated, "conceptual competence cannot be inferred from linguistic competence alone."

Much more investigation is needed of the relationship between language and intelligence. In particular, what is needed is joint research on this question by linguists, psychologists and specialists in measurement. Given our present knowledge, however, we can only wonder whether from an administrator's point of view (admissions and placement in particular) the number of tests frequently required of non-native speakers helps or hinders the attempt to arrive at a meaningful description of either linguistic or intellectual competence. Likewise, from the point of view of the students who take these tests the purpose and scope of these various tests is not clear. Put more succintly, the title of this paper represents the problem--language testing and intelligence testing: friends or foes?