

# The Predictive Validity of Final English Exams as a Measure of Success in Iranian National University Entrance English Exam

Taher Alavi

Allame Tabatabaee University, Tehran, Iran

Email: neta.7529@yahoo.com

**Abstract**—This study is an attempt to investigate the probable relationship between English scores of high school and pre- university with the English scores of the learners in (INUEEE) Iranian National University Entrance English Exam. To find out the relationships, a sample of 42 students randomly were selected out of the whole population of Pre-university students in Naghadeh, West Azerbaijan. The students final English scores in grade three and pre-university 1 and 2 were collected from the school archives and all sets of scores were correlated with the students' scores on INUEEE. All the statistical analyses were done by SPSS. Results of the analyses showed that all the correlations were significant at (.01) level of significance. However, since we were indented to use these sets of (high school and Pre-University) scores for INUEEE score prediction, we failed to do so due to low correlation coefficients between the variables. Therefore, we can strongly state that there is no guarantee that those students who grade high in high school and pre-university English exams will grade high in INUEEE as well.

**Index Terms**—Iranian National University Entrance English Exam (INUEEE), predictive validity

## I. INTRODUCTION

In Iranian educational system, any university candidate should pass through a test battery which includes the selected courses that the learners have already covered at high school and pre-university level. Iranian Measurement Organization (IMO) has been in charge of preparing and administering tests for university candidates for several decades nationwide. This test battery namely, Iranian National University Entrance Exam (INUEE) has got a screening function to select the most appropriate students for the majors that they have applied. Among these selected courses, there is an English exam aimed at measuring the English proficiency of the candidates. This test is comprised of 25 multiple-choice items, each measuring a different sub-component of English proficiency (Vocabulary – Structure – Reading). These items cover the contents of two course books (English Book 3 and Pre-university Book 1-2). This study is an attempt to examine the predictive validity of these three final exam scores (grade three and pre-university 1 - 2) as a measure of success in the students later performance in (INUEEE).

### A. Description of Tests 'Contents

Normally, test constructors follow a fixed procedure in the process of test construction for (INUEE). In English test construction for (INUEE), they devote the first five items to structure, ten items to vocabulary, five items to reading skill and the last five items to cloze passage in multiple choice format. Since they have to construct tests based on the content of three books (English textbook 3, Pre- university English textbook 1 and 2), they have to consider the share of items assigned to each of the books mentioned above. Normally, 40 percent of the items are constructed from Book 3 and the rest 60 percent from Pre – university textbooks. In the final exams on Book 3 and Pre-university 1 and 2, different types of test items are employed (completion – multiple choice – matching – essay). The effect of guessing is considerably reduced in the final exams, since the number of multiple choice items is few.

### B. What is Validity?

A test is considered to be valid if it measures what it purports to measure according to Hughes (2003). The concept of validity can be approached from several dimensions, each of which displays a different perspective to the concept of validity.

Hughes (2003) categorizes validity under four main headings which are: Content validity, Construct validity, Face validity and Criterion-related validity (concurrent and predictive validity).

Content validity is related to how adequately a test covers representative behavior which it is concerned with (Hughes, 2003). Having a table of specifications of the course content can be helpful in judging the content validity of a test; however, everything in the table of specifications may not be expected to appear in the test. A test which reflects the major aspects in a table of specifications is likely to measure what it claims to measure; thus, have content validity (Hughes, 2003).

Bachman (1990, p. 255) defines construct validity as “the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs.” Alderson, Clapham and Wall (1995) mention different ways for establishing the construct validity of a test. Some of these are assessing whether the test is based on its underlying theory, internal correlations among different components of a test, multitrait-multimethod analysis, convergent-divergent validation and factor analysis.

Face validity that is not considered to be a scientific concept by many authors; however, is regarded as quite important. Bachman and Palmer (2000, p. 42) state that face validity is the appearance of validity, or the extent to which the test appeals to the test takers and test users. They also argue that the notion of 'test appeal' is essentially a function of authenticity and instructiveness. Poor items, unclear instructions or unrealistic time limits may negatively affect the face validity of a test (Alderson, Clapham & Wall, 1995).

Criterion-related validity, according to Fulcher and Davidson (2007, p. 4), demonstrates whether there is a relationship between a particular test and a criterion to which we wish to make predictions. They also argue that validity evidence is the strength of the predictive relationship between the test score and that performance on the criterion. There are two types of criterion-related validity: concurrent and predictive.

