

# An introspective study on the construct validity of GRE tests: A plea for more qualitative research in language assessment

Tavakoli, Mansoor 

University of Isfahan, Iran ([Mr.tavakoli14@gmail.com](mailto:Mr.tavakoli14@gmail.com))

Farshid, Majid

University of Isfahan, Iran ([Farshidmajid@yahoo.com](mailto:Farshidmajid@yahoo.com))

Rezazadeh, Mohsen

University of Isfahan, Iran ([MIrezazadeh@gmail.com](mailto:MIrezazadeh@gmail.com))



ISSN: 2243-7738  
Online ISSN: 2243-7746

OPEN ACCESS

**Received:** 2 October 2012

**Available Online:** 11 March 2013

**Revised:** 13 February 2013

**DOI:** 10.5861/ijrset.2013.351

**Accepted:** 14 February 2013

## ***Abstract***

This study attempted to explore the notion of construct validity of Graduate Record Exam (GRE) tests (especially when it is used in some Iranian universities as a high-stakes test) taking a qualitative perspective. High-stakes assessment is any assessment whose outcome has life-changing implications for the test taker (Rover, 2001). To do so, five university instructors of English, all having post-graduate degree, were selected to take a GRE test. From different parts of the GRE test the analogy and the quantitative as well as the vocabulary parts of the GRE test were chosen for the experiment. It is because these parts, it seems, render GRE tests inappropriate in a situation like that of Iran. The data were then gathered via think-aloud technique, an introspection method. The study revealed some insights about the construct validity of GRT tests from the test takers' perspectives. The subjects unanimously reported that the mathematics questions in the quantitative part had little or nothing to do with the notion of English proficiency. In other words, they believed that GRE tests were not a good indicator of one's English knowledge. The participants also reported they had problems answering the analogy part of the test for a good number of reasons. This was partly had to do with the fact that the analogy questions needed more time to be solved and partly with the fact that they were more of intelligence questions than English proficiency items. In the end, it was concluded that GRE tests suffer from both construct under-representation and construct-irrelevant problems. Consequently, GRE tests should not be used as a high-stakes test in Iranian universities.

**Keywords:** construct validity; graduate record exam; introspection techniques; qualitative paradigm; validation; think-aloud technique

## **An introspective study on the construct validity of GRE tests: A plea for more qualitative research in language assessment**

### **1. Introduction**

Test validation is vital for all language tests especially for high-stakes tests. High-stakes assessment is any assessment whose outcome has life-changing implications for the test taker. Admission tests for universities or other professional programs, certification exams, or citizenship tests are all high-stakes assessment situations (Rover, 2001). One essential component of the test validation process is construct validation. Construct validity concerns the extent to which performance on a test is consistent with predictions that we make on the basis of a theory of abilities, or constructs (Bachman, 1990). The notion of construct validity has come to be recognized by the measurement profession as central to the appropriate interpretation of test scores (Bachman, 1990).

As a proficiency test, GRE tests aim at measuring test takers' English communicative language ability (CLA). The notion of CLA has been operationalized differently throughout the history of language testing. One of the most detailed models of CLA is that of Bachman (1990). In his model of communicative language ability, Bachman made a distinction between what constitutes 'knowledge' and what constitutes a skill (Fulcher & Davidson, 2007). The three components of CLA for Bachman are language competence (knowledge); strategic competence (the capacity for implementing the components of language competence in contextualized communicative language use) and psychological mechanisms. Although in Iranian universities Test of English as a Foreign Language (TOEFL) and International English Testing System (IELTS) tests are usually used to measure candidates' English proficiency, GRE is sometimes used for the same purpose. Since rarely has GRE been used as a high-stakes test in Iran, this article is an attempt to investigate candidates' reactions to the use of GRE as a high-stakes test in Iran.

In a seminal paper presented in ALTE Barcelona Conference, Lazaraton (2004) argued for more qualitative approaches for test development and validation though she asserted that they are not without their problems. It is, however, further argued that, "the most important methodological development in language-testing research over the last decade or so has been the introduction of qualitative research methodologies to design, describe and, most importantly, to validate language tests" (Lazaraton, 2004 p. 51). The same concern was also stated by other researchers (e.g. Bachman, 2000; Banerjee & Luoma, 1997; Taylor & Saville, 2001). They noted the significant role of such qualitative methodologies in exploring 'the complex relationships that exist among test performance, test-taker characteristics and strategies, features of the testing process, and features of testing tasks, to name a few' (cited in Lazaraton, 2004, p. 51). In line with these ideas, the present study taking a qualitative perspective aims at investigating the validity of GRE test which is recently used in Iranian academic context as a high-stakes test.

