

The Impact of Construct-Irrelevant Factors on the Validity of Reading Comprehension Tests

Zahra Ahmadjavaheri¹, Mitra Zeraatpishe²

Received: 12 November 2019

Accepted: 16 February 2020

Abstract

Test validity can be jeopardized to the extent that test scores are contaminated by factors other than those intended by the test. Many method specific factors and examinees' personal characteristics can influence test performance. In this study the contribution of four construct-irrelevant factors, namely, exam anxiety, reading anxiety, ambiguity tolerance, and attitudes towards reading to reading comprehension test performance was examined. Correlation and multiple regression analysis showed that reading anxiety and attitudes toward reading significantly contribute to reading test performance while exam anxiety and ambiguity tolerance do not. This finding shows that examinees' reading anxiety and attitudes toward reading can invalidate the reading comprehension test scores to a certain degree. Implications and limitations of the findings are discussed and some recommendations for test administrators and teachers are provided.

Keywords: validity, reading comprehension test, exam anxiety, reading anxiety, ambiguity tolerance, attitudes toward reading

1. Introduction

Validity is the most important characteristic of tests. Validity is defined as the degree to which a test measures what it is supposed to measure (Borsboom, Mellenbergh, & van Heerden, 2004). The variance in test scores could be attributed to three sources: the ability being tested, random errors of measurement, and undesirable systematic errors that contaminate the test scores. To the degree that a test is influenced by systematic errors the test becomes invalid. Addressing random errors is the concern of reliability. Messick (1989) states that there are two threats to validity: construct under-representation and construct-irrelevant variance. Construct under-representation occurs when the assessment is too narrow and limited and the content is not sampled thoroughly in the test. In the case of construct-

¹ English Department, Mashhad Branch, Islamic Azad University, Mashhad, Iran, (*corresponding author*), Email: mitra.zeraatpishe@yahoo.com

² English Department, Mashhad Branch, Islamic Azad University, Mashhad, Iran

irrelevant variance, the assessment is too broad and several irrelevant factors creep into the measurement.

Test performance might be influenced by the differences in individual characteristics that are irrelevant to the construct of interest. There are several systematic factors that could influence test scores such as background knowledge, sex, age, cultural background, and cognitive characteristics. If performance on tests is influenced by these factors the test has become invalid. Bachman and Palmer (1996) stated that examinees' characteristics are related to test performance to some extent. They further suggest that it is impossible to prevent the relation of individuals' characteristics and their test performance. Therefore, it is necessary to factor out their impact on performance and prevent test bias and omit their overlap with the trait the test is designed to test.

An important step in test validation is to ascertain that test performance has not been unduly affected by factors other than those intended by the test. This area of research in language testing has been referred to as test bias (Bachman, 1990) and differential test functioning (Baghaei, 2009; Holland & Wainer, 1993). DIF occurs when two groups of examinees with the same ability have different probabilities of correctly answering an item. Suppose an item has turned out to be easier for boys than girls while in general boys and girls are the same in the ability being measured. This is an instance of DIF or test bias as performance on the item is influenced by an irrelevant factor, namely, examinees' gender rather than the ability of interest.

Previous research has demonstrated that background knowledge influences reading and listening comprehension test scores. For example, Hale (1988) and Chavanachart (1984) in the context of English for Specific Purposes showed that familiarity with the content area accounts for performance on reading and listening comprehension tests (cited in Bachman, 1990). Chacevych, et al. (1982, cited in Bachman, 1990) showed that performance on cloze tests is affected by examinees' familiarity with passage content. Baghaei and Motallebzadeh (2010) had the same finding in the context of C-Test.

Hansen and Stansfield (1981/1984) investigated the relationship between field-dependence/independence cognitive style and test performance. Field independence is the tendency to separate details from the surrounding context while field-dependent individuals tend to see the separate details as a whole integrated unit. They found a significant correlation between field-independence and performance on cloze tests while it was uncorrelated with course grades and grades on oral and written tests of Spanish. Researchers found that field-independent people perform better on cloze tests than field dependent people. In another study, Chapelle (1988) found that field-independence cannot predict cloze, dictation, multiple-choice reading comprehension, and essay writing among nonnative speakers of English. However, for native-speakers field-independence was significantly correlated only with cloze.