In concurrent validity the test scores are compared with another measure for the same examinees and both measures are administered at about the same time (Alderson, Clapham & Wall, 1995).

The second type of criterion-related validity, predictive validity according to Alderson, Clapham and Wall (1995), can be tested for the same examinees by comparing a test score with another measure, which is collected after the test has been given. It is common to look for predictive validity in a proficiency test because predictive validity analyses are important in checking whether the main objective of the proficiency exam, which is to evaluate an examinee's ability to successfully perform in a future course, is achieved or not.

Crocker and Algina (1986, p. 224) define predictive validity as “the degree to which test scores predict criterion measurements that will be made at some point in the future.” In terms of a proficiency test, predictive validity refers to the extent to which a test can be appropriately used to draw inferences regarding proficiency.

Brown (2004, p. 24) asserts that predictive validity of an assessment becomes important in the case of placement tests, admissions assessment batteries, language aptitude tests, and the like. He also argues that the assessment criterion in such cases is not to measure concurrent ability but to assess (and predict) a test-taker's likelihood of future success.

## II. REVIEW OF RELATED LITERATURE

### Studies on Predictive Validity

A study was conducted by Prapphal (1990) to find out the predictive validity of three sub-tests of the National English Entrance Examination in Thailand on academic achievement in Freshmen General English and English for Academic Purposes courses at two different universities. The study involved 264 randomly selected science students who had taken the National English Entrance Examination in Thailand in 1982. The results indicated that all three tests correlated significantly and substantially with university English achievement. Since the content of all three tests involved general English, the three tests account for more variance with the General English Course than with the English for Academic Purposes Course. Prapphal (1990) also suggested that test format can play an important role in predicting future academic achievement in English.

Another study conducted by Prapphal (1990) examined the relationship between the test of General English (GE) which aimed at assessing the students' ability in understanding general English and the English for the Academic Purposes Test (EAP) which had a more discipline specific content. The formats of both exams were the same. The study was conducted with 320 Chulalongkorn University students. Significant indirect relationships between the subskills of General English and English for Academic Purposes were found. This study suggests that all language subskills are related to one another, no matter what the format is. A transfer of subskills from one content (General English content) to another (English for Academic Purposes content) is possible.

A third study that Prapphal (1990) conducted involved 100 first year students. The study was conducted to find out to what extent did the EAP subtests, the EAP Department Test and the University English AB Entrance Examination which assessed the general proficiency predicted the academic achievement which is represented by GPA. The results showed that even if all the tests were able to predict the academic achievement, the EAP tests were more successful when compared with the General English Test. It is suggested that EAP tests may predict achievement in EAP programs more effectively than General English tests.

A study carried out by Stofflet, Fenton, and Strough (2001) examined the predictive validity of the Alaska State High School Graduation Qualifying Examination (HSGQE) and Benchmark Examinations on the performance on California Achievement Tests (CAT). The results showed a strong and direct relationship between performances on the Benchmark Test or HSGQE Reading scores and Writing scores and performances on the CAT Total Reading scores and Total Language and Arts scores, respectively.

Dooley (1999) carried out a study which aimed at answering the question of whether IELTS is an accurate predictor of performance and success in Business, Science and Engineering students. Business was particularly chosen as 'linguistically demanding' as opposed to Science and Engineering which was considered to be 'less linguistically demanding'. This provided the opportunity to compare students in different disciplines. The study was conducted on a

total of 89 students in their first years. The results indicated that the only consistently positive correlation was found between IELTS reading subset scores and success in business which was considered to be 'more linguistically demanding' discipline.

A study conducted by Educational Testing Unit researchers, Ramist Lewis and McCauley-Jenkins (2002) investigated the correlations between SAT II Subject Tests and freshmen GPA. The results pointed out that English composition had a correlation of .51 with freshmen GPA. This was the highest correlation among SAT II Subject Tests. French, German, Hebrew, Latin and Spanish sub-tests showed lower correlations with freshmen GPA.