Introspection techniques are usually utilized in the literature to tap subjects' mental processes. These techniques are used in qualitative research and have always provided invaluable insights about what actually goes on in the minds of the subjects when they perform a particular task. Mackey and Gass (2005) hold that introspective methods which were first used in the field of philosophy and psychology are used to tap participants' reflection on mental processes. In using introspection the assumption is that what takes place in consciousness can be observed in much the same way that one can observe events in the external world. Verbal reporting, a technique of introspection, consists of gathering protocols, or reports, by asking individuals to say what is going through their minds as they solve a problem or complete a task. The verbal reporting, in turn, could be realized by both self-report, and self-observation, as well as self-revelation (also known as 'think-aloud'). Using think-aloud technique, the participant provides an on-going report of his or her thought processes while s/he is engaged in performing a particular task.

The purpose of the present paper is two-fold. First, an attempt is made to shed more light on the issue of using participants' beliefs on the components of communicative language ability in language test validation. Second, through utilizing introspective procedures, the paper aims at investigating strategies test takers use in dealing with different parts of GRE test.

## 2. Review of the related literature

### 2.1 Approaches to construct validity

There have been several approaches to test validation. Bachman (1990) came up with the following two comprehensive approaches to construct validation. The first approach is the co-relational one which is derived from a family of statistical procedures that examine the relationships among variables, or measures. The approach that has been used most extensively in construct validation studies of language tests is to examine patterns of correlations among test scores, either directly, or, for correlations among large numbers of test scores, through factor analysis. Correlation coefficients, once calculated, can provide empirical information for supporting or rejecting specific interpretations and uses of test scores. This is because when we interpret a test score as an indicator of a given language ability, we do so on the basis of the hypothesized relationship, specified in the operational definition, between the test score and the ability.

The other model commonly utilized in the co-relational approach is Multi-trait-multi (MTMM) design in which two or more methods are used to measure two or more traits. The correlations among pairs of scores are often interpreted as convergent and divergent validity. The co-relational approach to validating the construct of a test, generally speaking, encompasses Multi-trait-multi (MTMM) design, factor analysis, causal modeling, and multiple linear regression analysis. The second comprehensive approach to construct validity mentioned by Bachman (1990) is the experimental one. This approach makes use of experimental (or quasi-experimental) designs to examine the relative effects of different treatment, with the treatments consisting of different methods or techniques for language teaching, or different contexts of language learning. That is, the focus has generally been on examining the treatments themselves, with the tests that are used being viewed primarily as instruments to detect differences in individuals that can then be attributed to one treatment or another. Brown (2005) divides the experimental approach in to two categories of differential- groups and intervention studies. In the former the performance of two groups on a test is compared and in the latter the performance of one group of students is taken into account. Regardless of the technique used, the basic strategy in demonstrating construct validity is the same. The tester conducts an experiment to investigate the degree to which the test is measuring the construct for which it was designed.

### 2.2 Problems with statistical approaches to construct validation

Although the quantitative analysis of patterns of test performance and of relationship between treatments and test performance constitute powerful methods for investigating specific hypotheses in the process of construct validation, it would be naïve to assume that they are the panacea in the process of construct validation for the simple reason that in language testing we are dealing with human beings and not with a bunch of figures and charts. Skehan (1990) holds that the methods of investigation in language test construct validation look rather repetitive, for example a reliance on 'after -the-event' statistical procedures such as correlation, multi-trait multi method analysis, and factor analysis. He also maintains that missing from this field of enquiry is the use of non-statistical methods to better understand the underlying processes in language performance.

