International Journal of Social Sciences (IJSS)
Vol.6, No.4, 2016

The Impact of Teachers' Training on the Reliability of Tests and Assessments in Governmental and Non-governmental Sections

Heydar Nosratzadeh[1]
Department of Foreign Languages, Damavand Branch, Islamic Azad University, Tehran, Iran

Received 20 October 2016
Revised 21 November 2016
Accepted 14 December 2016

Abstract: *Assessment is considered as one of the fundamental elements in the field of foreign language acquisition. In order for communication take place, adequate number of vocabulary is needed to be known by the learners. The salient role of vocabulary in the field of foreign language acquisition resulted in the publication of several hundreds of papers and dozens of books. Due to the dominant role of vocabulary in this area, this research investigated two variables considered to be effective in the learning and retention of second language vocabulary: word exposure frequency, and elaboration of word processing. Word exposure frequency that refers to the number of times the learners are exposed to a word in input was considered as one of the independent variables, and elaboration of word processing that refers to the quality and richness of processing new words in the textual input was considered as the second independent variable. Depth and richness of word processing were investigated based on the involvement load hypothesis offered by Laufer and Hulstijn (2001).*

Keywords: *Assessment, teachers, validity, reliability.*

Introduction

In language testing and assessment, there has been a shift from a psychometric, reductionist language testing paradigm to an edumetric, anti-reductionist language assessment paradigm. Wolf, Bixby, Glenn and Gardner (1991) remark that testing culture is related to employing tests/exams merely to determine achievements/grades while an assessment culture is related to using assessments to improve instruction and promote student learning. To Shepard (2013), a worldwide focus on the use of innovative assessment, such as performance-based and criterion-referenced assessments, has established questionable motives for teachers. As for the importance of language testing and assessment in second language acquisition, it may suffice to mention that based on previous research teachers normally spend a minimum of one-third of their instructional time on assessment-related activities (Bachman, 2014). The fact of the matter is that the provision of good instruction and student learning seem to be highly associated with the quality of assessments used in classrooms. For these reasons, as Earl (2013) explains, a major task of the teachers is to integrate their assessments with their instruction and students learning so as to meet the needs of the twenty-first century goals such as preparing students for lifelong learning skills.

The crux of the matter is that the 'what' and 'how' of language testing and assessment is not separated from the 'why' of testing and assessment. That is to say, apprehending the 'what' and performing the 'how' necessitates appreciation of the background and reasoning behind the actions taken, that is, the 'why'. Each of these aspects is derived from language-related matters and from general education and assessment and testing cultures. According to Davidson (2007), current assessment developments should be taken into account in the any debate on language assessment literacy, in particular the support for 'assessment *for* learning' (AfL) approaches in many parts of the world. A study conducted by Marshall and Drummond (2006) showed that only one-fifth of AfL lessons could increase their learning autonomy. To help teachers to understand the purpose of autonomy for all learners, it is worth examining what is meant by autonomy, how AfL practices can improve learner autonomy, and as well

[1] Email: nosratzadeh_14@yahoo.com

the importance of students seeking affiliation and a feeling of belonging within the classroom community of practice.The present study as an initial attempt describes a research project designed to collect exploratory and empirical data from language teachers as well as experts in language testing, and to figure out what learning needs teachers have in language testing and assessment. The intention is to implement the outcome of the needs analysis to give empirically derived content to the concept of assessment literacy and develop the design and provide the construction of new materials that can be used in language testing education programs in the context of Iran.

Statement of the Problem

There are researchers and teachers who emphasize the benefits of making teachers know the new testing and assessment paradigm (Marso & Pigge, 1993). Marso and Pigge (1993) believe that using innovative assessment techniques is, especially, useful in the area of language learning. However, few researches has investigated the usefulness of assessment literacy for language teachers. A major problem is that most of the teachers who conduct assessments for summative purposes use poorly constructed, objective paper and pencil tests (e.g., multiple-choice tests) that simply measure students low-level knowledge and skills (see Bol & Strage, 1996). It has been well documented that such poorly designed tests can lead to surface learning, and therefore produce a mismatch between classroom assessment practices and teaching/learning goals (Heritage, 2013).