Some studies on the relationship of TOEFL with other English Proficiency Tests have been cited in Marvin and Simner (1999) in order to justify the use of the TOEFL for decision making. In addition, according to Marvin and Simner (1999) a relationship between TOEFL scores and first year performance in university English courses can be possible; however, the relationship may not continue beyond first year. Pack (1972) (in Marvin & Simner, 1999) carried out a study on 402 students and found out that, TOEFL scores were "significantly related to the grade obtained in the first English course taken, however, they are not related to grades obtained in subsequent English courses nor are they related to the probability that an examinee will graduate" (Hale et al. p. 161) (in Marvin & Simner, 1999).

A study conducted by Huong (2001) investigated the predictive validity of IELTS scores. The relationship between IELTS scores and subsequent academic performance was examined. 202 Vietnamese students who studied in different Australian universities were participants. Huang found a significant and positive correlation between IELTS scores and first and second semester GPA's which was considered to be satisfactory by Alderson, Clapham and Wall (1995). Moreover, among the four sub-tests of IELTS (Listening, Reading, Writing and Speaking) the highest correlation was observed between Reading and first semester GPA; Reading and Listening subtests compared to Writing and Speaking had higher correlations with first and second semester GPA's. The findings suggested a correspondence with the first and second semester GPA in terms of both IELTS total scores and sub-test scores.

Breland, Kubota and Bonner (1999) carried out a study in order to examine the relationship between scores on the SAT II: Writing Subject Test and performance in writing in the first year of university. 222 students participated with all the required writing samples; however, more cases were available for some variables when compared to others. The results of the study revealed high correlations between SAT I Verbal score and university course grades. Also, a high correlation was achieved for SAT II: Writing Test. However, the SAT Writing Test Essay score had a lower correlation for predicting course grades when compared to SAT I Verbal score and SAT II : Writing Test.

In a study, Heard and Ayers (1988) examined the validity of the American College Test (ACT) in predicting success on the Pre-Professional Skills Test (PPST). PPST is designed to measure proficiency in reading, writing and mathematics. ACT, which is used for admission, consists of English, mathematics, natural science, social science subtests in addition to a composite test score. 202 students took part in the study. These students had taken the PPST as a requirement for admission to the Professional component of the teacher education program at Tennessee Technological University. The students had also completed the ACT. It was concluded that the ACT composite score was the best predictor of success on three tests of PPST. ACT composite scores, subtest scores together with GPA in college English courses improved the prediction of achievement. These results indicate that scores from ACT are a reasonable predictor of success on the PPST.

### III. METHODOLOGY

#### A. *Research Questions*

To investigate the predictive validity of these tests the following research questions were raised.

H1: Is there any relationship between the final English score in grade three of high school and INUEE's English subset?

H2: Is there any relationship between the final English score in pre-university(book 1) and INUEE's English subset?

H3: Is there any relationship between the final English score in pre-university (Book 2) and INUEE's English subset?

H4: Is there any relationship between the average scores of final English scores on (grade three of high school and pre-university 1 – 2) and INUEE's English subset?

#### B. *Subjects and Instruments*

The pool of subjects was 42 students at pre-university level .The subjects were randomly selected out of the whole population of pre-university students in Naghadeh. The sample consisted of students in different fields of study .The subjects were all male students in two classes who had already attended in all three exams in question. Their English scores in the previous years were collected from the school archive. Having collected all the students' scores in all the exams, we calculated reliability of the tests using KR-21 reliability formula.

#### C. *Data Analysis*

The English scores of the subjects in grade three were correlated with their English scores in (INUEEE).We correlated the subjects' English scores in Pre-university -1 with their English scores in (INUEEE).Later, the subjects' English scores in Pre-university -2 with their English scores in (INUEEE) were correlated. In the final step, we

calculated the mean of all three exams in (grade three – pre-university -1 and pre-university -2) and correlated it with the subject's English scores in (INUEEE). All the statistics analyses were done by SPSS.

IV. DISCUSSION

The correlation coefficient used in predictive validity studies is called a validity coefficient. Crocker and Algina (1986) define validity coefficient as a correlation coefficient between two variables: a test score and a criterion score. As an answer to the question of what an acceptable validity coefficient can be is answered by Hughes (1989) who says that a validity coefficient around 0.40 is the highest correlation expected in predictive validity studies. Cronbach (1990), states that a correlation as low as 0.30 may definitely have a practical value and correlations below that value may help improve decisions. The reason for having such low validity coefficients is because students who are below the cut scores in the test are not included in predictive validity studies. Still, Cronbach (1990) argues that test validities ranging from 0.30 to 0.50 contribute considerably despite the fact that they may wrongly predict many students. Whereas, validity coefficients around .40 can be accepted as sufficient with any suitable external measure, higher validity coefficients may be expected between the measure and the criterion if both are measuring similar traits. For example, an English proficiency exam is expected to yield a higher predictive validity coefficient across the grades obtained from a freshmen English course.

