

### Iranian EFL Teachers' Language Assessment Literacy (LAL) under an Assessing Lens

Mitra Janatifar<sup>1</sup>, S. Susan Marandi<sup>2\*</sup>

<sup>1</sup> *PhD candidate, Faculty of Literature and Languages, Alzahra University, Tehran, Iran*

<sup>2</sup> *Associate Professor, Faculty of Literature and Languages, Alzahra University, Tehran, Iran*

Received: 2017/12/31

Accepted: 2018/06/25

**Abstract:** Despite being trained in pre-service teacher education programs, most EFL teachers are underprepared when faced with language assessment-related activities. Part of the problem emanates from the fact that Language Assessment Literacy (LAL) as a construct has not been well defined by experts. The purpose of this study was to pinpoint the components of LAL in the Iranian EFL context using an adapted version of Fulchers' (2012) LAL survey with two types of constructed and closed response items. The participants were 280 English language teachers from seventeen different provinces in Iran. Exploratory and confirmatory factor analyses and cross validation were used to define LAL as a construct. Furthermore, qualitative data analysis procedures were employed to analyze the data obtained from constructed response items. The results indicate that LAL in the Iranian context is comprised of four factors, namely: test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects (which mainly includes social and ethical aspects of language testing/assessment), and reliability and validity. Furthermore, the results show that the EFL teachers in this study believe that besides the theoretical issues of assessment, they should also receive hands-on skills-based instruction in language assessment. These results can have direct implications for future teacher education programs with the aim of enhancing EFL teachers' LAL.

**Keywords:** Language Assessment Literacy, Factor Analysis, Validation, Teacher Education.

---

\* Corresponding Author.

## Introduction

Assessment can play a pivotal role in promoting learning and also policy making. The link between instruction and assessment is so strong that "the quality of instruction in any ... classroom turns on the quality of the assessments used there" (Stiggins, 1999, p. 20). Stiggins (1991) was the first scholar who attracted everyone's attention to the importance of assessment literacy by making a strong confession that "We are a nation of assessment illiterates." More than a decade later, the same concern was raised by Popham (2004), who termed the lack of appropriate training in assessment as "professional suicide" (p. 82). Enhancing teachers' language assessment literacy can highly benefit instruction. Teachers who assess their instruction, for example, can realize their students' needs, monitor learning and instructional processes, diagnose student learning difficulty, and confirm their learning achievement (Gronlund & Linn, 1990). Despite such a great importance attached to assessment in most educational settings, research literature demonstrates that many teachers feel that they are not adequately prepared to deal with assessment issues and that they need some assistance in implementing various classroom assessments and in making assessment-related decisions (Mertler, 1999; Mertler & Campbell, 2005); therefore, there is a need for more emphasis on language assessment literacy in teacher education programs. Language teachers who are assessment literate can enhance the quality of their instruction and respond to their students' instructional needs more effectively.

Although Language Assessment Literacy (LAL) "is derived from the generic *assessment literacy* concept which refers to the knowledge and skills required for performing assessment related actions" (Inbar-Lourie, 2013a, p. 1; emphasis in original), the problem is that "there are still many uncertainties as to the nature of this concept particularly with regard to the relative weight of the literacy components, that is, which is essential, negligible, both essential and negligible, or even dispensable" (Inbar-Lourie, 2013a, p. 3). Clarifying LAL as a concept becomes crucial when prioritizing content for assessment courses or programs intended to enhance the acquisition of assessment literacy principles (Brindley, 2001; Brown & Bailey, 2008; Inbar-Lourie, 2008a).

In Iran, Language Assessment Literacy (LAL) is pretty much underexplored. However, there are two previous research projects related to Iranian EFL teachers' assessment literacy which merit mention here; in one study, test characteristics, test development, and testing language components/skills were found to be the top three areas that Iranian EFL teachers felt they needed to improve (Farhady & Tavassoli, 2015). In another study, the relationship

between Iranian EFL teachers' assessment literacy and washback effect was analyzed. The current study, however, aims at determining the components of LAL as a construct using both quantitative (confirmatory and exploratory factor analyses and cross validation) and qualitative data analysis procedures. The results can be used by policy makers and teacher educators in organizing and developing more effective EFL teacher education programs in the future in Iran.

## **Review of Literature**

### **Language Assessment Literacy (LAL)**

Despite the increasing importance of LAL, a major concern which still remains is determining a basic core of literacy in language assessment. This would include a range of skills related to test production, test score interpretation and use, and test evaluation and also the roles and functions of assessment in education and society (Inbar-Lourie, 2013b). LAL may be defined as:

The knowledge, skills and abilities required to design, develop, maintain or evaluate large-scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals (Fulcher, 2012, p. 125).

Melone (2013) put instructional issues at the core and defined LAL as "language instructors' familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language" (p. 329). On the other hand, Scarino (2013) emphasized the central role of teachers in assessment; therefore, LAL has also been defined as "the assessment of student achievements, teacher knowledge, understanding and practices of assessment" (p. 310). Finally, LAL may be regarded as a repertoire of competences used for understanding, evaluating and creating language tests and analyzing test data (Phill & Harding, 2013).

