

An Evaluation of Language Subtests of the Entrance Examinations at the Ph.D. Level

Parviz Birjandi *

Roxana Aminzadeh*

Abstract

The Ph.D. entrance examinations of Islamic Azad University in Iran are administered annually to select the best applicants for Ph.D. programs. In consideration of the important decisions, which are made on the basis of the results of these examinations, the present study was conducted to provide a qualitative evaluation of field-specific language subtests (FLSs) of four entrance examinations administered in 2003. In order to accomplish the objectives of the study, the views of four different groups of participants -- field-specialists, Ph.D. students, language testers, and applied linguists -- were obtained with respect to Ph.D. students' general-academic language needs, suitability of the current FLSs of Ph.D. entrance examinations of seventeen fields of Humanities with regard to their construction, use, and scoring procedures, and the desirable characteristics of a new general-academic English subtest. Furthermore, this study involved the qualitative analysis of the content and structure of the items appearing on FLSs of Ph.D. entrance examinations of four fields of Humanities. The results indicated that FLSs were not consistent with Ph.D. students' general-academic language needs. Moreover, there was no commonly agreed code of practice or set of standards, which was

* Allameh Tabataba'i University.

• Islamic Azad University, North Tehran.

followed by examination boards in the processes of test construction, use, and scoring. The results further indicated the need for developing a standardized Performance-based General Academic (PGAE) test battery to be administered alongside existing FLSs in Ph.D. entrance examinations. This need could be due to the fact that current tests alone could not cover the full range of general academic language ability required in Ph.D. programs and thus were not appropriate measures for admissions decisions.

Key Terms: Evaluation, General Academic Language Ability, General Academic Language needs, Performance-based Language Test, Target Language Use (TLU) Task, Test Task.

Introduction

It is a truism of educational and psychological measurement that constructing mental measures is extremely difficult. Usually it is only after considerable effort by item writers, examination of content by experts, piloting, and statistical analysis of items that a test constructor can feel that a test is likely to measure what it was designed to measure, and even then there is no guarantee that other researchers will feel equally satisfied.

Despite these considerable problems, tests are often used in educational settings to make important decisions: Determining levels of academic achievement, suitability for further education, or fitness for entry into desirable professions are typical examples. Language tests, too, are valuable tools for providing information that is relevant to making decisions about individuals, such as determining whether

individuals have the adequate level of language ability necessary for handling academic tasks in the course of their education. The importance of such decisions in determining the future lives of testees places an awesome responsibility on test makers, and it is to be expected that they would have developed a code of practice to ensure that their work conforms to the highest standards possible.

There are a number of such codes of practice published, probably the most well-known of which is the American Psychological Association (APA) Standards (1999). These standards provide specific criteria for the evaluation of tests, testing practices, and the effects of test use. Observing these standards allows test developers to construct measurement devices, which represent data of high credibility.

Moreover, in recent years, there has been increasing agreement among testers that language proficiency tests should, where possible, be related to candidates' future language needs (Bachman & Palmer, 1996). For example, if doctors are to be tested for their capacity to use English in an English-speaking hospital, it is considered only sensible to test them on the kinds of English that are used in the ward and the consulting room. No one will question the good sense of such English for specific purposes (ESP) testing. The issue, however, becomes less clear-cut when the language test is aimed at a less sharply defined audience, such as students proposing to embark on graduate studies at different disciplines. Many testers consider that a language test for such students should contain samples of the kinds of language tasks required of them in their academic work (Clapham, 1996). Thus, the major consideration in developing such language tests would be an indication of common general-academic language tasks among

students of different disciplines. This provides a shift from traditional language testing to performance assessment, where learners have to demonstrate their practical command of language.

Ph.D. entrance examinations of Islamic Azad University in Iran are administered annually to select the best applicants for Ph.D. programs. In consideration of the important decisions, which are made on the basis of the results of these examinations, the present study was conducted to provide a qualitative evaluation of field-specific language subtests of four examinations (Business Management, Educational Management, State Management, and Strategic Management) administered in 2003.