In a similar vein, Weir (1990) claims that in much of the recent American literature, construct validity is looked upon from a purely statistical perspective. It is viewed mainly as a matter of a posteriori statistical validation of whether a test has measured a construct which has a reality independent of other constructs. Farhady (2006) tries to get our attention to the inherent weakness of co-relational approaches to construct validation as follows:

*“It was soon realized that correlation coefficient is just the degree of go-togetherness and communality between the two tests. In no way a single correlation can be interpreted as an indication of the existence of an underlying construct”.* (p.155)

He then specifically doubts the usefulness of factor analysis in establishing the construct validity of a language tests:

*“In spite of its power and utility, mathematical complexities and methodological diversity of the factor analytic have led to obtaining controversial results. The same set of data may lead different factor patterns depending on the techniques utilized. For instance, the most convincing evidence Oller (1979) and others offered to support [Unitary Competence Hypothesis] UCH was obtained through inappropriate application of principal component analysis which is mathematically designed to force the data toward a single factor. When another technique was used on the same data, different patterns emerged and UCH went under question”.* (p. 155)

Bachman (1990) is of the opinion that the quantitative paradigm is necessarily limited:

*“A more critical limitation to co-relational and experimental approaches to construct validation, however, is that these examine only the products of the test taking process- the test scores- and provide no means for investigating the processes of test taking them.”* (p. 269)

He then quotes Messick (1989) and states his concern over the importance of investigating the processes of test-taking:

*“In numerous application of... techniques for studying the process, it became clear that Different individuals performed the same task in different ways and that even the same individual might perform in a different manner across items or on different occasions.... That is, individuals differ consistently in their strategies and styles of task performance.”*(p. 256)

In a nutshell, what language test researchers have noticed is the fact that test validation cannot and should not be based on mere statistics. On the one hand, there has been an overreliance on statistical procedures in the construct validation of language tests, and on the other hand, some test developers have begun doubting the efficacy of statistical techniques in this field of inquiry (Skehan, 1991). They argue that the biased view of some language test developer in overreliance on statistical techniques in the process of construct validation has removed from the field of language testing what can be extremely revealing and informative in this field of inquiry.

This being so, there seems to be a possibility that having test-takers introspect about the mental processes they engage in while doing tests can be revealing and informative. It is also possible that the use of introspective techniques will come to be regarded as a standard part of test validation. Only when quantitative paradigm is complemented with qualitative paradigm can we claim that a particular test measures what it is supposed to measure. It is time for language testing researchers to embark on utilizing qualitative empirical research procedures to better understand the processes test takers utilize in taking language tests.

Language test researchers have been rather slow in making use of qualitative empirical research procedures in order to investigate the strategies involved in taking language tests. Bachman (1990, pp. 126-127, Chap. 5) cites some of these studies: The pioneer was Cohen (1984) who first reported on the examination of perceived strategies employed by language test takers and their reactions to different item and test types. The method he utilized was verbal self-report with which he identified a variety of strategies test-takers employ in completing cloze and multiple-choice reading texts. For the cloze the strategies consisted of guessing, using the immediate context, and translation. For the multiple-choice reading test, subjects reported reading the questions before the passage, reading answer alternatives before the item stem and so on. Cohen also reported that subjects at different ability levels employed different strategies. In addition, students' reactions to different tests varied in

terms of whether they liked them, felt they were difficult, or what they were perceived to measure.

Another study utilizing qualitative empirical research procedures was that of Grotjahn (1986). In his study into the 'C-test' using self-report data and logical task analysis, Grotjahn derived a model for the problem-solving process in C-test taking, which included a classification of different cognitive and meta-cognitive strategies. Although the number of studies in the qualitative paradigm is limited, the usefulness of qualitative empirical research procedures in better understanding of what test takers actually do when they take tests is becoming clear. This research approach has a great deal to offer. It can provide evidence for construct validation by complementing the quantitative analysis of tests. In line with the few studies which were carried out in the qualitative paradigm, this study was an attempt to investigate the usefulness of this approach in the validation of a high-stakes test, i.e. GRE tests.

### **3. Research questions and hypotheses**

As noted earlier, research on construct validation has overwhelmingly relied on the use of quantitative and statistical techniques. A question which needs further investigation is whether the qualitative approaches such as think-aloud techniques will also give support to the construct validity of GRE tests. Accordingly, the study was designed to make sure whether GRE tests are valid measures of language proficiency. To objectively examine the construct validity of GRE test, in this study, the communicative language ability (CLA) model proposed by Bachman (1990) was adopted. In Bachman's model, CLA is made up of three components: language competence, strategic competence, and psycho- physiological competence. The language competence which includes organizational and pragmatic competence accounts for the knowledge of language test takers might have. Strategic competence, in turn, has to do with the ability of test takers in using the components of language competence in contextualized communication. The psycho- physiological competence enables the test takers to physically execute language knowledge.

