SEFAD, 2018 (40): 171-190
e-ISSN: 2458-908X
DOI Number: https://dx.doi.org/10.21497/sefad.515260

The Utilization of the European Standards for Defining Educational Assessment: Teacher-Tester Attributes and Directors' Control*

Assist. Prof. Dr. Nurdan Kavaklı
İzmir Democracy University Faculty of Education
The Department of English Language Teaching
nurdankavakli@gmail.com

Prof. Dr. İsmail Hakkı Mirici
Near East University Atatürk Faculty of Education, Northern Cyprus
Department of Social Science and Turkish Education
hakkimirici@gmail.com

Abstract

This study aims to scrutinize the utilization of the European guidelines in testing and assessment practices of non-formal English language schools. Providing insights from a mixed-methods research design, the quantitative data were gathered from the English language teachers, who were also working as test (-item) developers at three private institutions renowned for quality with the highest course attendee capacity and branches in Turkey to reveal teacher-tester attributes, whereas qualitative data were gathered from the directors of these private institutions to screen directors' control. The results have yielded that (1) the kinds of assessment in use allow for feedback on the performance of the on-going educational system; (2) the overall evaluation of the total program, and assessment of educational systems are taken into consideration in testing procedures to some extent; (3) what is good for the individual in assessment does not thoroughly align with the United Nations Convention on the Rights of the Child; (4) the assessment applied in the selected private institutions does not mainly cover standardized tests. The results are discussed, and laced with suggestions to improve the quality of current testing and assessment practices by the exploitation of the European Framework of Standards for Educational Assessment (AEA- Europe 2012) regarding non-formal private institutions as the arteries of Turkish education economy.

Keywords: Language testing, educational assessment, AEA- Europe, non-formal education, EFL.

Eğitsel Değerlendirmeyi Tanımlamada Avrupa Standartlarının Kullanımı: Öğretmen-Ölçen Yordamı ve Yönetici Kontrolü

Öz

Bu çalışma, yaygın eğitim veren İngilizce kurslarının ölçme ve değerlendirme uygulamalarında Avrupa ölçütleri kullanımını irdelemeyi amaçlamaktadır. Karma yönteme dayalı olan bu çalışmada, nicel veri Türkiye'de en yüksek katılımcı kapasitesine, belli bir

* This research is based on a PhD thesis entitled "Cefr Oriented Testing and Assessment Practices in Non-Formal English Language Schools in Turkey" submitted to Hacettepe University Graduate School of Social Sciences in 2018.

Sending Date / Gönderim Tarihi**:** 25/05/2018
Acceptance Date / Kabul Tarihi**:** 31/10/2018

kaliteye ve sayıca en fazla şubeye sahip İngilizce kurslarında aynı zamanda sınav hazırlayıcı olarak çalışan İngilizce öğretmenlerinden toplanmıştır. Öte yandan, nitel veri ise aynı kurumlardaki yöneticilerden toplanmıştır. Nicel veri ile öğretmen-ölçen yordamına ulaşmak hedeflenmiş, nitel veri ile de yönetici kontrolünün sürece etkisi dikkate alınmıştır. Buna göre, çalışmanın sonuçları göstermiştir ki (1) kullanılan ölçme yöntemleri halihazırdaki eğitim sistemi hakkında geribildirim sağlamaktadır; (2) müfredatın tamamının ve eğitim sisteminin değerlendirilmesi bir noktaya kadar dikkate alınmaktadır; (3) ölçme sürecinde bireyin iyiliği için yapılanlar Birleşmiş Milletler Çocuk Hakları Konvansiyonu ile tam bir uyum içinde değildir; (4) belirlenen İngilizce kurslarında yürütülen ölçme ve değerlendirme faaliyetleri ölçünleştirilmiş sınav uygulamalarından uzaktır. Çalışmanın sonuçları tartışılmış ve güncel ölçme ve değerlendirme uygulamalarının kalitesinin geliştirilmesi için çeşitli tavsiyeler sunulmuştur. Bu noktada, ilgili kurumları Türkiye eğitim ekonomisinin arterleri olarak görerek Avrupa Eğitsel Değerlendirme Birliği'nin sunduğu çerçeveye (AEA- Europe 2012) başvurulabileceği önerilmiştir.

Anahtar Kelimeler: Ölçme ve değerlendirme, eğitsel değerlendirme, ölçünleştirilmiş sınav, yaygın eğitim, İngilizce'nin yabancı dil olarak öğretimi.

1. INTRODUCTION

Recently, there has been an ongoing increase in the demand of English language learning through private institutions as formal education is somehow limited. According to the British Council's report, it is supposed that by the year 2020, the number of the adult English Language Learners (henceforth ELLs) is expected to rise to about 2 billion from 1.5, meaning that 1 out of 4 is to be using the language across the world (Pearson English 2014). For Graddol (2006: 101) nearly a third of world population are expected to learn English simultaneously. This expectation of significant growth is the case for English both within and outside English-speaking countries. For this reason, Turkey as a non-English speaking country and with its EFL context, holds English as a part of school curriculum, and supply courses paid for privately in language learning centers.