### **The significance of Language Assessment Literacy (LAL) in teacher development programs**

According to Hill (2017), there are two reasons for the growing recognition of the importance of language teachers' LAL: the increasing use of assessment for accountability purposes with the devolution of responsibility for assessment to classroom language teachers, and the shift in Classroom-Based Assessment (CBA) from assessment of learning to assessment for learning. Regarding the significance of language teachers' assessment literacy, Scarino (2017) believed that the global movement of people has caused communication to be increasingly multilingual and multicultural and this has caused language teachers to embrace a shift from a communicative to an intercultural orientation to language learning. This change in orientation of theoretical constructs affects the purposes of teaching, learning and assessment and therefore, "a sophisticated form of assessment literacy for teachers of languages becomes crucial" (p. 21). Tsagari and Vogt (2017) believed that the recent focus on CBA necessitated the professionalization of EFL teachers in relation to language assessment issues and topics. They also called for constant professional development for language teachers, in order to keep them updated with the challenges in classroom-based foreign language. Moreover, the movement from an era of assessment which entailed "comparing students with other students based on achievement to an era when we compare student performance to pre-set standards" (Stiggins, 2006, p. 3) also puts more responsibility on the shoulders of teachers, who need to deal more seriously with LAL. Although LAL as a concept seems to be associated more with language testers (Popham, 2009; Malone, 2013), the need is felt to develop LAL for teachers informed by "a holistic approach that goes beyond a mere knowledge-based concept of LAL" (Tsagari and Vogt, 2017, p. 43).

Malone (2013) dealt with the gap which exists between the ways in which the content of language assessment courses are viewed by language testing experts and language teachers. Jeong (2013) approached the same issue from course instructors' point of view who are either language testing experts or non-language testers. She found out that while the former group primarily emphasize technical issues, the latter put more emphasis on classroom assessment practices in language testing/assessment courses. Apart from the instructors' point of view, LAL can also be approached from the viewpoint of other stakeholders such as administrators, for example, when making admission decisions related to IELTS English proficiency test (O'Loughlin, 2013). In another study, Phil and Harding (2013) demonstrated how policy makers' lack of understanding of language

testing/assessment issues, as well as assessment tools and the purposes they serve can lead to meaningful misconceptions. When dealing with LAL, it is important to find a balance between theoretical aspects and practical know-how besides taking ethical issues into consideration; moreover, professional beliefs and practices within local communities of practices including language teachers, course instructors, university administrators and policy makers must also be accounted for (Taylor, 2013).

"Teacher assessment literacy has only recently found its way into the agenda of language assessment community"(Razavipour, Riazi, & Rashidi, 2011, p. 156) and this calls for more research to be done, with the aim of elaborating the concept of language assessment literacy and its constituent elements, especially in our context. There were two previous research projects related to Iranian EFL teachers' assessment literacy. Based on the findings of one of these studies, test characteristics, test development, and testing language components/skills were the top three areas that Iranian EFL teachers felt they needed to improve and were considered as part of LAL construct (Farhady, & Tavassoli, 2015). In another study, the relationship between Iranian EFL teachers' assessment literacy and washback effect was analyzed. The results of the latter study indicated that English language teachers' assessment literacy levels were unsatisfactory and far from researcher' expectations. "The extremely low assessment literacy of EFL teachers observed calls for a thorough overhauling of both pre-service and in-service assessment training courses"(Razavipour, Riazi, & Rashidi, 2011, p.160).

This study aimed at defining the nature of LAL in the Iranian context based on EFL teachers' beliefs. Therefore, the following research question was put forward:

- What components do Iranian EFL teachers believe must be included in a language testing/assessment course?

## **Methodology**

In order to fulfill the purposes of this phase of the study, a mixed-method research methodology was used to specifically answer the following research question: What components do Iranian EFL teachers believe must be included in a language testing/assessment course?

## ***Participants***

In order to answer the research question in this study, 280 Iranian English language teachers who came from seventeen different provinces in the country participated in this study.

Convenience sampling which involves "the selection of the most accessible subjects" (Marshall, 1996) was used in this study. The participants completed an adapted version of the Fulcher's (2012) LAL survey (Appendix A). Convenience sampling, which involves "the selection of the most accessible subjects" (Marshall, 1996) was used in this study. The demographic information of the participants can be found in Table 1.

**Table 1. Demographics of the Participants**

|                            | <i>Categories</i>    | <i>N</i> |
|----------------------------|----------------------|----------|
| <i>Gender</i>              | Male                 | 85       |
|                            | Female               | 195      |
| <i>Age</i>                 | 21-25                | 110      |
|                            | 26-30                | 90       |
|                            | 31-35                | 53       |
|                            | 36-40                | 15       |
|                            | 41-45                | 7        |
|                            | 51-55                | 1        |
|                            | Missing              | 4        |
| <i>Education level</i>     | High school graduate | 7        |
|                            | BA                   | 93       |
|                            | MA                   | 148      |
|                            | PhD                  | 17       |
|                            | Other                | 14       |
|                            | Missing              | 1        |
| <i>Teaching Experience</i> | 0-5 years            | 152      |
|                            | 5-10 years           | 65       |
|                            | 10-20 years          | 27       |
|                            | More than 20 years   | 5        |
|                            | Missing              | 31       |

### ***Data Collection and Analysis***

According to Fulcher (2012), the LAL survey deals with the issue of "What are the assessment needs of language teachers?" (p.118). However, due to the research context, some slight modifications were made in the demographic section of Fulcher's (2012) survey; for example, the question "Which is your home country?" was eliminated, since all the participants were Iranian. The researchers used Fulcher's (2012) LAL questionnaire since it was the only valid LAL questionnaire existent at the time of the study. The Cronbach's alpha in Fulcher's study was reported to be 0.93.