If assessment is implemented appropriately, it offers learners, teachers and practitioners in testing useful information about student performance and the extent to which learning objectives have been attained (Stoynoff & Chapelle, 2005). As explained by Malone (2011), assessment and teaching can be integrated with each other and form an association in which assessment and teaching support and expand each other. Nevertheless, when language teachers do not have sufficient training in assessment, this relationship cannot develop. Point taken, the current study is an initial attempt to create and develop a general view of the assessment needs of language teachers in the Iranian context that can be used as a basis for the establishment of new educational materials through both closed- and constructed-response items and can provide a number of innovative design features that endorse teachers to express needs independently of the predigested response options. As such, it tries to fill a gap in the field, and offer further substantive definition to the construct of assessment literacy.

Significance of the Study

As for the effectiveness of assessment in education, assessment specialists discuss that if the ultimate aim of educational reform in the current era is to advance students' broader knowledge and skills, the effective task of the teachers is to implement a variety of assessment methods in evaluating students' learning, without considering whether the assessment has been carried out for formative (i.e., enhancing instruction and learning) and/or summative purposes (i.e., summing up achievement) (for more information, see Wiliam, 1998; Masters, 2013). Teachers are highly responsible to assess students using the most effective assessment methods with regard to the students' needs and situation. Plake (1993) rightly states that teachers devote 50 percent of their time on assessment-related activities; however, irrespective of the amount of time spent, classroom assessment is a highly effective teaching function due to the fact that it could contribute to every other teacher function.

The studies on assessment literacy could develop teachers' knowledge base of assessment literacy especially through pre-service programs by providing the teachers with specific course work, including assessment topics in content and methods courses, and offering real-world chances to enable teachers to implement what they have acquired. As Tao (2014) explains, possessing deeper knowledge and understanding of such a process will help teachers to better design a variety of assessment methods to enhance instruction and promote students' learning (i.e., formative purposes) and summarize students' learning achievements (i.e., summative purposes).

Studies on assessment literacy could pave the way for assessment culture to be an integral part of the culture of educational settings like schools. This aim can be achieved if the members of language policy and language teaching departments accept the beliefs and assumptions regarding the nature of assessment and its role in the learning process. To Inbar-Lourie (2008b), achieving assessment literacy

means that teachers must experience a "profound perception change" (p. 293) since they adopt that all knowledge and meaning is socially constructed (fulcher, 2014). The central issue is that teachers' knowledge of high quality classroom assessments is directly related to the students' ability of acquiring lifelong learning skills and/or higher-order thinking skills that, in turn, help them to accomplish the goals of educational reform and equip them with necessary skills for the twenty-first century. Thus, it is extremely important to better understand the nature of teachers' classroom assessment literacy development so that the professional in language testing and assessment in general, and English language teachers as the key agents in assessment practice, in particular can take effective remedies to approach the issues in a timely manner.

Theoretical Framework

The present study follows the guidelines from four theoretical framework explained including Fishbein and Ajzen's (2010) reasoned action theory, Davidson's (2001) agentic theory, Bandura's (1986) social cognitive theory and Vygotsky's (1978) sociocultural theory to investigate teacher classroom assessment literacy. Based on reasoned action theory (Fishbein & Ajzen, 2010), personal beliefs of the teachers are assumed to be a good predictor of their behaviors/ performances. Put it another way, personal beliefs or attitudes of teachers can strongly affect the ways in which they employ their classroom assessment. The present study also follows the guidelines proposed by Davidson's (2001) agentic theory based on which teachers are deemed to be essential and powerful agents of their own behaviors/performances. That is to say, teachers are the principal actors and contributors to the achievement and effectiveness of conducting classroom assessment. As such, teachers play a crucial role in implementing high quality assessments. Within social cognitive theory (Bandura, 1986), the key concept is self-efficacy that is regarded to affect every phase of the individuals' personal evolution from cognitive processes to affective ones. Self-efficacy aims at finding out if the teachers consider changing their assessment implementation, and if they have adequate motivation to apply what they have learned. To Bandura (1997), self-efficacy is referred to as the individual's perceived ability to perform a pedagogical task.

Research Questions

The present study is undertaken in two distinct phases: A quantitative phase and a qualitative phase since a mixed methods approach can offer additional benefits for the understanding and application of the present research work. In order to examine the assessment knowledge base and belief of the language teachers and the application of their assessment literacy in L2 learning context, the current study sets out to provide answers for the following questions:

1. What are the teachers' needs regarding their assessment literacy in Iranian context from experts' and teacher trainers' opinions?
2. To what extent are experts' opinions and teacher trainers' opinions different from teachers', themselves, opinions regarding their assessment literacy needs?
3. Is there any significant relationship between teachers' assessment training courses and their language assessment literacy?
4. Is there any significant relationship between teachers' assessment training courses and their perceptions of language assessment?