Using Bachman's (1990) model of CLA, the present study addressed the following four questions:

- a. Is the content of the GRE test compatible with the participants' expectation regarding language proficiency?
- b. Does the GRE test exactly measure the components of language proficiency as proposed by Bachman in 1990?
- c. Is the construct validity of GRE confirmed?
- d. What strategies do language students engage with while taking mathematics and analogy parts of GRT tests?

### **4. Method**

#### *4.1 Participants*

To be manageable a convenience sampling procedure was adopted and five participants were selected. Out of the twenty instructors of English in both state and private universities in Kermanshah province, Iran, 5 instructors consisting of one female and four male instructors with the mean age of 33 were randomly selected. Three of the participants were already PHD candidates of TEFL and the others had their masters' degrees in TEFL.

#### *4.2 Instruments*

The GRE was used in the present study. GRE is required as an admission test in many U.S, Canadian, and

UK universities. It is a multiple-choice examination designed to measure the verbal, quantitative (mathematics) and analytical skills. The verbal section, claimed to measure candidates' ability to use words as tools in reasoning, consists of sentence completion, analogy, reading comprehension, and antonym questions. The analogy questions are intended to test the testees' understanding of the relationships among words and ideas. First, a pair of word is presented to the testees, then they are required to choose another pair which is related in the same way, but many relationships are likely to be found. The two terms can be synonyms; one term can be a cause, the other the effect, one can be a tool, the other the worker who uses the tool.

The quantitative (or mathematical) section, claimed to measure the candidates' arithmetic, elementary algebra, and geometry ability, consists of quantitative comparison, discrete quantitative, data interpretation, and discrete quantitative questions. The candidates, it is claimed by the GRE test developer, do not need to know any more advanced mathematics to be able to answer quantitative questions. Finally, the analytical section consists of analytical and logical reasoning questions. In this section the testees are required to analyze a passage or set of conditions, reasoning out its implications, and answer a question or group of questions based on it.

As a proficiency test, GRE tests aim at measuring test takers' English communicative language ability (CLA). The notion of CLA has been operationalized differently throughout the history of language testing. One of the most detailed models of CLA is that of Bachman (1990) based on which a distinction is made between what constitutes 'knowledge' and what constitutes a skill (Fulcher & Davidson, 2007). The three components of CLA for Bachman are language competence (knowledge); strategic competence (the capacity for implementing the components of language competence in contextualized communicative language use) and psychological mechanisms.

#### *4.3 Data collection procedure*

A qualitative design was adopted in the present research. Before the exam, the test takers were not informed about the purpose of the study. Once they were finished with the test, they were interviewed to verbalize the strategies they used in answering the questions. Then they were asked to write their opinion regarding the relevance of the exam items to their English proficiency. (See appendix 1) .The study was carried out using a technique of introspection method .i.e., think-aloud technique. That is to say, the participants were required to say what was going through their minds as they were answering the items of the GRE test. To facilitate the analysis of data, the researcher audio taped what the participants said during the think-aloud process.

#### *4.4 Procedure*

The test used in the present study was a GRE sample test taken from GRE General Test (Kaplan, 2003). The researcher only included the analogy and the quantitative as well as the vocabulary parts of the GRE test because among the many parts of GRE these seem not to conform to the components of CLA as proposed by Bachman (1990) (See the appendix). The data were then gathered via think-aloud technique, a procedure of data collection in introspection method (see the introduction for further information regarding think aloud technique). To have easy access to the data, what the test takers said was thoroughly tape- recorded and then transcribed.

Before taking the test, the participants were informed about what they were supposed to do. More specifically, they were required to answer the following questions:

- Do you think these GRE questions can really measure your knowledge of English? Why, why not?
- What type of questions seems dubious? Why?
- What strategies do you use to answer these GRE questions?
- Do you believe that either mathematics or analogy reasoning have anything to do with language knowledge?