The LAL survey includes three sections and can be found in Appendix A. The first part of the survey is one closed-response item (question 3) which required respondents to decide which topics should be included in a language testing/assessment course. The part of the survey which included one question (question 3) was measured on a 5-point scale from

*unimportant to essential*; the responses were analyzed quantitatively using confirmatory and exploratory factor analyses and cross validation. The second part of the survey included seven constructed response items; questions 1, 2, 4, 5, 6, 7, and 8. They dealt with some language assessment related issues such as the number and quality of language testing and assessment courses that the respondents had taken, language testing and assessment books they had studied, topics they considered essential in a practical language assessment book, how they rated their own language assessment knowledge, and further comments they might have had on what should be included in a book on language assessment. The data were collected from May to September 2015 from 280 ELT teachers in Iran and were gathered both electronically (23 participants) and in a face-to-face manner (257 participants). All the factor analysis assumptions including normality, linear relations, factorability, and sample size were checked prior to performing them.

The research question was partially answered using the data gathered from one closed-response item (question 3) in the LAL survey. In this question, respondents were required to decide which topics should be included in a course on language testing. Data gathered from this item was measured on a 5-point scale from unimportant to essential. Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and cross validation were used to analyze data gathered from responses to this question. The exploratory phase determined the underlying factor structure of LAL and the confirmatory phase reassured the researchers that the relationship observed between factors and their underlying latent construct LAL is plausible. First, in the EFA phase, the data was entered in to the *Statistical Package for Social Sciences* (SPSS, version 19) software and four LAL factors and their corresponding descriptive values were identified. Then, in the CFA phase, the results of the EFA were entered in Linear Structural Relations software (LISREL, version 8.71) to make sure that the relationship observed between factors extracted at the EFA and their underlying latent construct LAL does in fact exist. The overall reliability of the third question in the LAL survey is reported by Fulcher himself to be as high as .93. Cronbach's alphas for each of the four factors comprising LAL in Fulcher's study were test design and development (.89), large-scale standardized testing (.86), classroom testing and washback (.79), validity and reliability (.94). In order to assess the predictive performance of the model achieved in the EFA and CFA phases, cross validation was also performed. In the cross validation study, first the participants were divided into two equal groups (138 participants in each sample) and the factor analysis was run for each group separately.

The research question was also partially answered by data gathered from question 1 (i.e., When you last studied language testing, which parts of your course did you think were most relevant to your needs?) and question 2 (i.e., Are there any skills that you still need?) in the LAL survey. These two questions were analyzed qualitatively based on the procedure suggested by Miles and Huberman (1994, pp. 58-69) and the results were presented using descriptive statistics. Firstly, the researchers constructed a data-coding matrix comprising of the four factors obtained in the EFA and their respective items. Next, the elements in the matrix were used to analyze the data obtained from constructed response questions in the second section of the LAL survey and the results of the data obtained from questions 1 and 2 were presented descriptive statistics. For example, the entire coding matrix for the first factor is demonstrated below:

- A. Test Design and Development
  - 1. Writing test specifications/blueprints
  - 2. Writing test tasks and items
  - 3. Evaluating language tests
  - 4. Interpreting scores
  - 5. Test analysis
  - 6. Selecting tests for your own use

Questions 9 to 14 and question 8 in the LAL survey were related to participants' demographic information and were analyzed by using descriptive data analysis. Questions 4 to 7 were analyzed qualitatively by using content analysis.

## Results and Discussion

The research question was answered based on the data gathered from participants' responses to questions 1, 2, and 3. Exploratory Factor Analysis (EFA) was used to analyze the data gathered from participants' responses to question 3 in the LAL survey at this stage of the study. All the factor analyses assumptions including normality, linear relations, factorability, and sample size were checked prior to performing factor analyses procedures. EFA can be described as the "orderly simplification of interrelated measures. By performing EFA, "the underlying factor structure is identified" (Shur, 2006, p. 1). EFA "can be useful for refining measures, evaluating construct validity, and in some cases testing hypotheses" (Conway & Huffcutt, 2003, p.147). The Cronbach's alpha in this study was .83 (compared to .93 in Fulcher's study). The Kaiser-Meyer-Olkin measure of sampling adequacy was also suitably high (.75), since according to Williams, Brown, and Onsmann, (2012), a KMO index higher



than 0.50 is considered suitable for factor analysis studies. According to Brown (2001) since both Cronbach's alpha and Kaiser-Meyer-Olkin measure of sampling adequacy were within a fairly acceptable range, varimax rotation should be used in EFA to analyze the data. The results of the EFA can be found in Appendix B with items listed in the left hand column. As the table demonstrates, the eigenvalues which emerged for the four factors were all greater than one, accounting for 51.21 % of all the reliable variance: 26.33% for Factor one, 9.62 % for Factor two, 8.92 % for Factor three, and 7.36 % for Factor four. Based on the EFA results, the reliability and descriptive statistics for the four factors obtained in the current study are as follow in Table 2.

**Table 2.** *Reliability and Descriptive Values for the Four Factors based on the EFA Results*

| Factor  | Cronbach's $\alpha$ | M    | SD  | SE   |
|---|---------------------|------|-----|------|
| Test design and development                               | .75                 | 4.03 | .9  | .51  |
| Large-scale standardized testing and classroom assessment | .7                  | 3.76 | .94 | .57  |
| Beyond-the-test aspects                                   | .65                 | 3.79 | .98 | .569 |
| Validity and reliability                                  | .76                 | 4.1  | .87 | .51  |
| Total   | .83                 |      |     |      |

Based on the EFA results, Iranian EFL teachers' LAL as a construct appears to be comprised of four factors namely: test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects (which was mainly related to social and ethical aspects of language testing/assessment), and validity and reliability. The results of this study also corroborate the findings of another study in which test characteristics, test development, and testing language components/skills were identified as the top three areas that EFL teachers felt they needed to improve (Farhady & Tavassoli, 2015). However, some differences were observed in the pattern in which items were loaded on factors in the current study and in Fulcher's (2012) study. These differences can be found in Table 3.