Ambiguity tolerance (AT) is defined as the ability to stay calm in situations where things are not clear (Chapelle & Roberts, 1986). Intolerance of ambiguity is the tendency to perceive ambiguous situations as sources of threat. Some psychologists have argued that cognitive styles such as ambiguity tolerance are a generalized personality trait that remains relatively stable across domains. It is assumed that tolerance of ambiguity is related to cloze test performance where there is a lot of ambiguity in responses. That is, those who are more tolerant in ambiguous situations and stay calm should perform better on cloze tests.

However, it should not be related to multiple-choice item performance where everything is clear.

Chapelle and Roberts (1986) studied the relationship between tolerance of ambiguity and performance on language proficiency tests. Their findings showed that ambiguity tolerance has a low significant correlation with multiple choice tests of English proficiency and non-significant correlations with cloze tests. The correlation between ambiguity tolerance and dictation was significant. Some researchers have argued that some level of ambiguity tolerance is facilitative in test performance while excess levels are debilitating. The stance that a moderate level of ambiguity tolerance is more beneficial to second language learning than high or low ambiguity tolerance was confirmed by El-Koumy (2000). Using the MAT-50 (Norton, 1975) as a measure of ambiguity tolerance, El-Koumy classified the subjects, 150 EFL students, into three groups: high level of AT, middle-AT, and low-AT. The middle level of AT group attained significantly higher scores on the reading comprehension subtest of TOEFL than both high- and low-AT groups.

Other factors that affect language test performance are foreign language anxiety and exam anxiety. Birenbaum (2007) investigated that most students experience some levels of anxiety during an exam. When anxiety affects test performance, it becomes a problem. Test anxiety can interfere with the students' ability to perform adequately and prevent students from demonstrating their knowledge on examinations. Some students have the skills and knowledge to do very well in testing situations, but their excessive anxiety impairs their performance. Among many factors, anxiety can explain the differences in reading comprehension performance among learners and has a tremendous effect on foreign language reading.

In the context of foreign language testing, Zheng and Cheng (2018) demonstrated that cognitive test anxiety and test confidence are significant predictors of English language test performance while classroom anxiety was not. Likewise, Ganschow et al. (1994) showed that students with high levels of anxiety had poorer language skills. Ganschow and Sparks (1996) also reported that students with low anxiety performed better than those with high anxiety. In another study, Salehi and Marefat (2014) showed that final exam performance is negatively related to both exam anxiety and foreign language classroom anxiety. In meta-analyses by Seipp (1991, more than 126 trials) and Hembree (1988, more than 526 trials) test anxiety was negatively correlated with success in university studies ($r = -0.21$). However, maximum correlations in single studies reached about $r = -0.30$; explaining about 9% of the common variance.

Reading comprehension is also related to more specific components of foreign language anxiety such as reading anxiety and attitudes towards reading. Previous research has demonstrated that FLRA affects the reading process and also reading performance (Sellers, 2000; Shi & Liu, 2006). Some variables are related to students' reading behavior, including students' attitudes toward reading (Greaney & Hegarty, 1987). Eagly and Chaiken (1993) defined attitudes toward reading as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (p.1). The purpose of the present study is to examine the contribution of four variables including exam anxiety, reading anxiety, ambiguity tolerance, and attitudes toward reading to reading comprehension test performance.

2. Method

2.1 Participants and Setting

The participants of the study included 217 (170 females) undergraduate students of English as a foreign language in Ferdowsi University, Tabaran University, and Islamic Azad University in Mashhad, Iran. The age range was 19 to 58 with mean of 22.90 and standard deviation of 5.84.

2.2 Instrumentation

2.2.1 Exam Anxiety

The Persian translation of the Cognitive Test Anxiety Scale (PCTAS) by Baghaei and Cassidy (2014) was used. PCTAS is the revised version of the Cognitive Test Anxiety Scale (CTAR-17) with 17 items. They used the Rasch rating scale model to examine the psychometric qualities of the scale and its reliability.

2.2.2 Reading Anxiety

The researcher used the Persian version of the Foreign Language Reading Anxiety Scale, which is translated by Baghaei, Hohensinn, and Kubinger (2014). They used Rasch model analysis to provide evidence of validity for the scale.