The main objectives of the study were as follows:

1. To specify Ph.D. students' general-academic language needs in the course of their education;
2. To examine whether current field-specific language subtests (FLSs) are constructed on the basis of those specified general-academic language needs;
3. To identify whether FLSs conform to the American Psychological Association (APA) Standards in the processes of their construction, use, and scoring; and
4. To investigate the need for constructing a standardized Performance-based General Academic English (PGAE) subtest to be administered, alongside FLSs at Ph.D. entrance examinations;

In pursuit of these objectives, this study undertook to provide empirical evidence for the following research questions:

1. Is there any correspondence between the items of field-specific language subtests (FLSs) and Ph.D. students' general-academic language needs?
2. Do FLSs conform to APA Standards in their construction, use, and scoring?
3. Is there any need for developing a standardized Performance-based General Academic English (PGAE) subtest to be administered alongside FLSs at Ph.D. entrance examinations?

Method

Subjects

Four groups of subjects were involved in this study:

1. Twenty-five male and female university instructors and heads of departments of Ph.D. programs at seventeen fields of Humanities (History, Accounting, Psychology, Sociology, Geography, International Relations, Educational Management, Business Management, State Management, Strategic Management, Journalism, Economics, Consultation, Political Science, Physical Education, International Law, and Social Communication Science) at IAU, Science and Research Campus;
2. Sixty-five male and female Ph.D. students, majoring in the aforementioned seventeen fields of Humanities at IAU, Science and Research Campus;
3. Sixteen male and female language testers working at different branches of IAU; and

4. Eleven male and female applied linguists working in a variety of fields at different branches of IAU;

The subjects were all native speakers of Farsi and were selected through “purposeful sampling” (patton, 1990, p. 169).

Instrumentation

Two groups of instruments were utilized in this study. The initial group consisted of four measures.

The first measure (Questionnaire 1) was a detailed Persian questionnaire consisting of thirty closed and open format questions. This questionnaire was administered to a number of university instructors and heads of departments at Ph.D. programs of seventeen fields of Humanities at Islamic Azad University (IAU).

The second measure (Questionnaire 2) was another Persian questionnaire, which consisted of nineteen closed and open format questions. It was intended to elicit information from Ph.D. students, majoring in seventeen fields of Humanities at IAU.

The third measure (Questionnaire 3) was a detailed English questionnaire, consisting of sixteen closed and open format questions. This questionnaire was administered to a number of language testers who had experience in teaching ESP courses at different branches of IAU.

The fourth measure was a letter intended to seek the views of a number of applied linguists on theoretical issues of language and language testing.

It is worth mentioning that the three questionnaires utilized in this study were developed following the five stages proposed by Gillham (2002). These stages were as follows: (1) Defining the

objectives of the survey; (2) Pre-piloting (by semi-structured interview); (3) Drafting the items; (4) piloting and validating the questions (following the think-aloud technique proposed by Alderson and Crawshaw, 1990); and (5) Administering.

The next groups of instruments used in this research were exact copies of the field-specific language subtests (FLSs) of Ph.D. entrance examinations of four fields of Humanities at IAU, administered in 2003. These tests were deliberately chosen for quality control purposes since they had been used for admissions decisions.

Procedure

The present research was conducted in five stages. Initially, Questionnaire 1 was administered to a number of university instructors and heads of departments at Ph.D. programs of seventeen fields of Humanities. The intention was to elicit information on Ph.D. students' general-academic language needs, the suitability of the current field-specific language subtests (FLSs) of Ph.D. entrance examinations with regard to their construction, use, and scoring, and the desirable features of a new general-academic English subtest to be administered alongside current FLSs.

At the second stage, Questionnaire 2 was administered to a group of Ph.D. students, majoring in seventeen fields of Humanities, with the major aim of determining their general-academic language needs in the course of their education. Furthermore, the respondents' views concerning the suitability of current FLSs of Ph.D. entrance examinations and the need for developing a general-academic English subtest were sought.