**Table 3.** *Comparison between the Present Study and Fulcher's Study*

|                 | <i>The present study</i>                                  | <i>Fulcher's study</i>           |
|-----------------|---|----------------------------------|
| <b>Factor 1</b> | Test design and development                               | Test design and development      |
| <b>Items</b>    | D, F, E, H, I, G  | D, F, E, C, B, M                 |
| <b>Factor 2</b> | Large-scale standardized testing and classroom assessment | Large-scale standardized testing |
| <b>Items</b>    | Q, P, N, O, R   | Q, P, N, H, G, V, W, L           |
| <b>Factor 3</b> | Beyond-the-test aspects                                   | Classroom testing and washback   |
| <b>Items</b>    | V, U, W, C  | S, R, O, U, T, I                 |
| <b>Factor 4</b> | Validity and reliability                                  | Validity and reliability         |
| <b>Items</b>    | J, K, L   | J, K                             |

With the aim of better canvassing LAL as a construct, CFA was used to verify the factor structure obtained in the exploratory phase. CFA "allows the researcher to test the hypothesis that a relationship between observed variables and their underlying latent constructs exists" (Shur, 2006, p. 1). CFA is generally used as an analytic tool with the purposes of developing and refining measurement instruments, assessing construct validity, identifying method effects, and evaluating factor invariance across time and groups (Brown, 2015). The four-factor structure model (test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects, and validity and reliability) obtained in the EFA phase was used as the initial model to perform CFA to choose a final model with adequate fit. Although what should be reported in CFA is not universally agreed upon, there is some unanimity among researchers who have addressed this issue (e.g., Barrett, 2007; Bentler, 2007; McDonald & Ho, 2002; Thompson, 2004). In order to test a priori hypotheses about relations between observed variables and latent variable(s) in CFA, first some model estimates are derived and then the researcher needs to evaluate some model fit indices.

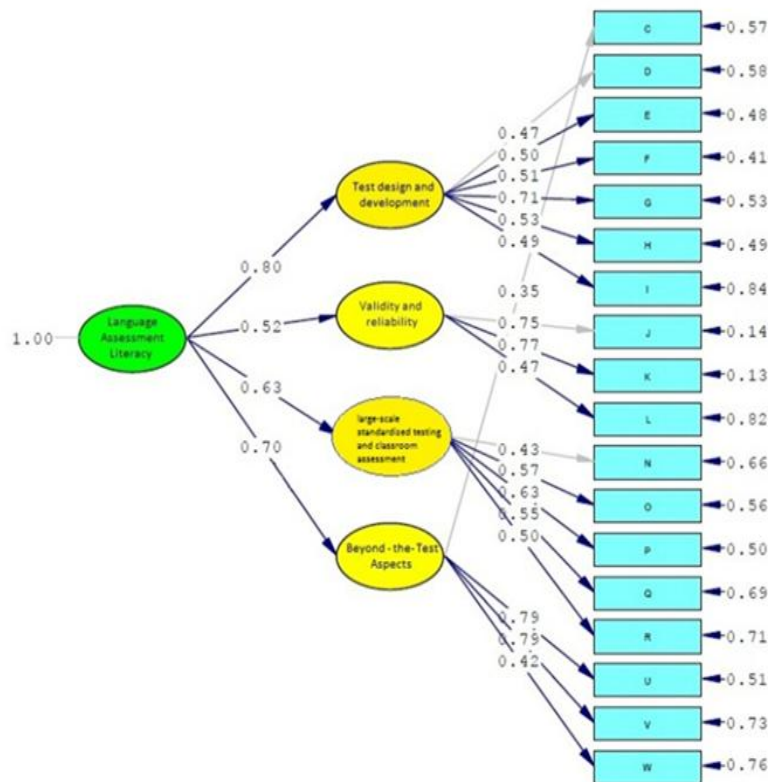
Apart from Chi-square, some other ancillary indices of global fit such as the goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), the Comparative Fit Index (CFI), the Root-Mean-Square Error of Approximation (RMSEA), Normed Fit Index (NFI), and Non-Normed Fit Index (NNFI) must also be evaluated (Jackson, Gillaspay Jr, & Purc-Stephenson, 2009). Regarding the cutoff rates for the above mentioned indices, it has been suggested that when  $N \geq 250$  a cutoff value close to 0.96 for CFI, the conventional cutoff Chi-P value of 0.05, and a cutoff value close to 0.06 for RMSEA are acceptable (Yu, 2002). Some researchers have suggested  $NNFI \geq 0.95$  and  $NFI \geq 0.95$  as the threshold; however, due to their non-normed nature, recommendations as low as 0.80 as a cutoff have been preferred; for GFI and AGFI, the acceptable range is between 0 and 1 (Hooper, Coughlan & Mullen, 2008). The results obtained for some of the goodness-of-fit indicators in the CFA phase of the study can be found in Table 4.