Mousavi (1999) argues that in discussing the coefficient of validity, one central issue is its magnitude. How high should a validity coefficient be? He also states that no general answer to this question is possible, since the interpretation of a validity coefficient must take into account a number of concomitant circumstances. The obtained correlation should be high enough to be statistically significant at some acceptable level , such as the .01 or .05 levels.

Hatch and Farhady (1982, p. 224) state that if the correlation between the two scores is low, our best guess about performance on the new test is the mean. Thus, as the correlation coefficients obtained are too low, using correlational analysis for predictive validity do not suit this particular case.

Jafarpur (1999, p. 143) provides a range for the magnitude of coefficient correlation arguing that any acceptable magnitude is dependent on the purpose of research. He states that if a sound decision is indented to be made, there should be a high correlation coefficient between the variables.

Considering the acceptable range for any correlation to be predictably valid, we can state that the magnitudes of two correlations (Pre 1 and 2) obtained are too low to be valid for predictive decisions and the magnitudes of the other two correlations are average and not enough for important predictive purposes. In the table of correlations, the highest correlation exists between grade three and INUEEE (r =.494) which is of an average value for correlation. The second highest correlation exists between student's average scores on the three final exams and INUEEE(r = .449), (average) which is exactly what we expected to obtain prior to this investigation. The third highest correlation exists between Pre-1 and INUEEE, (r = .343) (low) and the last correlation between Pre-2 and INUEEE (r = .329) (low) indicating that they should be ignored in this predictive analysis. Though all of this correlation is significant at (.05) and (.01) levels of significance, they are not able to efficiently predict the learner's success in INUEEE.

Coefficient of Determination

Henning (1987, p .69) argues that the correlation coefficient has another interesting property of use to us in Language test development .Since correlation is a measure of the relatedness of two variables, the square of the correlation coefficient tells us the proportion of variance in one variance accounted for by the other variable. Mousavi (1999, p. 45) states that coefficient of determination gives the percentage of variance in one variable that is associated with the variance in the other. You can think of a test score as being the result of many separate elements. Some of these elements are characteristics of the person taking the test .Some are characteristics of the situation in which the test is taken, and some are characteristics of the test itself. Based on what we mentioned above, we shall calculate the coefficient of determination for any of the correlations to see how much of the variables being correlated are the same and how much are different.

TABLE - 1

Variables	Correlation coefficient (r)	Coefficient of Determination 2 (r)
Three and INUEEE	.494	.244
Pre 1 and INUEEE	.343	.117
Pre 2 and INUEEE	.329	.108
Mean and INUEEE	.449	.101

As was illustrated in the table above, none of the Coefficient of Determinations are of great magnitudes, indicating that the variables in mind do not have much common variances.

V. CONCLUSION

The results of the study show that there is a positive relationship between each of the exams and INUEEE, separately and in combination. All the hypotheses that we had raised were confirmed in different levels of significance. However, as mentioned above, lack of high correlation between the variables do not allow the researchers to use the exams in

question for predictive purposes. So, Iranian English teachers who are teaching in high school and pre-university level should be aware that, the tests which they construct for final exams (high school and pre-university) do not have very high relationship with INUEEE, though the contents are the same. This study is not intended to dwell on the possible reasons of this low relationships. However among the possible reasons that can be investigated for further research are different ability levels that each of the exams is measuring or, lack of correlation between different forms of tests that are used in each of these exams.