2.2.3 Ambiguity Tolerance

Tolerance of Ambiguity Scale was developed by Ely (1995). The 12-item scale was translated by the researchers into Persian and was validated using the Rasch model. The scale has a reliability index of 0.74 (Cronbach alpha).

2.2.4 Attitudes toward Reading

Attitudes toward Reading Scale was translated by Akbari, Ghonsooly, Ghazanfari, and Shahriari (2017). This test consists of 35 items on a 5-point Likert scale. The test has two sets of parallel items on first language and second language.

2.2.5 Reading Comprehension Test

The researcher used the reading comprehension section of the Preliminary English Test (PET), for measuring reading ability of the participants. PET is an English language examination provided by Cambridge Assessment English (previously known as Cambridge English Language Assessment and University of Cambridge ESOL examinations). The level of the test is B1 which is an intermediate-level test and is designed for learners who have mastered the basics of English and now have practical language skills for everyday use.

3. Analysis and Results

Table 1 shows the descriptive statistics for the variables of the study. Since the number of items in the instruments and the method of scoring is not the same comparison with the mean is not possible.

Reliability of the instruments was examined using Cronbach's alpha. Table 2 shows alphas for the tests. Except for ambiguity tolerance all the scales have acceptable reliabilities

above 70. The low reliability for the ambiguity tolerance scale could be due to the small number of items in this instrument.

Table 1: Descriptive Statistics for the variables of the study

	N	Min	Max	Mean	Std. Deviation
Exam Anxiety	217	15.00	68.00	32.64	11.48
Reading Anxiety	217	30.00	89.00	57.98	13.24
Ambiguity Tolerance	217	.00	43.00	22.10	7.25
Attitudes Toward Reading	217	58.00	141.00	99.19	12.54
Reading Performance	217	1.00	17.00	9.51	3.83

Table 2: Reliabilities for the Scales

Scale	Number of items	Alpha
--------------	------------------------	--------------

Exam Anxiety	17	.93
Reading Anxiety	20	.82
Ambiguity Tolerance	12	.66
Attitudes Toward Reading	30	.81
Reading Performance	20	.70

The coefficients of correlation of students' reading comprehension scores and GPA (Grade Point Average) with exam anxiety, ambiguity tolerance, attitudes towards reading, and reading anxiety were computed. Table 3 shows the results.

Table 3: Correlation coefficients between Reading Comprehension and other Variables

	Exam Anxiety	Ambiguity Tolerance	Attitudes Reading	Reading Anxiety
Reading Comprehension	-.14*	.05	.23**	-.37**
GPA	-.28**	.11	.16*	-.22**

**Correlation is significant at the 0.01 level (2-tailed).

*Correlation is significant at the 0.01 level (2-tailed).

As Table 3 shows, reading comprehension is significantly correlated with reading anxiety and attitudes towards reading. It is also significantly correlated with exam anxiety but the correlation is very low although statistically significant. Ambiguity tolerance is not associated with performance on reading comprehension test. This pattern is observed for GPA too, except that GPA is more strongly correlated with exam anxiety than with reading anxiety. Multiple regression analysis was run to examine the combined contribution of the independent variables (exam anxiety, reading anxiety, ambiguity tolerance, and attitudes toward reading) to the prediction of reading comprehension test scores. Findings showed that the four predictors account for 27% of the variance in reading comprehension test scores ($R^2=.27$, adjusted $R^2=.25$, $F=19.70$, $p=0.00$). Reading anxiety had the greatest contribution (Beta=-.51, $p=0.00$), followed by attitudes towards reading (Beta=.36, $p=0.00$). Ambiguity tolerance (Beta=-.08, $p=.23$) and exam anxiety (Beta=-.01, $p=.86$) had no significant contribution.

4. Discussion

According to validity theory, tests should only measure a single ability of examinees; otherwise the test will not be valid. Nevertheless, it is impossible to prevent the inclusion of some irrelevant method specific and individuals' characteristics factors into the test scores (Baghaei & Aryadoust, 2015; Baghaei & Ravand, 2019; Hohensinn & Baghaei, 2017). A test or item can be considered to be biased if one particular section of candidate population is advantaged or disadvantaged by some features of the test or item which is not relevant to what is being measured. Among these characteristics, cognitive characteristics of test takers such as ambiguity tolerance, exam anxiety, attitudes toward reading, and reading anxiety have received less attention over the past few decades.