**Table 4.** *Fit Indices of the Four-Factor Model of Language Assessment Literacy (LAL)*

| <b>Model</b> | $\chi^2$ | <i>df</i> | $\chi^2/df$ | <i>RMSEA</i> | <i>CFI</i> | <i>AGFI</i> | <i>NNFI</i> | <i>GFI</i> | <i>NFI</i> |
|--------------|----------|-----------|-------------|--------------|------------|-------------|-------------|------------|------------|
|              | 338.99   | 131       | 2.58        | 0.076        | 0.91       | 0.84        | 0.90        | 0.88       | 0.87       |

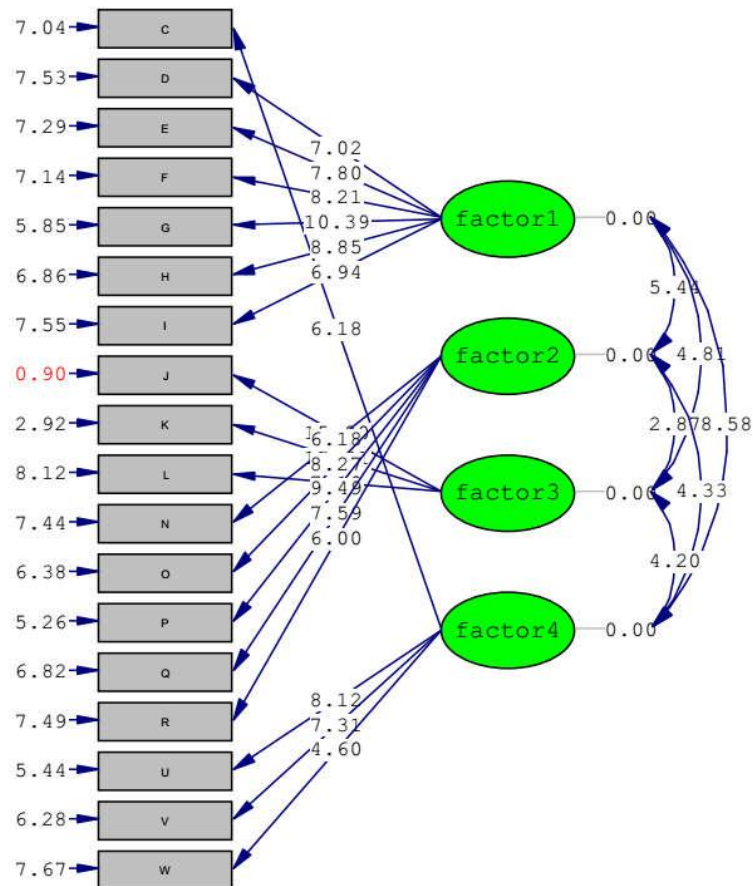
*Note.*  $\chi^2$  =chi-square; *df* =Degree of Freedom; *RMSEA*=Root Mean Square Error of Approximation; *CFI*= Comparative Fit Index; *AGFI*= Adjusted Goodness-of-Fit Index; *NNFI*= Non-Normed Fit Index; *GFI*=Goodness of Fit Index; *NFI*= Normed Fit Index.

The Chi-square calculated at this stage was significant [ $\chi^2 (131) = 338.99, p > .05$ ]; one reason for this significant  $p$  value is that  $\chi^2$  value is sensitive to the sample size (Tabachnick & Fidell, 2013). Therefore, the value for  $\chi^2/df$  was calculated and it was 2.58 which according to Loo & Loewen (2004) is within the satisfactory level ( $2 < \chi^2/df < 5$ ). Therefore, the above mentioned four-factor model was the final model for LAL construct (Factor 1: items G, F, E, D, H, I; Factor 2: Items P, Q, R, O, N; Factor 3: items V, U, W, C; and Factor 4: items J, K, L). The four-factor structure model of LAL obtained using CFA, factor loadings and correlations between factors can be found in Figure 1.



**Figure 1.** *Confirmatory Factor Analysis (CFA) Results*

In order to assess the predictive performance of the model achieved in the EFA and CFA phases, a cross validation study was also performed. First, the participants were divided into two equal groups (138 participants in each sample) and the factor analysis study was run for each group separately. The cross validation results for the first half of the participants group can be found in Figure 2.



Chi-Square=291.18, df=129, P-value=0.00000, RMSEA=0.096

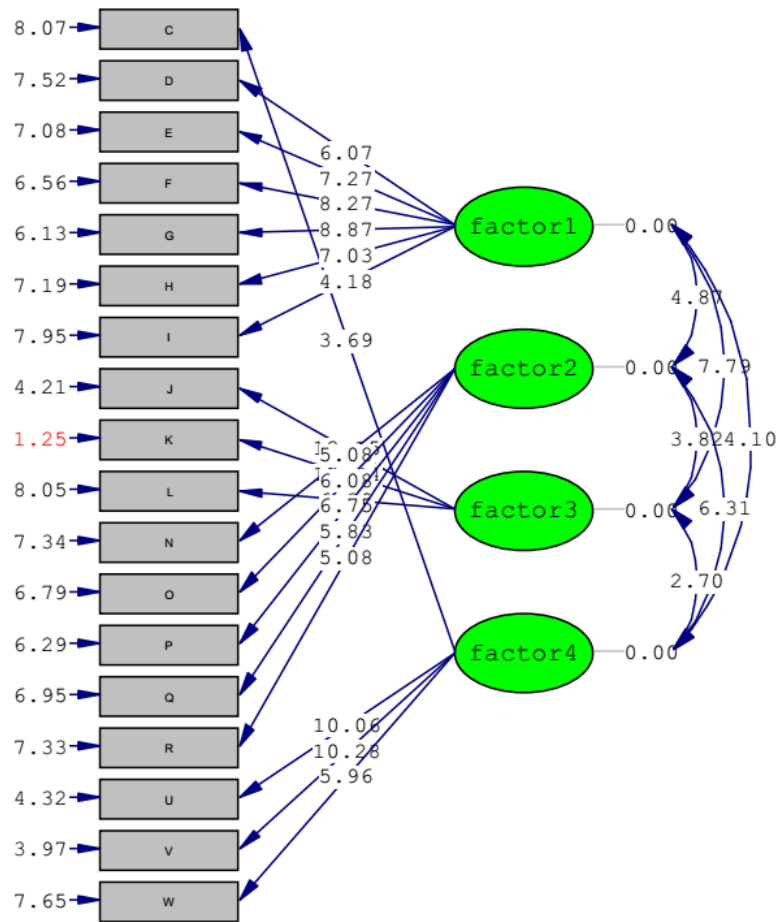
**Figure 2.** Cross Validation Results for the First Half of the Participant Group