#### REFERENCES

- [1] Alderson, J. C. and., Clapham, C. and Wall, D. (1995). *language test construction and Evaluation*. Cambridge: Cambridge University Press.
- [2] Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- [3] Bachman, L. F. and Palmer, A. (2000). *Language Testing in Practice*. Oxford: Oxford University Press.
- [4] Breland, H., M., Kubota, M., Y., Bonner, M., W. (1999). "The Performance Assessment study in Writing: Analysis of the SAT II: Writing Test." Retrieved in Jan 5, 2003 from [http://www.collegeboard.com/prod\\_downloads/about/news\\_info/cbsenior/yr2002/pdf/seventeen.pdf](http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2002/pdf/seventeen.pdf)
- [5] Brown, H. Douglas. (2004). *Language assessment Principles and Classroom Practices U.S.A*. White Plains, NY: Pearson Education.
- [6] Crocker, L., Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando: Holt, Rinehart and Winston, Inc.
- [7] Cronbach, L. J. (1990). *Essentials of Psychological Testing*. (5th ed.). New York: Harper Collins Publishers, Inc.
- [8] Dooney, P. (1999). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. In K. Martin, N.
- [9] Fulcher, G. and Davidson, F. (2007). *Language Testing and Assessment*. London: Routledge.
- [10] Green, S. B., Salkino, N. J., Akey, T. M. (1997). *Using SPSS for Windows*. (2nd ed.). New Jersey: Prentice Hall, Inc.
- [11] Hatch, E. and. Farhady. (1982). *Research Design and Statistics for Applied Linguistics*. Rowley, Massachusetts: Newbury House.
- [12] Heard, S. A., Ayers, J. B. (1988). "Validity of the American College Test in Predicting Success on the Pre-Professional Skills Test." *Educational and Psychological Measurement*. 48, pp.197-200.
- [13] Henning, G. (1987). *A Guide to Language Testing: Development, Evaluation, Research*. Rowley, Massachusetts: Newbury House.
- [14] Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- [15] Huong, T., T., T. (2001). "The Predictive Validity of the International English Language Testing System (IELTS) Test." *Postscript* 2, 1, 66-96. Retrieved from: [www.idp.edu.au/conference/conf2001/pres/28\\_1400\\_GlobCapEng/LngTst\\_IELTSschViet\\_pres\\_Huong.pdf](http://www.idp.edu.au/conference/conf2001/pres/28_1400_GlobCapEng/LngTst_IELTSschViet_pres_Huong.pdf)
- [16] Jafarpur, A. (1999). *Statistics in the Linguistic sciences*. Shiraz, Iran: Shiraz University Press.
- [17] Marvin, L., Simner, C. (1999). "Postscript to the Canadian Psychological Association's Position Statement on the TOEFL." Retrieved Dec 5, 2002 from [www.cpa.ca/documents/TOEFL.html](http://www.cpa.ca/documents/TOEFL.html).
- [18] Mousavi, S. A. (1999). *A Dictionary of Language Testing*. Tehran, Iran: Rahnama publications.
- [19] Prapphal, K. (1990). "The Relevance of Language Testing Research in the Planning of Language Programmes." Retrieved Jan 5, 2003 from <http://pioneer.netserv.chula.ac.th/~pkanchan/html/testres.htm>
- [20] "Predictive validity" Retrieved Oct 25, 2002 from <http://ericae.net/seltips.txt>
- [21] Ramist, L., Lewis, C., & McCauley-Jenkins, C. (2002). "Validity of the SAT II Science Tests." *Science Insights*, 6, 5. Retrieved Dec 5, 2002 from <http://www.nas.org>
- [22] Stofflet, F., Fenton, R., Strough, T. (2001, April). "Construct and Predictive Validity of the Alaska State High School Graduation Qualifying Examination: First Administration." A paper presented at the 2001 American Educational Research Association Convention. Retrieved Dec 5, 2002, from [http://www.asd.k12.ak.us/Depts/assess\\_eval/hsgqe/](http://www.asd.k12.ak.us/Depts/assess_eval/hsgqe/)

**Taher Alavi** was born in Nagade, West Azerbaijan, Iran (September 23, 1975). He has got M.A in TEFL/TESL from the university of Allame abatabaee, Tehran, Iran. He got his B.A in the same field from Shiraz University, Iran.

He has been teaching English in different local universities and language institutes since 2000. He has been offering different courses in the field of Language teaching methodology, and language testing for language students. At the present time he is the SUPERVISOR in his own language Institute, and teaching English for the advanced level students. So far he has developed a book which was published in Germany. *ON THE EFFECT OF THEMATIZATION ON THE COMPREHENSION OF SENTENCES WITH DIFFERENT VERB CATEGORIES AT INTERMEDIATE LEVEL*. (Germany, VDM, 2011)

Mr Alavi is the one of the leading members of AzarELTA (English language teachers association in West Azerbaijan, Iran). He is very interested in playing football in his free time.