



**A Study of Construct Validity of an English Achievement Test**

**Lorena Parra Gavilánez <sup>1</sup>, Ximena Calero <sup>2</sup>**

**1 Universidad Técnica de Ambato, lparra@uta.edu.ec**

**2 Universidad Técnica de Ambato, xcalero@uta.edu.ec**

**ABSTRACT**

The present article examined the extent of construct validity of a final achievement test given to Ecuadorian university students of B1 level who study English as a Foreign Language. In order to obtain the data and analyse it, qualitative research methods and the interpretative paradigm led this investigation. A matrix based on Bachman and Palmer's framework (1996) was created to collect the information of characteristics of input reflected in the test items. A group of experts in building achievement tests participated in the validation of the test, showing a level of agreement in their judgments. The results of the study revealed that there was a good match between some test items and some specific areas of the framework. However, there were sections of the test that could not be validated effectively which led to conclude that the test has partial construct validity. It was suggested that there is a need for training the teachers regarding to the design and validity of achievement tests as part of their professional development.

**Keywords: Construct validity, achievement test, test items, characteristics of input**



## **Un estudio de la validez de constructos de una prueba de logro en inglés**

### **RESUMEN**

El presente artículo examinó el alcance de la validez de constructo de una prueba de rendimiento final otorgada a estudiantes universitarios Ecuatorianos de nivel B1 que estudian inglés como lengua extranjera. Con el fin de obtener los datos y analizarlos, los métodos de investigación cualitativa y el paradigma interpretativo condujeron a esta investigación. Se creó una matriz basada en el marco de Bachman y Palmer (1996) para recopilar la información de las características de la entrada reflejada en los ítems de la prueba. Un grupo de expertos en la elaboración de pruebas de aprovechamiento participó en la validación de la prueba, mostrando un nivel de acuerdo en sus juicios de valor. Los resultados del estudio revelaron que había una buena concordancia entre algunos ítems de prueba y algunas áreas específicas del marco usado para la validación de la prueba. Sin embargo, hubo secciones de la prueba que no pudieron ser validadas efectivamente, lo que llevó a concluir que la prueba tiene parcial validez del constructo. Se sugirió que es necesario capacitar a los maestros con respecto al diseño y la validez de las pruebas de aprovechamiento como parte de su desarrollo profesional.

**Palabras claves:** validación del constructo, prueba de aprovechamiento, ítems de la prueba, características de entrada.



## **1. INTRODUCTION**

Achievement tests are essential part of classroom assessment in the teaching and learning process due to its importance for teachers to obtain evidence of the learners' achievements in relation to the instructional objectives. Based on this evidence, teachers make judgments and decisions about student's competence, promotion, weaknesses and strengths. According to McNamara (2000), an achievement test is an instrument of evaluation which purpose is directly related to the teaching and learning process that serves as a powerful making decision tool and support to reach the learning outcomes during or at the end of a course study. Achievement tests as instruments that contribute to classroom assessment should meet qualities of validity that guarantee the capability of measuring the extent to which the learning objectives are achieved.

One of this qualities is construct validity which refers to the appropriateness of score interpretation as an indicator of the ability or the construct that teachers or test developers want to measure taking into account that for justifying such interpretations, it is essential to provide evidence of the ability to be measured (Bachman & Palmer, 1996). Then, a test that reflects construct validity will adequately assess the theoretical concepts that it is claimed to do.

Designing a valid test requires that teachers have the appropriate knowledge and training about the construction of instruments of evaluation and be aware of the process of validation to ensure that such instrument will provide reliable information about the learners' achievements.

For the Ecuadorian context the English language is a required subject in the tertiary education as a complement of the curriculum, and achievement tests play an important role to evaluate university students' knowledge gained through the learning process in an EFL program in which learners have to perform their ability to use the English language through a final achievement test at a B1 level according to the Common European Framework. In the case of the language university centre of my context, this final achievement test is designed by trained language teachers of the English program and the results obtained through this evaluation are taken into account to make decisions about the promotion on the students' performance. However, problems related to the



inadequacy of the test items are reported by the course teachers claiming alternatives that can determine the problem and find solutions for having a fair test that allows having reliable results.