As the figure demonstrates, the same four factors obtained in the original model also exist in the cross validation analysis of the first half of participants. The factor indices in the cross validation study on the first half of the participants are all significant and can be found in Table 5.

**Table 5.** Fit Indices for the First Half of the Participants Group

| GFI | CFI | NNFI | NFI | RMSEA |
|-----|-----|------|-----|-------|
| .87 | .89 | .87  | .83 | .096  |

The cross validation results for the second half of the participants group can be found in Figure 3.



Chi-Square=419.01, df=129, P-value=0.00000, RMSEA=0.128

**Figure 3.** Cross Validation Results for the Second Half of the Participants Group

The factor indices in the cross validation study on the second half of participants are all significant and can be found in Table 6. The results of the cross validation study confirm the four factor structure obtained in the original study here.

**Table 6.** Fit Indices for the Second Half of the Participant Group

| GFI | CFI | NNFI | NFI | RMSEA |
|-----|-----|------|-----|-------|
| .75 | .72 | .67  | .67 | .13   |

The high level of factor loading of the first factor (Test design and development) indicates that most of the participants believed that in addition to a thorough coverage of the theoretical principles of language assessment, more emphasis must be put on practical aspects and hands-on activities in language assessment courses.

Moreover, some items within factors were emphasized by participants more than others. For example, out of the six items which loaded on the first factor, item G,

(i.e., Interpreting scores) was considered by almost half of participants (49.2%) as essential. Out of the five items loaded on the second factor, item O (i.e., Classroom assessment) was considered by 37.7% of participants to be essential. Out of the four items loaded on factor three, item C, (i.e., Deciding what to test) received the highest percentage of 49.4% to be essential compared to other items. Out of the three items loading on factor four, item K (i.e., Validation) was considered by half of all participants (50.4%) to be essential.

In order to have a better understanding of factors and their corresponding items, educational level and teaching experience were used as the two axes based on which the gathered data were analyzed. Regarding the first factor, while 48.2% of BA level participants considered "Writing test tasks and items" (item E) to be "fairly important" it was considered by half of the PhD level participants (53.3%) as essential. "Classroom assessment" item O in factor two, was considered only by 37.9% of MA participants as "essential"; however, the same item was regarded by the majority of PhD level participants (60.0%) to be essential. This shows that despite moving to high academic levels, most PhD level participants still feel the need to learn more about classroom assessment. Interestingly, the "Reliability" item which loaded on the fourth factor received almost equal attention by all three educational levels of BA (45.2%), MA (54.3%) and PhD (46.7%) to be regarded as essential. However, the participants considered "Validation," which was another item of factor four, to be essential progressively as their educational level advanced (14.8% for BA, 58.9% for MA and 73.3% for the PhD level).

Besides educational level, teaching experience was another grid used for data analysis. "Writing test tasks and items" which loaded on factor one, was considered to be "essential" by only 34.5% of participants who had 0-5 years of teaching experience; however, for those with more than 20 years of teaching experience, it was 60.0%. "Ethical considerations in language testing" which was an item loaded on factor three was considered "essential" only by 28.0 % of participants with 0-5 years of teaching experience; however, it was regarded as "essential" by almost half of the participants with 10-20 years of teaching experience. In fact, this finding indicates that attention to ethical issues in language testing increases with teaching experience.

Apart from question 3 in the LAL survey, data gathered from participants' responses to questions 1 and 2 were also analyzed using a procedure suggested by Miles and Huberman (1994). Based on this procedure, the factor structure obtained at the factor analysis phase of the study was used to analyze open-ended questions 1 and 2. Based on this analysis, 89% of

the respondents regarded test design and development as a necessary skill which needs to be added to the language testing and assessment courses. It was followed by 5 % for reliability and validity and 4% for large-scale standardized testing and classroom assessment. The remaining 2% were miscellaneous.

Putting both forms of data analysis procedures results together which are factor analyses and the procedure suggested by Miles and Huberman (1994), it can be concluded that most participants emphasized the inclusion of more practical aspects in their language testing/assessment courses.

According to Scarino (2017), the need is felt to develop LAL programs for teachers informed "by a holistic approach that goes beyond a mere knowledge-based concept of LAL" (p. 43). It was also revealed that some issues in language testing/assessment such as "Writing test tasks and items," "Classroom assessment," and "Validation" were given more importance by participants as they moved to higher educational levels. "Ethical considerations in language testing" and "Writing test tasks and items" were also progressively given more importance as participants' teaching experiences increased. It then becomes crucial that language testing/assessment courses must be launched based on who the participants and what their needs are. According to Inbar-Louri (2008a), this target setting must be done ahead of launching the course. Vogt and Tsagari (2014) believe that the emphasis on test design and development "is prevalent only if traditional forms of testing are used, rather than more recent, informal, or alternative forms of assessment." Therefore, besides "Test design and development," the practical aspects of language testing/assessment which should be the focus of language testing/assessment courses, must also include more emphasis on alternative forms of assessment; this way we can witness a transition from a traditional approach to language testing/assessment to a more modern one.

