

June 26, 2008

A RESEARCH REPORT on METU-EPE (2007/June)
(English Proficiency Exam)

***Prof. Dr. Hüsnu ENGİNARLAR**
SFL, Director

Contents

- 1.0 Introduction
- 2.0 Reliability indices
- 3.0 Validity issues
 - 3.1 Predictive validity
 - 3.2 Construct validity
- 4.0 Item performance & overall difficulty measures
- 5.0 Future Work & prospects

* I gratefully acknowledge Gökçen Baskan's and Fatma Ataman's -both DBE instructors- help with the statistical work.

1.0 Introduction

METU-EPE, prepared by a committee appointed by the Director of School of Foreign Languages, is administered 4 or 5 times a year to approximately 8.000 candidates. The Committee prepares 4 versions of the exam every year and measures are taken to ensure test equivalence.

Therefore, the results reported in this study on METU-EPE June, 2007 should be generalizable to different versions of the exam.

The exam is given in two stages, where:

Stage I Language Use & Reading (Grammar, reading and vocabulary embedded)

and

Stage II intends to test Listening Comprehension, Note-taking & Writing

2.0 Reliability indices

METU-EPE 2007

Stage I - Cronbach's alpha : .94

Stage II - (Listening) - Cronbach's alpha : .86

Note-taking & Writing sections, graded by SFL instructors had inter-scorer reliability coefficients of .90 and .85, respectively in the previous year, not calculated for June, 2007. The reliability figures reported above are quite satisfactory.

3.0 Validity issues

Of the several types of validity mentioned in the literature related, this study investigated the predictive and construct validity of METU-EPE.

3.1 Predictive validity

Predictive validity, a kind of criterion –related validity, is quite important for proficiency exams since it attempts to investigate the relationship between the performance of test-takers in a test and their success in future tasks, possibly related to that test.

To explore the predictive power of METU-EPE (June, 2007) two future criteria were considered: GPAs (Grand Point Average) of 2007-2008, Fall Term and

ENG 101 Grades in the same term. TABLE -1 below displays the correlation coefficients between the METU-EPE 2007 scores and the GPAs in the Fall Term, overall and by faculty.

*A minimum of 24.5 pts/50 is required to qualify for Stage II.

TABLE - 1

	June - EPE
GPA /overall	.444*
GPA / EDUC	.441
GPA / ARCH	.299
GPA / ARTS & Sc	.29
GPA / ADM	.39
CPA / ENG	.441

*Correlations are significant at the 0,01 level.

TABLE - 2 displays the relationships between the JUNE – EPE scores and the ENG 101 grades.

	EPE - June
ENG 101 / overall	.551*
ENG 101 / EDUC	.459
ENG 101 / ARCH	.629
ENG 101 / ADM	.508
ENG 101 / ARTS & Sc	.489
ENG 101 / ENG	.566

*Correlations are significant at the 0,01 level

The overall GPA-EPE correlation (.441) in Table – 1 can be considered moderately high. Understandably, the relationship is weaker in the Faculties of Architecture and Arts & Sciences. It should also be mentioned that GPA includes ENG 101 grades, which might have contributed to the magnitude of the correlation.

The correlation coefficients in TABLE – 2 are naturally higher, but perhaps not as high as expected. A similar study had revealed a relationship of r : .65 earlier between EPE and ENG 101. Apparently, a term grade in course like ENG 101, requires more than proficiency in English.

All these statistics attest to the predictive power of METU-EPE. A reference should be made in this context to two other findings in earlier studies. In one study, the predictive power of all the achievement exams given in the academic year 2006 -2007 in the Department of Basic English was investigated. The r between the Department of Basic English yearly averages and subsequent

GPA was relatively high: .435 (a little below .444); however, the relationship between the Department of English annual averages and the ENG 101 scores was lower: .45 (.55 in the case of EPE).

It is worth noting that one test alone – EPE – carries a predictive power equal to or higher than a multitude of exams and tests given during the year. It is also interesting to note that achievement exams given during the preparatory education in the DBE have a pretty high predictive power of future success in the freshman year.

The correspondences between the June EPE grades and ENG 101 grades (.551) and GPAs (.441) are moderately strong, but that between ENG 101 grades and GPAs is quite high: .642. It seems ENG 101 term grades are stronger predictors of academic success than EPE scores in the first term of the freshman year (N.B: ENG 101 is included in the GPA).

Another criterion – related study conducted in the previous years involved 60 preparatory year students taking TOEFL CBT and METU-EPE almost concurrently, 10 days apart. The overall correlation was above .70.

3.2 Construct validity

TABLE - 3 displays the intercorrelation coefficients among the components of the METU-EPE June, 2007.