As a way to determine the usefulness of the test, the validation process represents a paramount stage due to the strong justifications it provides to use score interpretation as a reference to make decisions (Bachman & Palmer, 1996). This study attempted to focus on level of agreement that the experts had between the test items and the characteristics of input adapted by Bachman and Palmer's framework in order to determine the level of construct validity of the test. The present investigation addresses the following research questions:

- RQ 1: Does the final achievement test applied to B1 level university students of my context reflect construct validity?
- RQ 2: To what extent does the achievement test reflect agreement between the test items and the characteristics of input to determine its validity?

### **Achievement Tests**

An achievement test is a measurement tool used to evaluate students' progress in relation to a specific formal program in language education. The teaching and learning process has a strong relationship with classroom assessment due to the compromise and influence of testing when the different factors of this process reflect positive or negative effects on test takers (Cheng & Watanabe, 2004). Furthermore, achievement tests as tools that serve to diagnose the students' weaknesses and strengths to make decisions regarding the instructional progress, these type of tests also provide teachers feedback about positive and negative aspects of their instruction (Madsen, 1983; Shaaban, 2001). Green (2013), teachers have a big responsibility in their hands when constructing assessment tools that reflect connection between the content of the test and the learners' performance. At this point, achievement tests play an essential role for both, teachers and students being mandatory that teachers pay much attention to test construction ensuring its usefulness through the validation of the test. According to Alderson, Clapham, and Wall (1995), the construction of achievement tests should be based on the content of the curriculum and appropriate procedures that ensure the validity and



reliability of the test as well as the recognition of its purpose in relation to the syllabus and the learners' achievements.

### **Construct Validity**

Construct validity is one of the principles that plays an important role in the validation process of a test to determine the level of usefulness of this tool of evaluation.

According to Weir (1990); Alderson et al. (1995), one of the most relevant types of validity is construct validity due to its characteristic to gather evidence to support the contention that a given test indeed measures what is claimed to measure. When there is evidence that the significance of the measurement is given by the correspondent harmony between the theory and the constructs of a test, then we can say that the test has construct validity (Cumming & Berwick, 1996). For Gronlund (1988), the different types of procedures used to determine the construct validity of a test provide reliable evidence of the connection between the theoretical structure and the measured aspects of a test. Then, teachers and test developers should make predictions about performance on the test by interpreting the harmony a particular construct has in relation to the theory.

The term "construct validity" makes reference when the candidate's knowledge is inferred through the results of a test in which the test tasks are clearly designed according to the nature of the construct to be measured and such inferences depend on language theoretical hypotheses that need to be justified to give validity to the test (Bachman & Palmer, 1996). In short, the test tasks are verified to see if they measure the target language skill and sub skill in order to determine its validity and the comparison between the test tasks and the theory is one of the ways to assess the construct validity of the test (Brown, 2004).

In other words, the construct validity of a test is determined by the evidence obtained through the study of the influence of several factors that contribute to the general score interpretation as an indicator of the ability or construct to be measured (Birjandi & Mosalanejad, 2010).

### **Language Task Characteristics**

One of the ways to determine the construct validity of a score interpretation is considering the construct definition and the characteristics of the test task as suggested



by Bachman and Palmer (1996) framework, in which the level of correspondence between the construct or constructs of the target language and the test tasks are relevant factors to determine the usefulness of a test and its purpose for what it was designed.

Bachman and Palmer (1996, p. 46) state, “the characteristics of the tasks used are always likely to affect test scores to some degree, so that there is virtually no test that yields only information about the ability we want to measure”. Therefore, the students’ performance is affected not only by the test tasks, but also by the way teachers evaluate the language skills and the proportion of the tasks used to test the target constructs.

The tasks characteristics proposed by Bachman (1990) are related to setting, test rubric, input, expected response and relationship between input and response. All these characteristics are involved in describing the target language use (TLU), differentiating the test tasks to determine comparability and reliability, and comparing the characteristics of TLU to test authenticity (Bachman, 1990).