#### 4.5 Data analysis

The data collected for this study were analyzed qualitatively. That is to say, qualitative descriptions regarding the data were done so as to come to any possible patterns or themes or classification of their ideas. More specifically, first, to find out what the subjects think about the components of CLA the GRE sample test measures the written comments given by the participants were analyzed to by the researchers. Second, to see what strategies they actually used in answering the questions of thought the GRE sample test, right after the administration of the test the participant were interviewed (See the appendix for the written comments of the subjects).

### 5. Results and discussion

Qualitatively, the data gained from the think-aloud technique showed that the participants did not regard the GRE test as a good measure of English proficiency. In other words, GRE as a measure of general English proficiency test suffers from a number of inherent defects in its construction. As a matter of fact, there were some items in the test which, in the participants' views, had little or nothing to do with the components of the English proficiency; the analogy and the mathematical questions being the most dubious ones. Nearly all the participants complained that most of the questions were beyond their reasoning and mathematical abilities. The reasons put forward by the participants for this disability included lack of enough time, large numbers of the distracters, and little background in mathematics. More specifically, the participants had difficulty answering the analogy questions. The testees reported that although they knew what the words meant, they were not able to find the relationship(s) between the pairs. It seemed that answering analogy questions required that the candidates have some non-linguistic knowledge. The test takers believed that they were more of intelligence questions than English proficiency items.

#### 5.1 Research question 1

This study also set out to answer 4 questions. The first question concerned whether the content of the GRE tests is compatible with the participants' expectation regarding language proficiency. The answer was somewhat disappointing in that nearly all the participants reported that the test was measuring some irrelevant abilities having little or nothing to do with the components of language proficiency.

#### 5.2 Research question 2

The second question concerned whether GRE tests exactly measure the components of language proficiency as proposed by Bachman (1990). The authors could not find any evidence for illocutionary competence, nor did they find any evidence for sociolinguistic competence of the communicative language proficiency model as suggested by Bachman in 1990. We, however, found some items in the GRE test which tapped test-takers' strategic competence; especially in the mathematics and analogy questions. To our knowledge, nevertheless, no test specialist has yet claimed that intelligence be tested in language tests since it is related to general cognitive ability rather than aptitude.

#### 5.3 Research question 3

The third question tried to investigate whether GRE tests are constructively valid examinations. The answer to this question depends upon where the test is used. When GRE tests are used in a foreign language setting, like that of Iran, they could not be considered as a valid test in that most of the test-takers have a weak background in mathematics.

#### 5.4 Research question 4

Finally, the fourth question was raised to investigate the strategies used by the test takers in taking the

mathematics and analogy questions. By strategies it is meant the way subjects approached analogy and mathematics sections. With regard to the mathematics questions, some of the participants admitted that they had serious problems with the terminologies used and even they did not know how the terminologies had to be pronounced. In dealing with these questions, the subjects had difficulty with basic concepts of mathematics. The subjects also reported that some of the questions of this part, in fact, measured very trivial operations in mathematics. They also believed that the mathematics questions were either too easy or too difficult; in either case nothing related to English proficiency was measured.

As far as analogy questions are concerned, the subjects first tried to translate the first pair into their L1 (Persian), they then tried to discover the possible relationship(s) between parts of the pair, next they once again translated the other pairs in the responses into their L1. Once the translation was over, they had some kind of self talk, all in their L1, to find out how some other pairs could be ruled out so as to find the correct answer. The interesting thing was the fact they seldom knew what all the pairs meant. What we found in the qualitative analysis of the data confirms what language test researchers have been concerned with. Too much emphasis on quantitative approaches in test validation process will not do justice; they have to be complemented with qualitative approaches.

## **6. Conclusion**

The results of the qualitative data analysis clearly showed that mere reliance on statistical procedures in test validation of language proficiency tests in general and in GRE tests in particular cannot do justice. This is because test-takers' expectations regarding what a proficiency test ought to measure are, to a great extent, overlooked in the process of GRE test development. It was found that GRE tests suffer from both construct under-representation and construct-irrelevant problems when used as an English proficiency test and a high-stake test. GRE has construct-irrelevance problem: From the test takers' point of view, GRE tests seem to measure some constructs having little or nothing to do with the components of CLA. These irrelevant construct include intelligence, mathematics, general reasoning, knowledge of trivial words, and even knowledge of management.