At the third stage of the study, Questionnaire 3 was administered to a number of language testers who had experience in teaching ESP courses at universities. At this stage, it was intended to obtain information on the appropriacy of current FLSs of Ph.D. entrance examinations. Moreover, language testers were asked to make recommendations on the structure and content of a desirable general-academic English test battery.

At the fourth stage, an attempt was made to seek the views of a number of applied linguists on the model of language proficiency, which the general-academic English subtest of Ph.D. entrance examinations at different fields of Humanities had to operationalize.

At the fifth stage of the study, it was tried to obtain the exact copies of field-specific language subtests (FLSs) of Ph.D. entrance examinations of four fields of Humanities (Business Management, Educational Management, State Management, Strategic Management) at Islamic Azad University, administered in 2003. Due to the cooperation of the authorities, these measures were obtained easily. The items included in the tests were qualitatively evaluated with respect to the degree of observing APA Standards in their construction, use, and scoring. Moreover, all the items on the four measures were inspected in terms of their degree of correspondence with Ph.D. students' general-academic language needs.

In order to investigate the degree of conformity of test items with TLU tasks (Ph.D. students' general-academic language needs), a comparison was conducted between the tables of specifications of these four tests and Ph.D. students' TLU tasks and language areas required in Ph.D. programs.

Results and Conclusions

At the end of the fourth stage of the study, all the views collected from four groups of participants were weighed up.

Based on the analysis of the responses of the three questionnaires, the general-academic language needs (TLU tasks) which were common among Ph.D. students majoring in seventeen fields of Humanities were identified. The specified TLU tasks are presented in the order of importance (mainly based on Ph.D. students' views) in the following table:

Table 1. Ph.D. Students' TLU Tasks

TLU Tasks	Degree of Importance (%)		
	Ph.D. students' views	Ph.D. instructors' views	Language testers' views
a. studying new sources in the field	70.0	94.1	72.7
b. finding new pieces of information in Internet	62.0	82.4	27.3
c. studying international field-specific journals	62.0	76.5	63.6
d. translating field-specific texts (English – Persian)	60.0	66.6	27.3
e. conducting research for Ph.D. dissertation	54.0	68.8	45.5
f. attending international seminars	54.0	64.7	36.4
g. writing and/or translating articles for international field-specific journals	52.0	64.7	63.6
h. writing the abstract of Ph.D. dissertation in English	50.0	76.5	45.5
i. summarizing English texts in English	50.0	76.5	54.5
j. taking notes of English texts	---	---	63.6

k. having the ability to link ideas appeared in different English sources	---	---	63.6
l. extracting relevant sections of English materials	---	---	45.5
m. having the ability to organize ideas in English	---	---	45.5
n. evaluating English textbooks for possible importance	---	---	36.4
o. making contact with international scientific and academic centers	45.7	33.3	---
p. interacting with foreign specialists in Iran	11.4	---	---

Moreover, those areas of language ability, knowledge of which was acknowledged to be important in improving the quality of Ph.D. programs, are presented in the order of importance in Table 2 below:

Table 2. Language Areas Required in Ph.D. Programs

Language Areas	Degree of Importance (%)		
	Ph.D. students' views	Ph.D. instructors' views	Language testers' views

	86.0	100.0	100.0
a. Reading comprehension	78.0	100.0	90.9
b. Field-specific vocabulary	84.0	88.2	72.7
c. Translation ability (English – Persian)	50.0	64.7	63.6
d. Writing			
e. Grammar	30.0	41.2	63.6
f. Oral interaction	54.0	41.2	45.5
g. General vocabulary			
h. Listening comprehension	14.0	35.3	90.9
i. Translation ability (Persian – English)	50.0	35.3	45.5
	34.0	47.1	27.3

The analysis of the results further indicated that field-specific language subtests of Ph.D. entrance examinations were not consistent with Ph.D. students' general-academic language needs. In fact, no needs analysis was conducted prior to test development contrary to what has been repeatedly mentioned in the literature that needs analysis should be central to ESP testing.