### **Characteristics of Input**

The material used in a task to be processed in an effective way by the learners corresponds to input which characteristics respond to format that describes channel, form, language, length, type, speediness and vehicle; and language that includes organizational and pragmatic characteristics (Bachman & Palmer, 1996). Moreover, to determine the construct validity of a test we can define the competence, knowledge, skill and ability through the definition of the test task that the learners should be able to perform (Chapelle, 1998 as cited in Buck, 2001)

For achievement tests, the characteristics of tasks are fundamental descriptors that can help to determine the usefulness of the test due to the direct control and influence teachers have on it as a handmade tool used to evaluate the learners’ achievements.

### **Previous Studies**

“An Evaluation of the Construct Validity of Iranian National Test of English at High Schools” is one of the studies that investigated the construct validity of a final national test applied to last level secondary students. This study used item analysis and correlation coefficient by adopting quantitative method. It was determined a significant



difference between the two versions of tests concluding the lack of construct validity in the evaluation tools (Zhoghi, Rostami, & Gholami, 2016).

Khodadady (2014) examined the construct validity of C-tests using factorial approach. Undergraduate and graduate students' performance majored in English Language and Literature was analysed through the application of six tests and the principal component analysis in relation to the participants' answers. The C-tests, Decontextualized C-test and Spelling test revealed to have two dominant components which are language proficiency and direction specificity. Meanwhile the application of the S-test, TOEFL and Lexical Knowledge test confirmed the same two components reflecting construct validity as proficiency measures of language.

Phakiti (2008) investigated the construct validity of the theory of the strategic competence proposed by Bachman and Palmer (1996) by using the structural equation model. The relationship of L2 test takers' long-term strategic knowledge and actual strategy use were studied by applying a questionnaire after an achievement test given to Thai university students. The studied confirmed that trait metacognitive strategy use affects both the cognitive strategy use and the state of MSU, and hence this last directly affects a specific language test performance.

## **2. METHOD**

This study used the qualitative method by its exploratory, intuitive and nonlinear characteristics that allow to conduct an interpretative analysis of the document in study (Heigham & Croker, 2009). The interpretative paradigm and the documentary analysis led this research.

### **Participants**

As to the purpose of this study, five Ecuadorian experts in the area of building tests participated validating the final achievement test applied to Ecuadorian university students of B1 level of English as a Foreign Language. These five experts have been working in the university context where the study was carried out for almost a decade as English teachers and test makers.

### **Instruments and materials**



In order to answer the research questions of this study, the following matrix based on Bachman and Palmer's framework (1996) was designed to collect the information and determine the construct validity of the test through the task characteristics of input analysed in each section of the test.

**Characteristics of input.**

INSTRUCTION Compare the attached exam with the present matrix by marking with an X in the correct alternative		Matrix for defining the construct validity of an English achievement test through task characteristics: <i>(Adapted from Bachman and Palmer's framework)</i>																														
		<b>Characteristics of setting:</b> <i>Physical characteristics:</i> Location: classroom Materials: paper, pen, eraser, audio equipment <i>Participants:</i> Teachers of B1 English level <i>Time of each task:</i> Not specified																														
		<b>Characteristics of the test rubrics</b>																														
		Format												Language of input										Topical characteristics								
LISTENING		channel	form	language		length			type		vehicle		grammatical		textual		functional			sociolinguistics			personal	cultu	acad	tech						
task	item	aural	visual	language	non language	native	target	single words	phrases	sentences	item	prompt	live	reproduced	vocab	syntax	phonology	cohesion	conversational	ideational	manipulative	heuristic	imaginative	dialects	registers	idiomatic	cultural					
A	1																															
	2																															
	3																															
B	4																															
	5																															
	6																															
	7																															
C	8																															
	9																															
D	10																															