Research Hypotheses

Since research question one, two, and six are exploratory in nature, a priori research hypotheses cannot be formulated for them. Indeed, according to Holliday (2002), "producing, rather than testing, hypotheses are more often the outcome of qualitative research" (p. 35). Yet, the following null hypotheses were formed on the bases of the remaining research questions of the study:

- H_{01}: There is no significant relationship between teachers' assessment training courses and their language assessment literacy.
- H_{02}: There is no significant relationship between teachers' assessment training courses and their perceptions of language assessment.

- H_{03}: Teachers' scores on language assessment literacy could not be explained by: scores from teacher sense of reflectivity, scores from teachers' efficacy and scores from teacher perceptions of language assessment.

Definition of the Key Terms

Assessment Literacy

Stiggins (1991) suggested the term "assessment literacy" to describe if the teachers and educators know the difference between sound and unsound assessment. Similarly, Popham (2009) defined assessment literacy as the teachers' understanding of fundamental assessment-related principles and procedures that can impact their educational decision making. That is, assessment-literate teachers know the way to develop and/or select more suitable assessment tasks, employ a variety of assessment methods and interpret accurate assessment data, as well as know how to deal with any bias that may creep into their self-made assessment tasks.

Innovative Assessment Methods

McMillan (2014) mentioned that traditional or objective assessment tends to be associated with measuring lower-order thinking skills, while innovative or alternative assessment is associated with measuring higher-order thinking skills. Furthermore, alternative assessment is also referred to authentic assessment. Bachman and Palmer (1996) defined authentic assessment as a task involving in students' everyday lives. Furthermore, Newmann and Archbald (1992) refer to authentic achievement defined as student learning outcomes that have been assessed by authentic assessment tasks. In fact, they expanded the notion of authenticity by stating that the quality and use of assessment rely on the degree to which outcomes can represent worthwhile, appropriate and meaningful accomplishment.

Portfolio Assessment

A portfolio assessment includes the student work structured collection that documents students' application of knowledge and skill in a variety of authentic contexts. Portfolios generally require a variety of student work related to multiple standards or themes (e.g., reports, work samples, awards and certificates, career development documentation, self-reflection and evaluation pieces). It is in contrast to projects that typically require students to produce one product related to a few standards or themes. A more comprehensive view than projects of students' standards-based knowledge and skills can usually been provided by portfolio assessments. There are several portfolios examples in place today across the various levels of the education system. Portfolios have been found to be beneficial across subject areas and purposes, ranging from the statewide writing portfolio used and accountability system through several successful classroom-focused applications at the elementary and secondary education levels. Portfolios are also being used for purposes of professional certification. For example, the National Board for Professional Teaching Standards requires a teacher to successfully complete a portfolio in order to be certified as an accomplished teacher. While no state has endorsed portfolios as part of their adult education account- ability system, a number of adult education programs are beginning to develop and incorporate portfolios into classroom practice, including pro- grams that are using the EFF standards.

Performance-based Assessment

A great variety and range of assessment methods and tasks are included in performance-based assessment: projects, portfolios, performance tasks, computer simulations, scientific or mathematical inquiries, research reports, etc. To allow for more in-depth discussion, we consider three major types of performance-based assessment methods in this section: written scenarios, projects, and portfolios.

These assessment methods were selected as examples for many reasons. First, they have each been used widely and successfully in different educational programs across the nation. Second, because they differ in the scope of the task and the skills that they target, these assessment methods offer a glimpse of the range of possibilities associated with performance-based assessment. Finally, these assessment methods were selected due to their potential for reinforcing and measuring learning relative to the EFF standards.

Assessments that are composed solely of multiple-choice items, or that merely ask for definitions of terms and concepts, will provide no insights into the ability of candidates or teachers to actually apply the knowledge and skills within the assessment literacy domain. Indeed, for most of the domain, various forms of performance-based assessment, including observation and examination of artifacts, are the best—and we would say the only—effective means of measuring mastery. Students who have not been exposed to performance-based assessment methods will need a lot of preparation in order to be successful. The specific strategies for helping students succeed on on-demand scenarios differ somewhat from those needed for helping students succeed on projects and portfolios.