TABLE - 3

	LANGUAGE	READING	LISTENING	WRITING
LANG	1.00	-	-	-
READ	*.412	1.00	-	-
LIST	.446	.62	1.00	-
WRIT	.343	.352	.45	1.00

*Correlations are significant at the 0.01 level.

TABLE - 4 shows the degree of correspondence between STAGE - 1 and STAGE - 2 and also the predictive power of these two parts of the exam with respect to ENG 101 and GPAs.

TABLE - 4

	ENG 101	GPA	STG 1
STG 1	*.486	.376	-
STG 2	.469	.321	.499

*Correlations are significant at the 0.01 level.

The values on TABLE - 3 are interesting in several aspects. First of all, that the highest correlation does not exceed .62 indicates that these components are measuring different traits. Otherwise, we would have almost perfect correlations and wouldn't need to include all of these components. Moderate correlations observed above are normal since there is, of course, a common ground among these components. The lowest correlations are observed in the case of Writing, another justification to include the Writing component.

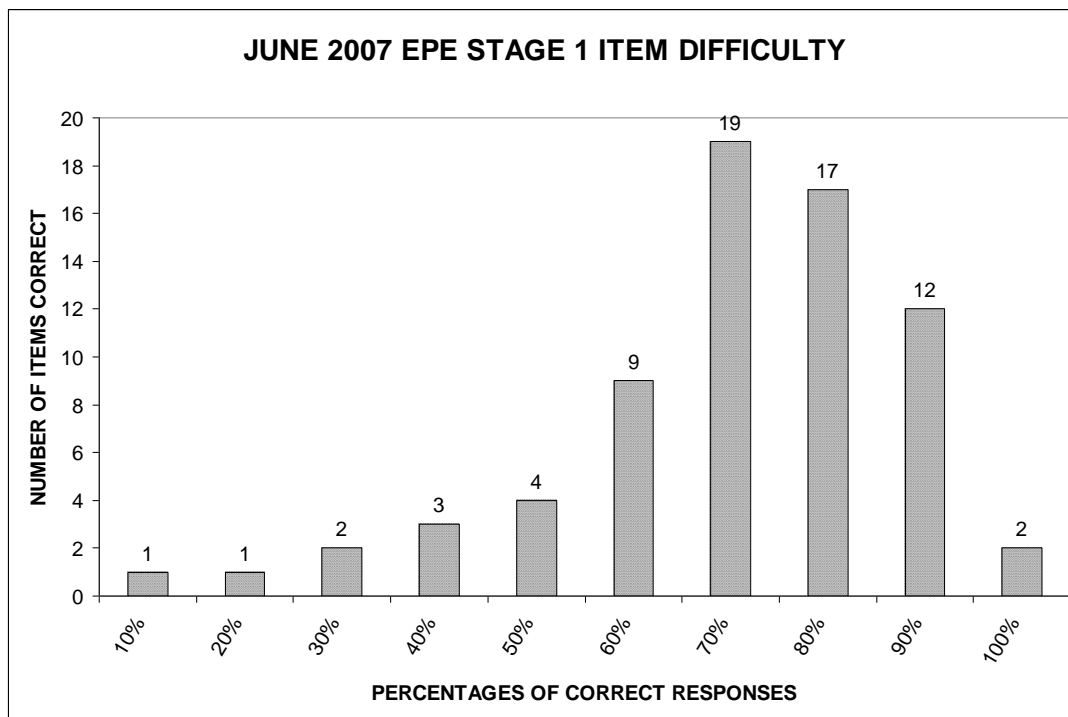
TABLE - 4 displays a similar relationship between STAGE - 1 and STAGE - 2 (r : .499), i.e., that these two parts are tapping different abilities and are thus essential.

Interestingly, the predictive power of STAGE - 1 is slightly higher, which implies Language Use & Reading Sections are to remain in the test for the foreseeable future.

4.0 Item Performance & Overall difficulty measures.

TABLE - 5 below displays the results of item difficulty level analysis with the number of questions answered correctly at every interval (10%) of percentage.

TABLE - 5 (STAGE - 1)