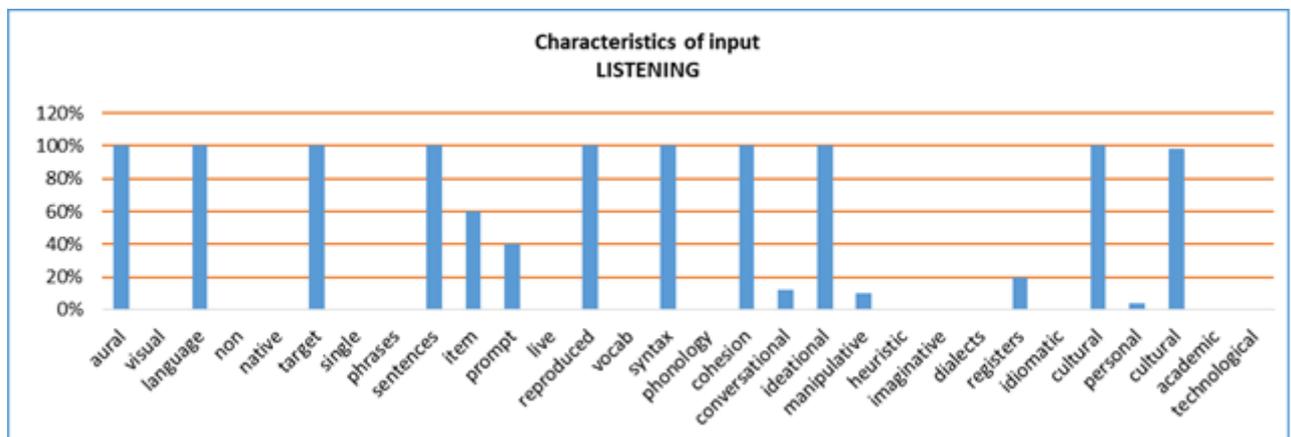
**Figure 1.** Characteristics of input

The final achievement test used in this study contains five sections divided in listening, grammar, vocabulary, reading and writing. Each section is scored over twenty points and the overall score of the test is evaluated over one hundred. The sixty items of the achievement test were aligned in terms of characteristics of input.

### Data Collection

The test, instrument of this studied was designed in 2016 by teachers who were teaching the B1 level of English at the university language centre where the investigation was carried out. The test was analysed with the pertinent authorization of the director of the institution.