Moreover, it was indicated that these tests did not conform to standards of testing in the processes of their construction, use, and scoring. The test items had mostly essay-type format, however, the scripts were seldom double-marked and there was no indication of the existence of predetermined scoring criteria. The researcher's inspection of APA Standards (1999) revealed that these facts were against the standards proposed in this document. Furthermore, no annual revision was conducted on these tests despite general dissatisfaction. The tests were not piloted before the main administration and thus, no statistical analyses were applied on the

results in order to determine the statistical properties of items and to standardize the tests. The absence of piloting is a cause of considerable concern, since piloting can provide corroboration of the value of the judgment of item writers. Furthermore, no serious procedures were followed to estimate the reliability and to ensure the validity of results in spite of the high-stakes nature of these tests. This is alarming since not much trust can be put in the results of a test whose reliability and validity are not guaranteed. These facts revealed that the related APA Standards were not observed.

The results of the analysis of data also indicated the need for developing a general-academic English subtest to be administered alongside existing FLSs in Ph.D. entrance examinations (PEEs). This could be due to the fact that current tests alone could not cover the full range of English proficiency required in Ph.D. programs and thus were not appropriate measures for selection purposes. It was further suggested that due to the high-stakes nature of this test, it should be standardized and should be developed in accordance with Ph.D. students' general-academic language needs.

These results are presented in the following table:

Table 3. Respondents' Views with Respect to the Need for Developing a General-Academic English Subtest

Statements	Respondents' Degree of Agreement (%)			
	Ph.D. instructors' views	Ph.D. students' views	Language testers' views	Applied Linguists' views
a. The need for developing a general English subtest to be administered alongside current field-specific subtests at Ph.D. entrance examinations	100.0	93.6	81.8	---
b. The need for standardizing the general English subtest employing an accepted set of standards in testing	93.3	89.4	81.8	---
c. The need for constructing the general English subtest in accordance with Ph.D. students' general-academic language needs (TLU tasks)	86.7	93.5	81.8	100.0

Considering the content of the general-academic English subtest it was acknowledged by most language testers and applied linguists that Ph.D. students majoring in different fields of Humanities need to be proficient in four language skills (reading, writing, speaking, and listening). However, with respect to practical limitations of Ph.D. entrance examinations and the views of university instructors and Ph.D. students themselves, it was concluded that listening and

speaking components should not be included in this test. It was further decided not to include lexis and grammar components and thus to measure these two language areas indirectly. Furthermore, due to the important role of translation ability (English – Persian) in Ph.D. programs, it was decided to include this component in the test.

On the whole, with regard to the results of the needs analysis conducted at the qualitative phase of the study (Tables 1 & 2), it was concluded that the general-academic English subtest of PEEs should contain the following components:

- (i) Reading
- (ii) Translation (English – Persian)
- (iii) Writing

The fifth stage of the study was carried out to yield information with respect to the nature of the language tests, which were used for selecting the best applicants for Ph.D. programs. To achieve this purpose, four field-specific language subtests (FLSs) of Ph.D. entrance examinations, which were administered by Islamic Azad University in 2003, were studied qualitatively.

In order to answer the three questions of the study, the tests were inspected with regard to their degree of conformity to Ph.D. students' general-academic language needs and standards of test construction (APA Standards, 1999).

A comparison was conducted between the tables of specifications of these four tests (Table 4) and Ph.D. students' TLU tasks (Table 1) and language areas required in Ph.D. programs (Table 2). Moreover, all the items in the four tests were evaluated in terms of item writing standards. Besides, they were studied in terms of

naturalness. In order to make sure that the items represented examples of good English, two American native speakers were asked to judge the authenticity of the items.

The number and percentage of items devoted to each language area in the four field-specific language subtests (FLSs) are presented in Table 4 below.