In this study, the contribution of four individual characteristics to reading comprehension test performance was examined. Exam anxiety, reading anxiety, ambiguity tolerance, and attitudes toward reading were specifically selected for examination in this study. Correlation and regression analyses showed that the four variables all together account for 27% of the variance in reading comprehension test performance. Among the predictors, reading anxiety and attitudes towards reading had significant contribution and exam anxiety and ambiguity tolerance were not associated with reading comprehension test performance. Unexpectedly, reading test performance had a very small correlation with exam anxiety. These findings suggest that construct irrelevant factors of reading anxiety and attitudes towards reading can explain reading comprehension test performance. In other words, reading comprehension test scores become less valid to the extent that they are affected by reading anxiety and attitudes towards reading. When test performance is unduly affected by anything other than the examinees' ability, the validity of score interpretations is violated. The relationship between test anxiety and test performances has been shown by some researchers like Pajares and Schunk (2001).

6. Conclusions

The results of the current study have some implications for the language teachers and learners. This study sheds light on the importance of affect and personality characteristics in EFL pedagogy and testing. In general, English teachers and learners can benefit from the findings of this study so that they can take some steps towards mitigating the impact of irrelevant factors on test performance. Alrabai (2015) states that "EFL teachers should behave as models and should avoid antagonistic behavior toward students, such as aggressively criticizing, overcorrecting errors or blaming students when they make mistakes, engaging students in competition, and offering public comparisons of the performance or grades of different students. Teachers should promote cooperative learning in which the students work together instead of competitively, show faith in students' abilities to succeed and make use of encouraging expressions for this purpose, provide students with positive feedback and appraisals of their performance, provide students with more control over their learning, and involve students in decision-making. Teachers should also tackle learners' beliefs and misconceptions that can evoke their feelings of anxiety" (p.21). To this list one can add assigning appropriate activities in accordance to learners' competency, controlling

the class in a way that students never fear from being laughed at by the others, and helping learners to react positively to each mistake they make.

These recommendations can be generalized to testing contexts. Teachers and test administration staff should provide a positive atmosphere during the testing session, assigning appropriate tasks similar to those in the test during the course, reducing the negative social consequences of failure, and promoting the view that each test is an opportunity to learn among students. , teachers should focus on the students' learning process rather than just their production and marks. Teachers should be less critical on the performance of students and provide them with more chances to show their ability. In other words, teachers can help their students by minimizing the stressful conditions in the test setting. They can establish positive rapport with such learners in an attempt to make learning more pleasurable and profitable for them.

One of the limitations of this study is that it is not possible to separate the contribution of the studied variables to reading comprehension test performance from their contribution to actual reading comprehension ability. In other words, we cannot conclusively state that reading anxiety and attitudes towards reading contribute to test performance. It is very probable that reading anxiety and attitudes towards reading contribute to reading comprehension. That is, those with higher reading anxiety and those with negative attitudes towards reading are poorer readers and one cannot conclude that they only perform poorly on the tests. That is, we cannot separate performance on reading comprehension tests from actual reading ability.

The study was restricted to quantitative methods in which data were gathered through questionnaires, only. It is suggested that other researchers use more varieties of techniques for data collection including interviews and diaries. Also, only four affective variables were considered in this work which can be expanded aiming at investigating the impact of other affective and cognitive variables on reading comprehension test performance. It is also interesting to investigate if the correlation between such factors and test performance is the same for all foreign language skills tests. This study focused on reading comprehension. However, we do not know if these variables are equally related to listening, speaking, writing and grammar tests too. This research was carried out on a limited available sample of Iranian EFL learners and cannot be truly generalized to the rest of the EFL learners in Iran. Therefore, the same procedure can be followed by other researchers to investigate the issue on other student populations. There is no doubt that larger samples provide better understanding about the interrelationship among affective variables and test performance.