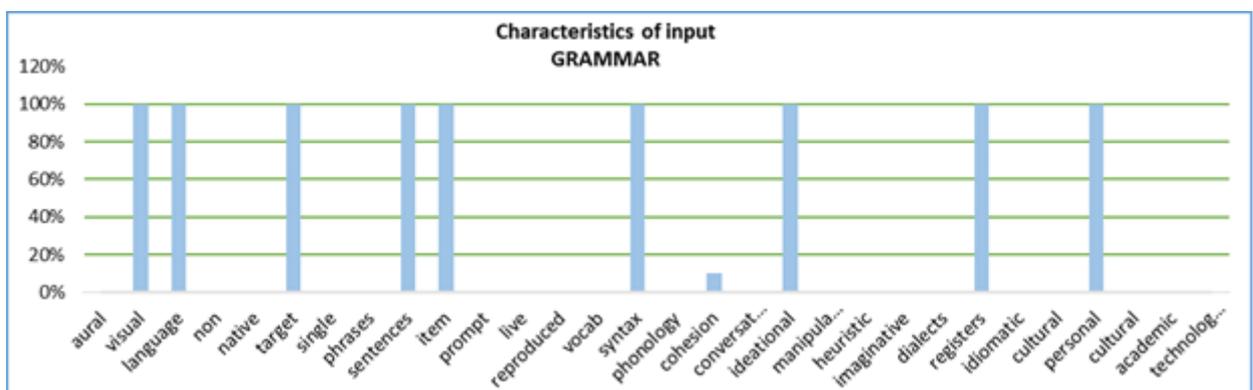
### 3. RESULTS





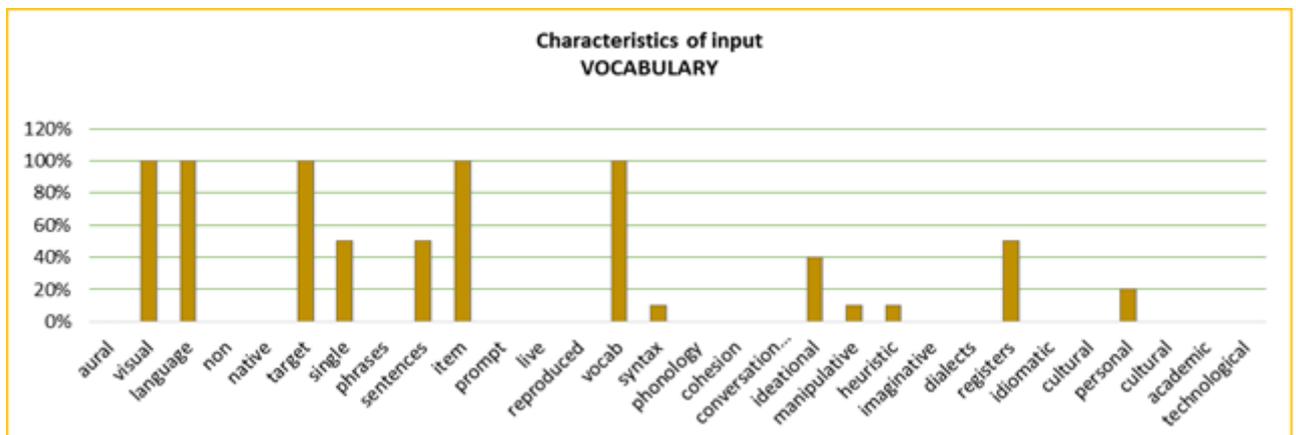
**Figure 2.** Characteristics of input in the Listening section of the test

As we observe, there is a 100% agreement in regard to the following characteristics of input found in the test as: aural channel, language form, target language, sentence length, and reproduced vehicle, syntax language, cohesion, ideational function and cultural references. The experts concluded that the 60% of the task types is built by items and the 40% by prompts. The items involved in textual language is conversational in a 12% while the items involved in functional language is manipulative in a 10%. The experts agreed that the test items include sociolinguistic characteristics as register in a 20%, cultural topics in 98% and personal topic in a 4%.



**Figure 3.** Characteristics of input in the Grammar section of the test

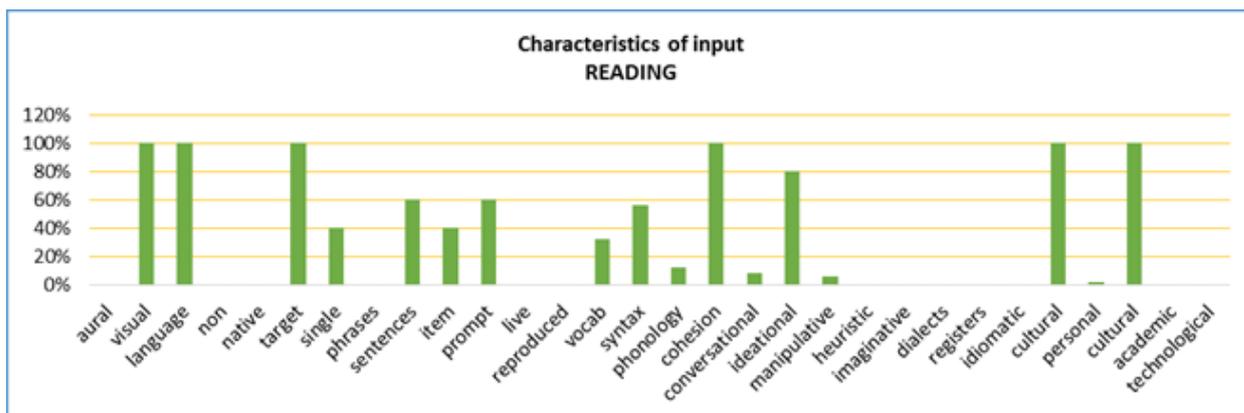
For the grammar section, the experts concluded that visual channel, language form, target language, sentences length, item type, syntax language, ideational function, register sociolinguistics characteristic and personal characteristics are met in the test items in a 100%, while knowledge of cohesion is met in a 10%. The rest of characteristics of input are not found in the item test of this section.



**Figure 4.** Characteristics of input in the Vocabulary section of the test



There is a 100% of agreement that the following characteristics of input met the test items as visual channel, language form, target language, item type and vocabulary language. The 50% of agreement corresponds to single words and another 50% to sentence length while a 10% is related to syntax language. The experts found that a 40% of test tasks corresponds to ideational function and another 10% to manipulative and heuristic function. The sociolinguistic characteristic is met through register in a 50%, meanwhile the topical characteristic determines to be personal in a 20% of consensus.