## Conclusion

This research aimed at illuminating language testing/assessment needs of Iranian EFL teachers. Hence, defining the concept of Language Assessment Literacy (LAL) as a construct in the Iranian context was the main aim of this research. An adapted version of Fulcher's (2012) survey with two types of constructed and closed response items was used in this study. Based on the EFA, CFA, and the cross validation results, LAL is comprised of four factors, namely: test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects (which mainly includes social and ethical aspects of

language testing/assessment), and reliability and validity. The first factor (Test design and development) accounted for 26.33% of all the reliable variance in LAL construct. It can be concluded that the EFL teachers of the current study believe that besides a thorough coverage of theoretical issues, it is necessary to pay special attention to hands-on activities and skills-based approaches to language assessment. Delineating the components of EFL teachers' LAL can have direct implications for future EFL teacher education programs. Each of the four aforementioned factors which comprise LAL must be accounted for by policy makers and teacher educators and have their fair share in both in-service and pre-service teacher education programs across the country.

The study suffered from certain limitations; first, the fact that participants voluntarily took part in the study and answered the LAL survey questions indicates that they tacitly believed in the importance of LAL and the great role it played in their teaching career. The ideas and/or reasons of those who did not have such a belief was not accounted for in this study. Second, the analysis of the data obtained from the constructed response items in the LAL survey depended on the items included in the closed response item in the LAL survey. This could have caused possible limitation in the way language testing/assessment topics and their subcategories were put together. Third, the convenience sampling procedure adopted by the researchers in this study resulted in the majority of participants being MA holders; making thorough cross comparison analyses among the participants would have been easier and more valid if the number of participants in different educational levels were equal.

## References

- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824.
- Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825-829.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing, 18*(4), 393-407.
- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge, UK: Cambridge University Press.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing, 25*(3), 349-383.



- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational research methods*, 6(2), 147-168.
- Farhady, H. & Tavassoli, K. (2015). *EFL teachers' professional knowledge of assessment*. Paper presented at the 37<sup>th</sup> international LTRC conference on From Language Testing to Language Assessment: Connecting Teaching, Learning, and Assessment (March, 18-20). Toronto, Canada.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Gronlund, N. E., & Linn, R. L. (1990). *Constructing objective test items: multiple-choice forms. Measurement and Evaluation in Teaching*. New York: MacMillan.
- Hill, K. (2017). Language Teacher Assessment Literacy—scoping the territory. *Papers in Language Testing and Assessment*, 6 (1), iv-vii.
- Hooper, D., Coughlan, J. and Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6(1), pp. 53 – 60.
- Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Inbar-Lourie, O. (2013a). Language assessment literacy. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 2923–2931). Oxford: Blackwell.
- Inbar-Lourie, O. (2013b). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307.
- Jackson, D. L., Gillaspay Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods*, 14(1),6-23.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345-362.
- Loo, R., & Loewen, P. (2004). Confirmatory factor analyses of scores from full and short versions of the Marlowe–Crowne Social Desirability Scale. *Journal of Applied Social Psychology*, 34(11), 2343-2352.

- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Marshall, M. N. (1996). Sampling for qualitative research. *Family practice*, 13(6), 522-526.
- McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological methods*, 7(1), 64.
- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120(2), 285-285.
- Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "Assessment Literacy Inventory." *Online Submission*. Retrieved from <http://files.eric.ed.gov/fulltext/ED490355.pdf>
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: A sourcebook. *Beverly Hills: Sage Publications*.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402.
- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62(1), 82-83.
- Popham, W. J. (2009). Assessment literacy for teachers: faddish or fundamental? *Theory Into Practice*, 48(1), 4-11.
- Razavipour, K. Riazi, A. Rashidi, N. (2011). On the Interaction of Test Washback and Teacher Assessment Literacy: The Case of Iranian EFL Secondary School Teachers. *English Language Teaching*, 4(1), 156-161.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327.
- Scarino, A. (2017). Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. *Papers in Language Testing and Assessment*, 6(1), 18-40.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-39.
- Stiggins, R. J. (1999). Are You Assessment Literate?. *High School Magazine*, 6(5), 20-23.
- Stiggins, R. J. (2006). *Balanced assessment systems: Redefining excellence in assessment*. Portland, OR: Educational Testing Service. Retrieved from

- [http://media.ride.ri.gov/PD/FA/Formative\\_Assessment\\_Module\\_1\\_Lesson\\_4/story\\_content/external\\_files/Stiggins\\_Balanced\\_Assessment\\_Systems%202006.pdf](http://media.ride.ri.gov/PD/FA/Formative_Assessment_Module_1_Lesson_4/story_content/external_files/Stiggins_Balanced_Assessment_Systems%202006.pdf)
- Suhr, D. D. (2006). *Exploratory or confirmatory factor analysis?* In SAS Institute Inc., *Proceedings of the Thirty-first Annual SAS Users Group International Conference, Cary, NC: SAS Institute Inc.* (pp. 1-17). Cary: SAS Institute. Retrieved from [http://140.112.142.232/~PurpleWoo/Literature/!DataAnalysis/FactorAnalysis\\_SAS.com\\_200-31.pdf](http://140.112.142.232/~PurpleWoo/Literature/!DataAnalysis/FactorAnalysis_SAS.com_200-31.pdf)
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6<sup>th</sup> ed.). Boston: Pearson.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language testing*, 30(3), 403-412.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
- Tsagari, D., & Vogt, K. (2017). Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402.
- Williams, B., Brown, T., & Onsmann, A. (2012). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), 1.
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral dissertation, University of California Los Angeles).