Dynamic Assessment

Teachers need to know how students are approaching, interpreting, and engaging in authentic literacy tasks to plan instruction. Teachers working with second-language students also need to know how they are using their two languages to make sense of the literacy tasks before them. It is my thesis that informal assessment measures, sometimes termed situated or dynamic assessment, can provide a more comprehensive profile of the second-language learners' literacy strengths and weaknesses than can formal assessment measures. Dynamic assessment is partly different and could be provided in both a formal and informal context (Lidz, 1987). As a construct, it refers to the opportunity to document what children can and cannot do with additional help. When it is used in a situated context, it allows teachers to document the progress children are making with and without support from themselves or other peers.

Research on Language Assessment Literacy

This section gives the body of exploratory research focused on the assessment knowledge base and personal beliefs about assessment that is thought to underpin classroom assessment literacy of teachers. It also discusses the research, which is still in its infancy stage, with regard to the relationship between teachers' classroom assessment literacy and assessment implementation. Fulcher (2012) conducted a need analysis study to collect empirical data from language teachers and discover what learning needs they have in language testing and assessment

Assessment Knowledge and Assessment Practice

The first underpinning classroom assessment literacy of teachers to be considered is their assessment knowledge base. Individual teachers' knowledge base has been acknowledged as a vital aspect that contributes to the success and effectiveness of assessment implementation (Bandura, 1997; Ajzen, 1991, 2005; Fishbein & Ajzen, 2010). It has also been widely reported that the quality of assessment has a relationship with the quality of instruction and students' learning (Tang, 1994; Boud, 2006; Biggs & Tang, 2007; Joughin, 2009; Earl, 2013). As such, there is a need for teachers to have a sound assessment knowledge base to enable them to implement high quality assessments. Assessment specialists have argued that the greater the assessment knowledge base teachers possess, the more capable they are of implementing quality assessments to enhance instruction and student learning (Stiggins, 1991a, 1995; Popham, 2006, 2009). Possessing an adequate assessment knowledge base can help teachers to have a better understanding of the process for conducting classroom assessment. Such a knowledge base will equip them with an appreciation of the assessment process. Within social cognitive (Bandura, 1989, 1997), reasoned action (Fishbein & Ajzen, 1975, 2010) and planned behaviours theories (Ajzen, 1991, 2005), a lack of requisite knowledge and/or skills of any individual teachers make them unable to carry out the intended tasks. In other words, if the teacher lacks knowledge and/or skills to implement the assessment, it is more likely that s/he fails to conduct the intended assessment and/or implements poor quality assessments.

Research Methodology

To testify the truth or falsity of the research hypotheses, the researcher first decides to collect and analyze data. The aims of the study are to 1) identify the teachers' needs regarding their assessment literacy in Iranian context from experts' and teacher trainers' opinions, 2) to tabulate the difference between teachers' opinions regarding their assessment literacy with those of experts and teacher trainers, 3) to examine the relationship between teachers' assessment training courses and their

language assessment literacy, 4) to examine the relationship between teachers' assessment training courses and their perceptions of language assessment, 5) to see what is the best predictor of teachers' language assessment literacy with respect to scores from teacher sense of reflectivity, scores from teachers' efficacy and scores from teacher perceptions of language assessment, and 6) to explore the teachers' reactions towards assessment training programs. To meet the objectives of the study, data was collected from various participants through different instruments during the six steps of the study.

Design of the Study

The present study makes use of both qualitative research and quantitative research because mixed method research yields a much more comprehensive result (Richards, 2003). During the first step of the study, using semi-structured interview, the researcher tries to categorize the commonalties among experts' opinions regarding teacher assessment literacy as well as the commonalities among teachers' trainers opinions, and in so doing, the researcher employs constant comparison analysis and classical content analysis. During the second step of the study, the study identifies the teachers' opinions regarding their needs and knowledge of their assessment literacy using open-ended questionnaire and interview and tabulate the frequency of the most commonly-cited factors using Descriptive Statistics and compare them with those of experts' and teacher trainers' opinions. During the third step of the study, the researcher, as an initial attempt in the context of Iran, designs a scenario-based inventory to measure teachers' knowledge of language assessment literacy. Then, the inventory is distributed to the teachers before and after treatment. Using Repeated Measured t-test, the difference between their knowledge and skills prior to and following the treatment is determined. During the fourth step of the study, the researcher, using t-test, tires to examine the difference between teachers' knowledge and skills of assessment literacy and their perceptions of assessment. During the fifth step of the study, the researcher, first, construct and validate the questionnaire of language assessment literacy using exploratory factor analysis and structural equation modeling. Then, using Multiple Regression, the study sets out to see how much of the variance of teachers' scores on language assessment literacy could be explained by their scores from teacher sense of reflectivity, teachers' efficacy and teacher perceptions of language assessment. During the last step of the study, the researcher explores teachers' reactions towards the provision of instruction in improving their knowledge and skills of new types of assessment procedures and techniques, using semi-structured interviews.