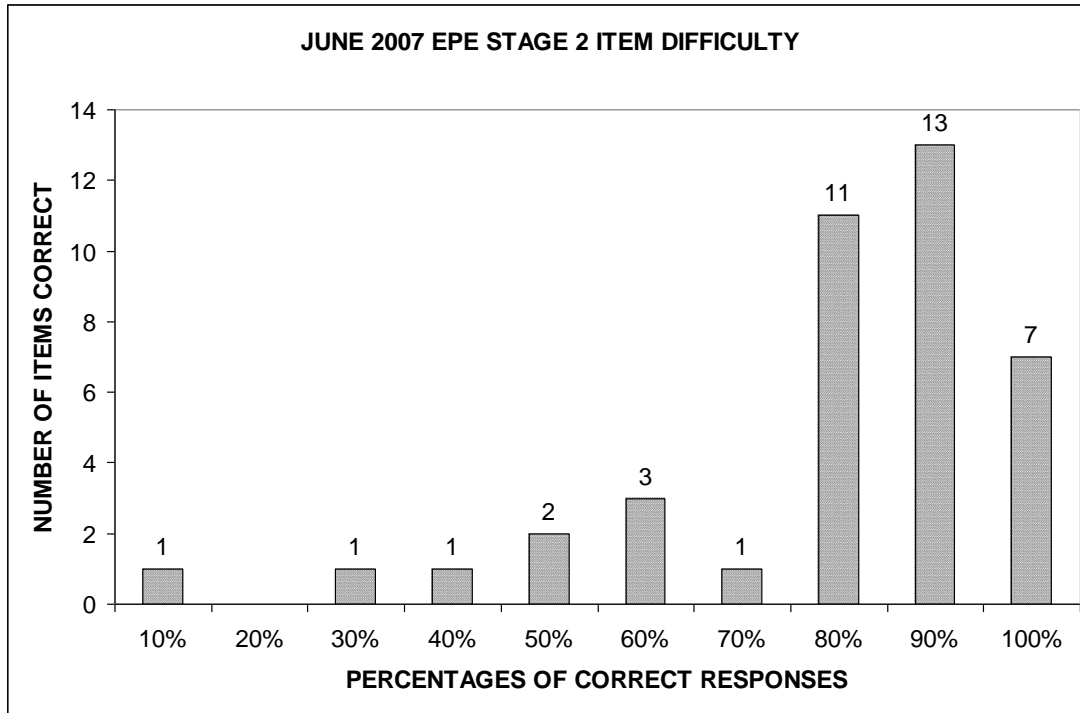


STAGE - 1 : Overall difficulty average: 0.645 (=65%)

STAGE - 1 : Overall discrimination index = 0.612

Finally, TABLE - 6 below exhibits the item performance values of the 40 – multiple – choice item Listening Component in STAGE – 2.

TABLE – 6



STAGE - 2 - Listening / overall difficulty level: 0.739 (%74)

STAGE - 2 - Listening / overall discrimination index : 0.61

The overall difficulty indices (%65 & %74) are quite satisfactory (acceptable range of difficulty : .30 - .70). It may be argued that LISTENING is on the easy side with an average difficulty index of %74; however, it must be noted that only those who pass STAGE – 1 take this component (usually, around 20% get eliminated in STAGE – 1). Only a few items are both extremities in TABLE – 5 need radical revision. Item discrimination indices are also very high. The overall item performance values, which seem to be quite good must have contributed considerably to the reliability and validity of the test.

5.0 Future work & prospects

Test construction and development is a dynamic process. Although METU-EPE seems to be a well functioning instrument in many ways, improvement efforts should continue. As far as content and construct validity go, we need to constantly compare EPE with the goals and objectives of the overall SFL curriculum and current developments in language acquisition. The Listening Component, both the multiple-choice part and the Note-taking

Section, also needs more development work to increase the degree of authenticity and validity. The Writing Section is essential and yet may be considered inadequate in some ways. Asking test-takers to write a paragraph of about 200 words is a compromise between needs and resources. Increasing the volume of writing would place severe constraints on the practicality of EPE. Any type of improvement is subject to means and resources. However, EPE should not exercise a negative impact on departmental syllabuses. For instance, there should be no reason why some groups cannot engage in essay writing as well. Test oriented teaching should be a concern only in the final part of the DBE syllabus.

We all know that METU-EPE is ‘incomplete’ in terms of direct skills coverage, i.e., speaking is not yet being tested directly. This is largely dependent on two factors: intensity of demand and motivation on one hand and accumulation of expertise within SFL and solution of issues related to practicality on the other. Departments, both DBE & DML are increasingly becoming learning organizations regarding the teaching and testing of speaking. This is a promising trend. However, practicality constraints weigh heavy against the incorporation of a speaking test for all. Studies in the near future might focus on pilot testing of speaking for some groups of students (for instance, for certain social science students and some groups of graduate students who need speaking skills more than the others.) But SFL needs more resources and flexible budgetary means for large scale speaking tests.

Another possible area of work is to investigate the discriminatory power of METU-EPE for students in different disciplines. METU-EPE seems to be working well with cut-offs of 59.5 and 64.5 for undergraduate and graduate students, respectively. But we don’t know how well it discriminates at higher intervals of proficiency. Interested departments and SFL can undertake joint studies to determine how sensitive METU-EPE is at distinguishing between different levels of proficiency for different fields of study.