Table 4. The Number and Percentage of Items Related to Each Language Area in FLS₁, FLS₂, FLS₃, and FLS₄

Test	Read.	Trans. (Eng. -- Per.) Total	Vocab.	Writ.
FLS ₁	--- 42 --- 100%	42 100%	--- ---	--- ---
FLS ₂	--- 68 --- 100%	40 58.8%	28 41.2%	--- ---
FLS ₃	5 8 62.5% 100%	3 37.5%	--- ---	--- ---

	5	7	---	---
FLS₄	12			
	41.7%	58.3%	---	---
	100%			

FLS₁ = field-specific language subtest (Business Management)

FLS₂ = field-specific language subtest (Educational Management)

FLS₃ = field-specific language subtest (State Management)

FLS₄ = field-specific language subtest (Strategic Management)

Read. = reading comprehension

Trans. (Eng. – Per.) = translation (English – Persian)

Vocab. = vocabulary

Writ. = writing

The results of the qualitative evaluation of these four tests testified the lack of harmony between their content and Ph.D. students' general-academic language needs. Moreover, none of the items of these four tests resembled real-world language tasks accomplished by Ph.D. students in the course of their education. As was mentioned earlier, the analysis of the responses to some of the questions of Questionnaires 1, 2, and 3 indicated the same results.

Besides, the defects mentioned in relation to the item writing process, indicated the lack of quality control procedures after drafting and then typing the test. Moreover, it was revealed that no particular standards had been observed in the construction of these tests. Most of the reading comprehension items were either unnatural or ungrammatical in terms of meaning or structure. In other words, the

items had not been appropriately developed to satisfy the requirements of such important tests and they included serious defects in terms of authenticity and grammaticality in the broadest sense of the word.

In general, these outcomes and the results of the analysis of Questionnaires 1, 2, and 3, discussed earlier, revealed that these field-specific language subtests did not conform to any standards (including APA Standards) in their construction, use, and scoring.

Furthermore, with respect to all the results of the study, it was evident that current non-standard field-specific language subtests of Ph.D. entrance examinations did not cover all areas of general-academic language ability required for Ph.D. students and thus failed to be appropriate measures for admissions decisions. This fact indicated the need for constructing a standardized Performance-based General Academic English subtest to be administered alongside current FLSs. It should be noted, however, that due to the high-stakes nature of these tests, field-specific language tests should also be developed following standards of testing and candidates' language needs. In this way, both the society and the individuals will enjoy the desirable outcomes in the long run.

Implications

It seems that there is reason to fear that field-specific language subtests of Ph.D. entrance examinations of some fields of Humanities at Islamic Azad University are not constructed in accordance with examinees' general-academic language needs. The results of this study revealed that no needs analysis has ever been conducted prior to the development of such high-stakes tests. Moreover, these tests do not cover the full range of general-academic language ability required for students in the course of their study. Thus, due to the influence of

students' high command of language ability on improving the quality of Ph.D. programs, it is essential that examination boards analyze candidates' language needs and identify those language areas which are mostly required for students in the course of their education. This can be best achieved through obtaining the views of different parties, such as field-specialists, Ph.D. students, language testers, and applied linguists. The specified language needs can further be utilized as a basis for constructing the items/tasks of a general academic test that may be used as the general English subtest of entrance examinations. This can be the first step for shifting from traditional test formats to performance-based assessment where examinees have to demonstrate their practical command of language.

The results of this study further indicated that no accepted standards are followed in the processes of test construction, use, and scoring. It is argued that the time is ripe for the development of a set of standards examination boards should follow, and for a public discussion of those procedures that would be most appropriate to ensure that standards were met.

As mentioned earlier, perhaps the best-known set of standards is the Standards for Educational and Psychological Testing, which is also referred to as the APA Standards. These standards were published in 1999 by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

However, the implication here is not to blindly follow these standards. As emphasized in the APA Standards document, the standards should be exercised with flexibility and caution. Where practical grounds are lacking, some compromises have to be made

with regard to the degree of observing the standards in the processes of test construction, use, and scoring.