References

- Akbari, H., Ghonsooly, B., Ghazanfari, M., & Shahriari, H. (2017). Attitude toward Reading: L1 or L2 or Both. *Sage Open*, 7, 1-10.
- Alrabai, F. (2015). The influence of teachers' anxiety-reducing strategies on learners' foreign language anxiety. *Innovation in Language Learning and Teaching*, 9, 163-190.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Baghaei, P., & Ravand, H. (2019). Method bias in cloze tests as reading comprehension measures. *Sage Open*, 9, 1-8.
- Baghaei, P. (2009). *Understanding the Rasch model*. Mashhad: Mashhad Islamic Azad University Press.
- Baghaei, P., & Aryadoust, V. (2015). Modeling local item dependence due to common test format with a multidimensional Rasch model. *International Journal of Testing*, 15, 71–87.
- Baghaei, P., & Motallebzadeh, K. (2010). The effects of the prior knowledge of passage content on C-Test scores: An investigation of text-level processing. *Journal of Teaching English as a Foreign Language and Literature*, 2, 25-39.
- Baghaei, P., & Cassidy, J. (2014). Validation of the Persian translation of the Cognitive Test Anxiety Scale. *Sage Open*, 4, 1-11.
- Baghaei, P., Hohensinn, C., & Kubinger, K. D. (2014). Persian adaptation of Foreign Language Reading Anxiety Scale: A psychometric analysis. *Psychological Reports*, 114, 315-325.
- Birenbaum, M. (2007). Assessment and Instruction Preferences and their Relationship with Test Anxiety and Learning Strategies. *Higher Education*, 53, 749-768.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Chapelle, C. (1988). Field independence: a source of language test variance? *Language Testing*, 5, 62-82.
- Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field dependence as predictors of proficiency in English as a second language. *Language Learning*, 36, 27-45.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.
- El-koumy, A. S. (2000). Effects Of Skills-based Versus Whole Language Approach on the Comprehension of EFL Students with Low and High Listening Ability Levels. *Education Resources Information Center (Eric)*, USA.
- Ely, C. M. (1995). Tolerance of Ambiguity and the Teaching of ESL. In J. M. Reid (Ed.), *Learning Styles in the ESL/EFL Classroom* (pp. 87-95). New York: Heinle & Heinle Publishers.
- Ganschow, L., & Sparks, R. (1996). Anxiety about foreign language learning among high school women. *The Modern Language Journal*, 80, 199–212.
- Ganschow, L., Sparks, R. L., Anderson, R., Javorshy, J., Skinner, S., & Patton, J. (1994). Differences in language performance among high-, average-, and low-anxious college foreign language learners. *The Modern Language Journal*, 78, 41–55
- Greaney, V., & Hegarty, M. (1987). Correlates of leisure- time reading. *Journal of Research in Reading*, 10, 3-20.
- Hansen, L. (1984). Field dependence-independence and language testing: evidence from six pacific island cultures. *TESOL Quarterly* 18, 311-24.
- Hansen, J., & Stansfield, C. (1981). The relationship of field dependent-independent cognitive styles and foreign language achievement. *Language Learning*, 31, 349-67.

- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58, 47–77.
- Hohensinn, C., & Baghaei, P. (2017). Does the position of response options in multiple-choice tests matter? *Psicológica*, 38, 93-109.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Lawrence Erlbaum Associates, Inc.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.
- Norton, R.W. (1975). Measurement of ambiguity tolerance. *Journal of Personality Assessment*, 39, 607-619.
- Pajares, F., & Schunk, D. H. (2001). Self-Beliefs and School Success: Self-efficacy, Self-concept, and School Achievement. *Perception*, 11, 239-266.
- Salehi, M, & Marefat, F. (2014). The effects of foreign language anxiety and test anxiety on foreign language test performance. *Theory and Practice in Language Studies*, 4, 931-940.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, 4, 27-41.
- Sellers, V. D. (2000). Anxiety and reading comprehension in Spanish as a foreign language. *Foreign Language Annals*, 33, 512 - 521.
- Shi, Y. Z., & Liu, Z. Q. (2006). Foreign language reading anxiety and its relationship to English achievement and gender. *Journal of PLA University of Foreign Languages*, 29, 59 - 65.
- Zheng, Y., & Cheng, L. (2018). How does anxiety influence language performance? From the perspectives of foreign language classroom anxiety and cognitive test anxiety. *Language Testing in Asia*, 8, 1-19.