Part one: the Qualitative Phase

Sample

A criterion-based selection method, rather than a random one, was chosen as the sampling method. In this method, the researcher specified the criteria essential to the purposes of the study, and sought out participants fulfilling those specific attributes (Le Compete & Preissle, 1993). The criteria set in this phase were: a) Being an English teacher, b) Being a PhD graduate/candidate in English TEFL, TESL, Applied Linguistics or Testing and Assessment, and c) Having experience of teaching for at least five years. A pool of 40 experts in testing and assessment is participated in this study based on selection sampling. The researcher selects 20 famous global experts and 20 well-known local experts. Experts have more than five years of experience in teaching English language and have published papers in testing and assessment prestigious journals. They have received their PhD in TEFL, TESL, Applied Linguistics or Testing and Assessment. Moreover, the researcher selects purposefully 40 teacher trainers who are active in setting up educational workshops for pre-service and in-service teachers around Iran and they hold PhD in TEFL and have been engaged in testing and assessment in some ways as to teacher training workshops or as to their PhD dissertation. The experts and teacher trainers are asked to fill out the open-ended questionnaire designed by the researchers and also some of them are asked to participate in face-to-face interview with the researcher regarding the teachers' needs and knowledge of assessment literacy. Finally, a pool of 150 teachers participates in this phase of the study. The study sets out to carry out at different schools of Iran. The subjects participating in this study are teachers who are teaching English language in secondary schools. The teachers are selected from both big cities including Tehran, Tabriz, Isfahan, Shiraz, Mashhad, Sannandaj, Rasht and Ahvaz and from small cities around the country. The participants are asked to reply to the open-ended questionnaire, and then 20 of them are randomly interviewed by the researcher.

Instrumentation

Two major instruments are used to gather data in the qualitative phase of the study. The first instrument used in this study is semi-structured interview. To seek out the beliefs of the experts, teacher trainers and English teachers regarding what shapes the assessment needs, knowledge and skills of the teachers they are interviewed in face-to-face individual interviews with the researcher. The method of the interview selected for this study is "qualitative interviewing". This method "is based in conversation, with the emphasis on researchers asking questions and listening, and respondents answering" (Rubin & Rubin, 1995, as cited in Warren, 2002, p. 83). In this method of interview, the interviewees are considered as meaning makers, not passive conduits for retrieving information from an existing vessel of answers.

Furthermore, among the diverse forms of interviewing such as telephone or face-to-face interviews, individual semi-structured or in-depth interviews, and focus groups discussions (Kairuz, Crump & O'Brein, 2007) - the kind of the interview conducted in this study is that of semi-structured. In this type of interview, the whole process of interviewing changes throughout the continuum of highly-structured to highly unstructured in that the predetermined questions are not necessarily asked in a fixed order but rather in a more flexible manner.

Data Collection Procedure

First, following the guidelines proposed by Dörnyei (2007), the researcher opts for a simple questionnaire consisting of three open-ended statements distributed to the experts, teacher trainers, and teachers in order to find out teachers' language assessment literacy. To the best knowledge of the researcher, no study has been carried out in the context of Iran to identify the major factors to which students majoring in English translation refer to regarding their assessment literacy. Since qualitative research is in line with grounded theory and anti-positive philosophies, the kind of methodology employed in this study is that of interpretive approach to analyzing data gathered by means of a simple open-ended questionnaire. Participants are allowed to write as many reasons as they know and they are asked to reflect on their previous experience. The questions are written in Persian Language and the students are required to reply to the questions in Persian Language since using their native language, they can express themselves more easily and precisely. Their writings are then translated into English, analyzed, and categorized. Using a simple writing task for data collection is most likely the best way to get the students express their thoughts about the three issues mentioned above. The researcher permits greater freedom of expression since the range of possible answers cannot be previously anticipated. The writing task resembles an unstructured interview, in which subjects can externalize their thinking with very little interference. The pupils receive the instructions both orally and in writing. The pupils have approximately thirty minutes to write down their answers. Finally, some of the experts and teacher trainers as well as 20 teachers are randomly interviewed in face-to-face interaction with the researcher. Qualitative data are collected through interviews from these participants who had already filled out the questionnaires to accomplish the first and second steps of the study. Before starting each interview, the researcher briefly discusses the purpose of the interview and the use of the audio-recorder.