**Figure 5.** Characteristics of input in the Reading section of the test

For the reading section of the achievement test, it is observable that the visual channel, language form, target language, cohesion, cultural sociolinguistic characteristic and cultural topic characteristic have the 100% agreement between the experts. In relation to the length of input, the level of consensus reaches 40% for the use of single words while for the use of sentences it gets 60%. Moreover, the specialists found that the tasks of this section are built by 40% of items and 60% of prompts, while the language of input is delivered by 32% of vocabulary, 56% of syntax, 12% of phonology and 8% of conversational characteristics. In terms of pragmatic characteristics, the tasks are found to be ideational in 80% and manipulative in 6%. The tasks reflect personal topic in a consensus of 2%.

#### **4. CONCLUSION**

In this section, the analysis was based on the level of agreement the testing specialists had when they validated the achievement test by aligning the characteristics of input drawn by Bachman and Palmer (1996) and the test items, focusing our attention on the most relevant statistical information.



Generally speaking, more than the fifty percent of consensus was met for the sections of listening and reading, meanwhile for the sections of grammar and vocabulary it was struggled to identify some characteristics of input.

Regarding the first question which addresses if the final achievement test analysed reflects construct validity, it was found that the test exhibits construct validity in specific sections of the test. This leads to conclude that the construct validity is fragmented in the achievement test which demonstrates that one of the characteristics of a useful test needs to be reinforced to apply a fair and accurate instrument of evaluation.

In relation to the second research question, which refers to the extent of agreement between the test items and the characteristics of input to determine the construct validity of the test, it is pertinent to say that the test tasks reflect a high percentage of agreement in terms of format. However, in terms of language characteristics the percentage of consensus is low especially for the grammar and vocabulary sections, tending to indicate that the nature of language used in the areas of language knowledge and topical knowledge are characteristics that need more attention in all the sections when designing the test.

For the listening and reading sections of the achievement test, the characteristics of input tend to have a good level of agreement for both, the format of input and the language characteristics. However, the pragmatics characteristics in the language of input used in the test are vaguely met in its functional and sociolinguistic different alternatives set to provide a variety of characteristics that may contribute with a useful test.

For the grammar and vocabulary components of the test, the format continues having a high level of consensus. Meanwhile, the language of input reflects to be a weakness in these sections of the test due to its significant absence of diversity language and pragmatic characteristics in the test tasks. The ideational function, register sociolinguistic quality and personal topic were the only pragmatic characteristics met in the test.



## **5. REFERENCES**

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge [England]; New York, NY, USA: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford: Oxford University Press.
- Birjandi, P., & Mosalanejad, P. (2010). An overview of testing and assessment. *Tehran: Sepahan Publication*.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson/Longman.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Cumming, A. H., & Berwick, R. (1996). *Validation in language testing* (Vol. 2): Multilingual Matters.
- Cheng, L., & Watanabe, Y. (2004). *Washback in language testing: Research contexts and methods*: Routledge.
- Green, A. (2013). *Exploring language assessment and testing: Language in action*. London: Routledge.
- Gronlund, N. E. (1988). *How to construct achievement tests*: Prentice Hall.
- Heigham, J., & Croker, R. A. (2009). *Qualitative research in applied linguistics : a practical introduction*. Houndmills, Basingstoke, Hampshire [England]; New York: Palgrave Macmillan.
- Khodadady, E. (2014). Construct Validity of C-Tests: A factorial approach. *Journal of Language Teaching and Research, 5*(6), 1353-1362.
- Madsen, H. S. (1983). *Techniques in Testing*. Oxford: Oxford University Press.
- McNamara, T. F. (2000). *Language testing*. Oxford [England]; New York: Oxford University Press.
- Phakiti, A. (2008). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language testing, 25*(2), 237-272.



**A Study of Construct Validity of an English Achievement Test**  
*Revista Publicando, 4 No 12. (1). 2017, 297-309. ISSN 1390-9304*

- Shaaban, K. (2001). *Assessment of young learners*. Paper presented at the English Teaching Forum.
- Weir, C. J. (1990). *Communicative language testing*. New York: Prentice Hall.
- Zhoghi, M., Rostami, G., & Gholami, H. (2016). An Evaluation of the Construct Validity of Iranian National Test of English at High Schools. *Journal of Applied Linguistics and Language Research, 3*(1), 185-196.