GRE has construct under-representation problem: GRE does not completely measure the components of CLA based on Bachman's (1990) model of communicative language ability; Sociolinguistic and, illocutionary knowledge from pragmatic competence and knowledge of phonology knowledge from grammatical competence are not truly represented in GRE tests. In conclusion it can be argued that GRE tests should not be used a high-stakes test in Iranian universities. Of course it is not indented to say that GRE is not a valid test at all. Rather it may seem valid when used in different context and for a different purpose. If language test developers employed qualitative approach in the test validation process, more effective tests could be achieved. Since in high-stakes tests the test-takers' lives are affected, greater care has to be exercised in developing such tests.

It was also concluded that, in adopting language tests, the context for which the test is designed must be taken in to account. A test developed for university admission of native speakers in American context, for example, might not be appropriate for the same purpose in another, quite different context. This is exactly what has happened regarding the utilization of the GRE tests in Iran. There are a number of potential limitations with this study which require caution in making claims of any kind. The learners who took part in the study were almost equal in terms of English proficiency. Different results might have been gained, had the participant been different with regard to their English proficiency. The other problem has to do with the fact that the whole GRE test was not presented in its totality; in other words, only the analogy and the quantitative section of the test were used. A comprehensive study might include all other parts of a GRE test.

Still another limitation of the present study relates to the number of the participants. Only five subjects took part in the current research. If more subjects participated in the study, a different result might have been obtained. Last but not least, the technique of data collection, i.e. think-aloud method, might be used along with other types of qualitative or quantitative techniques to shed more lights on the issue of construct validity of the GRE exams



as a measure of general English proficiency, especially when it is used in a foreign language context like that of Iran.

## 7. References:

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1 - 42.
- Banerjee, J., & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Language Testing and Assessment* (Vol. 7, pp. 275-287). Amsterdam: Kluwer.
- Brown, J. D. (2005). *Testing in language programs*. New York, NY: McGraw Hill.
- Farhady, H. (2006). Construct Validity. In H. Farhady (Ed.), *Twenty five years living with applied linguistics* (pp. 135-148). Iran, Tehran: Rahnama Publications.
- Lazaraton, A. (2004). Qualitative research methods in language test development and validation. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context: Proceedings of the ALTE Barcelona Conference, July 2001* (pp. 51-71). Cambridge: Cambridge University Press.
- Mackey, A., & Gass, S. (2005). *Second language research, methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Rover, C. (2001). Web-based language testing. *Language Learning and Technology*, 5(2), 84-94.
- Kaplan, R. (2003). *GRE exam*. New York: Simon & Schuster.
- Skehan, P. (1991). Progress in language testing: The 1990s. In C. Alderson & B. North (Eds.). *Language testing in the 1990s* (pp. 3-22). London: Macmillan.
- Taylor, L. B., & Saville, N. (2001, February). *The role of qualitative methods in setting and maintaining standards in language assessment*. Paper presented at the Setting Standards for Qualitative Research in Applied Linguistics Colloquium, American Association of Applied Linguistics Annual Conference, St. Louis, MO.
- Weir, C. J. (1990). *Communicative language testing*. London: Prentice Hall International Ltd.

## Appendix

Test takers' opinion regarding the GRE items

A: The first page can be used to assess the advanced learner's knowledge of both vocabulary and reading. Moreover, it seems that they are challenging items.

The second page is designed to assess the knowledge of vocabulary; however, there are two potential drawbacks. First, the vocab. items are out of context. Second. They seem to me like jigsaws or puzzles.

The third page has nothing to do with my SL proficiency. It requires my mathematics knowledge or appears to be something like intelligent test.

B) I have always wondered who has proposed the GRE as a language proficiency test for the first time. The first part of the test seems to aim evaluating your reading and vocabulary competence regardless of the difficulty, authenticity and most importantly the frequency of such text in real English context. The second part (analogy) relies heavily on the testee's logical capacity as well as his dominance over some infrequent words. The last part enhances the problem by focusing on mathematical intelligence and skill in which most language learners are not good enough

C) The questions seem odd to me in that everything except English proficiency is measured. I have always had difficulty understanding mathematics. When I came across the questions of the analogy part I was a bit shocked. In fact they remind me of those damn days in schools.

D) This is not the first time that I have seen these questions. I actually took several GER exams. At first they were difficult to do but practice makes perfect. But to be honest with you, they are not meeting my expectations of what an English proficiency should be about.

E) They are absolutely irrelevant t to English proficiency test.