## Appendix A

### Language Assessment Literacy (LAL) Survey

1. When you last studied language testing, which parts of your course did you think were most relevant to your needs?
2. Are there any skills that you still need?
3. Please look at each of the following topics in language testing.

For each one please decide whether you think this is a topic that should be included in a course on language testing.

|  | Unimportant | Not very important | Fairly important | important | essential |
|--|-------------|--------------------|------------------|-----------|-----------|
| A. History of Language Testing                 |             |                    |                  |           |           |
| B. Procedures in language test design          |             |                    |                  |           |           |
| C. Deciding what to test                       |             |                    |                  |           |           |
| D. Writing test specifications/blueprints      |             |                    |                  |           |           |
| E. Writing test tasks and items                |             |                    |                  |           |           |
| F. Evaluating language tests                   |             |                    |                  |           |           |
| G. Interpreting scores                         |             |                    |                  |           |           |
| H. Test analysis                               |             |                    |                  |           |           |
| I. Selecting tests for your own use            |             |                    |                  |           |           |
| J. Reliability                                 |             |                    |                  |           |           |
| K. Validation                                  |             |                    |                  |           |           |
| L. Use of statistics                           |             |                    |                  |           |           |
| M. Rating performance tests (speaking/writing) |             |                    |                  |           |           |
| N. Scoring closed-response items               |             |                    |                  |           |           |
| O. Classroom assessment                        |             |                    |                  |           |           |
| P. Large-scale testing                         |             |                    |                  |           |           |
| Q. Standard setting                            |             |                    |                  |           |           |
| R. Preparing learners to take tests            |             |                    |                  |           |           |
| S. Washback on the classroom                   |             |                    |                  |           |           |
| T. Test administration                         |             |                    |                  |           |           |
| U. Ethical considerations in testing           |             |                    |                  |           |           |
| V. The uses of tests in society                |             |                    |                  |           |           |
| W. Principles of educational measurement       |             |                    |                  |           |           |

4. Which was the last language testing book you studied or used in class? What did you like about the book? What did you dislike about the book?
5. What do you think are essential topics in a book on practical language testing?
6. What other features (e.g. glossary/activities etc) would you most like to see in a book on practical language testing?

7. Do you have any other comments that will help me to understand your needs in a book on practical language testing?

8. How would you rate your knowledge and understanding of language testing?

5 = very good

4 = good

3 = average

2 = poor

1 = very poor

9. And now, just a few quick questions about you. Are you male or female?

Female |  Male

10. What is your age range?

Under 20

21 – 25

26 – 30

31 – 35

36 – 40

41 – 45

46 – 50

51 – 55

56 – 60

61 – 65

Above 65

11. Please select your current educational level:

High School Graduate

BA degree

MA degree

Doctorate

Other

12. Which language do you consider your first language? What other languages do you speak?

13. Do you have experience teaching English? If "yes" please specify your experience:

0-5 years

5-10 years

10-20 years

more than 20 years

14. Where do you teach English?  Institute  University  High school

(You can select more than one option)

An online course in language assessment might be formed based on the data gathered from this survey; if you are interested in taking part in such a course, please provide the following information so that we can notify you later on. The information you provide will be strictly confidential and used only for research purposes.

Name:

Phone number:

E-mail

**Thanks**

## Appendix B

### Exploratory Factor Analysis Results

|                   | Component   |             |             |             | h <sup>2</sup> |
|-------------------|-------------|-------------|-------------|-------------|----------------|
|                   | 1           | 2           | 3           | 4           |                |
| SMEAN(G)          | <b>.717</b> | .113        | .106        | .165        | .541           |
| SMEAN(F)          | <b>.689</b> | .023        | .137        | .101        | .490           |
| SMEAN(E)          | <b>.675</b> | .073        | .094        | .093        | .455           |
| SMEAN(D)          | <b>.594</b> | .095        | .099        | .157        | .372           |
| SMEAN(H)          | <b>.576</b> | .140        | .030        | .340        | .365           |
| SMEAN(I)          | <b>.571</b> | .247        | .169        | -.265       | .485           |
| SMEAN(P)          | .124        | <b>.761</b> | .037        | .057        | .596           |
| SMEAN(Q)          | .077        | <b>.672</b> | -.019       | .185        | .490           |
| SMEAN(R)          | -.082       | <b>.640</b> | .151        | .204        | .458           |
| SMEAN(O)          | .270        | <b>.581</b> | .265        | -.074       | .484           |
| SMEAN(N)          | .186        | <b>.553</b> | .123        | -.036       | .355           |
| SMEAN(J)          | .194        | .065        | .076        | <b>.884</b> | .827           |
| SMEAN(K)          | .217        | .081        | .114        | <b>.876</b> | .832           |
| SMEAN(L)          | .105        | .261        | .295        | <b>.475</b> | .391           |
| SMEAN(V)          | .051        | .116        | <b>.873</b> | -.026       | .777           |
| SMEAN(U)          | .288        | .154        | <b>.679</b> | .115        | .579           |
| SMEAN(W)          | .032        | .088        | <b>.574</b> | .223        | .386           |
| SMEAN(C)          | .322        | .065        | <b>.422</b> | .045        | .287           |
| <b>Eigenvalue</b> | <b>4.74</b> | <b>1.73</b> | <b>1.60</b> | <b>1.32</b> | <b>9.39</b>    |