Data Analysis

The researcher of the present study opts for both constant comparison analysis and classical content analysis to analyze data. With reference to constant comparison analysis, he follows the three major stages of the constant comparison analysis (Strauss & Corbin, 1998). During the first stage (i.e., open coding), the data are chunked into small units. The researcher attaches a descriptor, or code, to each of the units. Then, during the second stage (i.e., axial coding), these codes are grouped into categories. Finally, in the third stage (i.e., selective coding), the researcher develops one or more themes expressing the content of each of the groups. With regard to classical content analysis, the researcher creates smaller chunks of the data, placing a code with each chunk. However, instead of creating a theme from the codes (as with constant comparison analysis), with classical content analysis, these codes then are placed into similar groupings and counted. The researcher, first, identifies whether each participant used a given code, then, he assesses whether each group used a given code, and finally, he identifies all instances of a given code. The researchers not only provides information regarding the

frequency of each code (i.e., quantitative information) but supplemented these data with a rich description of each code (i.e., qualitative information), which would create a mixed methods content analysis.

Part Two: the Quantitative Analysis

In this phase of the study, the researcher constructs and validates an inventory of teacher assessment knowledge and designs a questionnaire of language assessment literacy employing factor analysis to uncover the latent structure that underlined the large datasets. The researcher, indeed, aims at examining the teachers' knowledge of assessment prior and following the treatment to accomplish the foremost purpose of the research work.

Sample

The sample consists of 250 secondary school English teachers who have already responded to the open-ended questionnaire. The teachers are randomly selected from different schools around the country and they mostly have BA or MA degree in TEFL. They have already taken testing courses in their academic studies. The teachers are from both big cities and small cities, and they are selected from different schools including higher level schools such as schools for the gifted and ordinary schools such as public schools at deprived parts of the country.

Instrumentation

The first instrument used in this phase of the study is a scenario-based inventory of assessment that is designed and validated by the researcher. The inventory measures the teachers' knowledge regarding assessment procedures and techniques. The inventory is named Teacher Assessment Knowledge Inventory (abbreviated as TAKI) and can be used as an instrument for conducting further research in the domain of testing and assessment (Appendix B). The second instrument used is a questionnaire of assessment literacy that is called Language Assessment Literacy Questionnaire (henceforth, LALQ) developed and validated by the researcher (Appendix C). Although there are some questionnaires in this regard; however, since there are noticeable differences in the context of Iran with regard to testing and assessment culture, the researcher tries to include socio-cultural differences in the questionnaire. The last three questionnaires used in this study are teacher sense of flexibility, efficacy and teacher perceptions of assessment that have already been designed and validated by prominent researchers in Iran.

Data Collection Procedure

Prior to the treatment phase of the study, the researcher designed the relevant inventory of teacher assessment knowledge. The content of the inventory items are derived from the commonalities emerged from the interviews as well as the findings of the previous studies in this regard. The content validity of the inventory is checked by both local and global experts in testing and assessment. Before, the treatment, the participants are asked to respond to the inventory and their scores are calculated by the researcher. Moreover, to construct and validate the questionnaire of language assessment literacy, the current study adopts a straightforward procedure including three steps to ensure the reliability and validity of the questionnaire as much as possible. First, the researcher designs the question items with reference to the data emerged from the previous qualitative study. As the initial piloting, he asks two specialists in testing and assessment to indicate the problems of the items. Having received the feedback from the initial pilot group, the researcher carries out the final piloting during which the scale is administered to 150 students. The questions are written in English language. He includes one section devoted to demographic information. The typed scale containing questions on a single page in a Likert type scale is used with 'strongly agree' at one end and 'strongly disagree' at the other. The minimum and maximum scores are 1 and 5, respectively. The researcher works both on the psychometric and edumetric language testing paradigm covering such terms as reliability and construct validity and the shift in the definitions of reliability and validity from testing culture to dynamic assessment culture. The problem with the teachers is that although they assign grades for the students during the classroom; however, they finally evaluate students based on the final test and most of them, unfortunately, give mark to students' oral performances out of their performances in written

examinations. That is why students are concerned more about paper and pencil examinations than oral ones.