Due to the constant need to produce new examinations, Ph.D. entrance examinations are rarely, if ever, tried out on students or subjected to the statistical analysis of typical test production procedures. In this regard Standard 3.1. of the APA Standards document (1999) states:

Tests and testing programs should be developed on a sound scientific basis. Test developers should compile the evidence bearing on a test, decide which information is needed prior to test publication or distribution and which information can be provided later, and conduct any needed research. (p.25)

It goes without saying that piloting is costly in terms of the required time, energy, thought, and money on the part of test developers, test administrators, and subjects. However, the advantages of piloting significantly override the disadvantages of administering a poorly constructed test for making important decisions with respect to the future lives of people. Pilot studies can provide authorities with invaluable information regarding the quality of items, timing of tests, clarity of instructions, and the statistical properties of tests.

Test developers have a responsibility to provide evidence related to the reliability and validity of their tests, as well as manuals and norms, when appropriate, to guide proper interpretation. Piloting is a helpful device in this regard too, since it enables the test developers to explore the reliability and validity of individual items as well as the reliability and validity of the total test by yielding the necessary data for statistical analyses.

Another important implication of this study is that the significance of carrying out post hoc analyses of test results is no less than conducting pilot studies. Sometimes, those defects, which have survived the analysis carried out during the piloting process, are detected during post hoc analyses.

It is recommended that tests be monitored on a routine and regular basis. Item and subtest analyses should be conducted after each administration and descriptive statistics should be calculated--including the reliability indices. Post-test reports ought to contain information that will enable a test to be modified in the future if necessary.

The results of this study also indicated that a very limited time is devoted to language instructions in some fields of Humanities at university. However, it seems necessary to increase the amount of time and attention paid to foreign language instruction at university levels since in the modern world of today the knowledge of a foreign language is not a luxury anymore; it is a necessity which enables the individuals to keep abreast of the most recent achievements in science and technology.

It is further recommended that material developers incorporate teaching materials and exercises which may help students to enhance their general academic language ability and language instructors devote more energy and time to language areas which are most required for students in the course of their education.

To sum up, the present research was intended to provide some information with regard to the qualitative characteristics of the field-specific language subtests of Ph.D. entrance examinations of four fields of Humanities at Islamic Azad University. It was further

intended to identify the general-academic language needs of Ph.D. students majoring in seventeen fields of Humanities. While the efforts of authorities in examination boards for constructing tests is appreciated and while it is acknowledged that developing desirable tests is a cumbersome task, it is emphasized that much work needs to be done in this regard in our country. It is hoped that this study would be a starting point for evaluating all those tests, which are administered for making important decisions.

Bibliography

Alderson, J. C., & Clapham, C. (1992). *Examining the ELTS test: An account of the first stage of the ELTS revision project* (Research Report 2). The British Council, the University of Cambridge Local Examinations Syndicate, and the International Development Program of Australian Universities and Colleges.

Alderson, J. C., & Crawshaw (1990). In Alderson, J. C. *Validating questionnaires*. Lancaster University, Center for Research in Language Education.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: American Psychological Association.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Clapham, C. M. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.

Davies, A. (1981). Review of Munby, J., Communicative syllabus design. *TESOL Quarterly*, 15 (3), 332-6.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.

Gillham, B. (2002). *Developing a questionnaire*. London and New York: Continuum.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.

Griffie, D. T. (1997). Questionnaire construction and classroom research [The Language Teacher On-line 23.01].

Jones, R. L. (1985). Second language performance testing: An overview. In Hauptman, P. C., LeBlanc, R., & Wesche, M. B. (eds.). *Second language performance testing*. Ottawa: University of Ottawa Press.

Lynch, B. (1996). *Language program evaluation*. USA: Cambridge University Press.

Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. California: Sage Publications.

McNamara, T. F. (1996). *Measuring second language performance: Applied linguistics and language study*. New York: Addison Wesley Longman Limited.

Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods (p.169)*. Newbury Park: Sage.

Weir, C. (1983). *Identifying the language needs of overseas students in tertiary education in the United Kingdom*. Unpublished Ph.D. dissertation, University of London.