Data Analysis

Reliability

The internal consistency of LALQ is assessed with Cronbach's Alpha as the most widely used reliability estimate. It is used to assess the internal consistency of the whole items. Additionally, the internal consistencies of the new factors constructing the validated scale are also examined using Cronbach's Alpha.

Construct Validity

Since factor analysis departures from a correlation matrix, the researcher first creates a correlation matrix using SPSS version 20:00 for Windows by calculating the correlations between each pair of variables. The validity of the hypothesized factor structure of the scale is examined through Exploratory Factor Analysis (EFA) to identify the underlying factors by calculating the eigenvalues of the matrix greater than 1.0. Besides these analyses, the researcher decides about the number of factors to retain for rotation. The Scree Test, which is available in the most frequently used statistical software, involves examining the graph of the eigenvalues and looking for the natural bend or break point in the data where the curve flattens out. Since interpretation of the factors can be very difficult, a solution for this difficulty is factor rotation. Factor rotation alters the pattern of the factor loadings, and hence can improve interpretation. The researcher uses Varmix (orthogonal rotation), as the most commonly used orthogonal rotation, since based on convictional wisdom, orthogonal rotation produces more easily interpretable data. Varimax rotation results in a rotated component matrix and a transformation matrix. The rotated component matrix presents the 'post rotation' loadings of the original variables on the extracted factors, and the transformation matrix give information about the angle of rotation. The rotated component matrix illustrates the variables loaded in each factor so that the researcher comes up with the new factors. The new factors or constructs are given new labels by the researcher.

References

1. Assessment Quarterly subscription information. Retrieved June 12, 2014 from: http://www.tandfonline.com/loi/hlaq20
2. Assessment Reform Group. (2002). Assessment for learning: 10 principles.
3. Bandura, A. (1986). Social foundations of thought and action: A cognitive social theory. New York: Pretince Hall, Englewood Cliffs.
4. Bandura, A. (1997). Self-efficacy: The exercise of control. New York: Freeman.
5. Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. Science Education, 80 (2), 145-163.
6. Claxton (Eds.), Learning for life in the 21st century, (pp. 73–83). Oxford: Blackwell.
7. Davidson, D. (2001). Essays on actions and events (2 Ed.). New York: Oxford University Press.
8. Falchikov, N., & Boud, D. (2008). The role of assessment in preparing for lifelong learning: Problems and challenges. In A. Havnes & L. McDowell (Eds.), balancing dilemmas in assessment and learning in contemporary education (pp. 87-99). New York: Routledge.
9. Fulcher, G. (2012). Assessment Literacy for the Language Classroom. Language
10. Gipps, C. (1994). Beyond testing: Towards a theory of educational assessment. London: The Falmer Press.

11. Gipps, C. (2002). Sociocultural perspectives on assessment. In G. Wells & G.
12. Hornberger (Eds.), Encyclopedia of language and education: Language testing and assessment (pp. 225-239). NewYork: Springer Media.
13. Malone, M. (2008). Training in language assessment. In E. Shohamy & N.
14. Marshall, B., & M. Drummond (2006). How teachers engage with Assessment for Learning: Lessons from the classroom. Research Papers in Education 21, (1), 133–49.
15. Marso, R. N., & Pigge, F. L. (1993). Teachers' testing knowledge, skills, and practice. In S. L. Wise (Ed.), Teacher training in measurement and assessment skills. Buros Institute of Mental Measurements, University of Nebraska- Lincoln.
16. Retrieved June 4, 2013, from: http://k1.ioe.ac.uk/tlrp/arg/CIE3.PDF.
17. Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. Educational Measurement: Issues and Practice, 10, 3-6.
18. Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72(7), 534-539.
19. Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. Educational Measurement: Issues and Practice, 18(1), 23-27.
20. Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of Chinese college EFL teacher's experience. TESOL Quarterly, 43(3), 493-513.