However, Turkey's focus on quantity in formal education rather than quality has blossomed as a major factor of deficiency in English although assumed to be more practical. Therefore, what learners can do with the functional skills necessitated by the task seems more important than how well learners perform in the sense that they can effectively and efficiently use what they acquire as language skills (De Jong 2004: 58; Hulstijn 2007: 663). That is why after those years spent on English language (now it is 11 years until undergraduate education from 2nd to 12th grade), and hours of study ranging between 2 and 4 per week, many of them are still unable to have a simple act even in daily life conversations. Adumbrating that learners have had adequate grammatical and lexical knowledge, it is expected to be fair for them to have a good command of language. But this is not the case. Thus, a myriad of learners has decided to take further English language education by means of language schools/courses, study centers and/or other private institutions in Turkey, as well.

Contrary to ordinary, as the ratio of auditing in non-formal educational settings is rather low as to that of formal education, how testing and assessment practices are carried out by non-formal English language schools is somewhat blur. Of particular interest, the notion of progression in the field of testing and assessment by the European guidelines for non-formal educational settings is at the core as they are the centers enclosing a great number of ELLs for many reasons. However, there is a scarcity of empirical studies conducted on the utilization of the European guidelines in language testing and assessment practices of non-formal educational institutions. Therefore, this study aims to probe into the testing and assessment practices of private English language schools in Turkey, which are listed under the heading of non-formal educational institutions. Supposed to do so, how well they trace the guidelines and basic principles purported by the Association of Educational Assessment- Europe (hereafter AEA-Europe) is enlightened starting from the very beginning with the decision-makers at these private institutions. The importance of assessment in education and utilization of European standards in educational assessment with its broadest sense are also highlighted with the help of the European Framework of Standards for Educational Assessment (AEA- Europe 2012: 11) within the perspectives of teacher-testers and directors.

2. NON-FORMAL EDUCATION (NFE)

The United Nations Educational, Scientific and Cultural Organization (UNESCO) has announced a report on a move towards life-long education (UNESCO 1972: 134). This has led to a tripartite categorization of the education systems taking life-long learning as the core element (Colardyn 2002: 69; La Belle 1982: 160). Just because formal education systems are more conservative to adapt the socio-economic changes around them swiftly, there occurs a point of departure which highlights the distinctions among formal education (FE), in-formal education (IFE) and non-formal education (NFE) around the world (Fordham 1993: 2). Similarly, ensuring that formal education itself cannot respond to constant and rapid changes in economy, social life and technology, the Committee on Culture and Education has reported that non-formal education encompasses learning activities outside the formal education system, nestling young people and adult together in order to make them acquire and maintain abilities, skills and dispositions in a life-long learning concept (CoE 1999: 28). As a result, non-formal education has mushroomed as an educational force of the postmodern world, which develops into the worldwide educational industry (Romi-Schmida 2009: 257).

For the Turkish context, it is due to the radical changes in the general political environment of the late 1980s when the politics of education have undergone a gradual withdrawal of the state from education by creating new opportunities for the private sector (Demirer 2015: 307). Accordingly, the basic premises of national education in Turkey have also become inappropriate and inadequate to meet the demands of the current society, since many other countries finding it difficult to pay for the expansion of formal education. As the education has become more individualized, out-of-school education system is enlightened more than before. The results of a study conducted by the Council of Higher Education on higher education pupils have demonstrated that the proportion of the learners who are enrolled in a private institution in order to meet their further learning needs is 71.8% (CoHE 2007: 82). In this context, it is reported by the Association of Private Educational Institutions and Study Centers in Turkey that the number of private institutions has reached up to 1.500 in 2011, which is approximately 600-750 million Turkish lira revenues. That is why a sudden change in the Turkish education system with the total closure of some of these private institutions, namely *dershanes* in Turkish context, or returning them to the Basic High Schools, has caused some problems (Dolgunsöz 2016: 72). However, it is also emphasized that the opening of new private courses by the municipalities and other non-governmental organizations has resulted in an unfair competition by operating in a wrongful way (ÖZ-KUR-DER 2011). Supporting inequalities in education by such wrongful implementations, these private institutions have become more prevalent for those rushing in a competitive environment where success becomes hard to be accomplished, though (Silova-Budiene et al. 2006: 159; Southgate 2009: 165).

To broach into NFE in Turkey, it appears to be embellished with general and vocational technical programs. At this juncture, the institutes providing NFE could be listed as the practical arts schools, advanced technical schools, industrial practical arts schools, technical education centers, public education centers offering craftwork, literacy courses, tech-related courses and language-related courses, and apprenticeship training centers. The testing and assessment practices of these non-formal educational settings are held by MoNE whereas formal educational settings are conducted by the Measuring, Selection and

Placement Center (ÖSYM) in Turkey. MoNE is indirectly involved in the process of testing and assessment practices of the institutions serving for NFE in Turkey. In other words, the language certificate examination of the non-formal educational institutions is administered by MoNE in Turkey.

3. THE AEA- EUROPE: PURPOSE, GUIDING PRINCIPLES AND INSTRUMENT

The AEA- Europe serves as a platform where developments within the scope of educational assessment in Europe are discussed to cherish collaboration between individuals and related organizations. Therefore, it promotes educational assessment practices together with academic, professional and vocational contexts. Engaging individuals, agencies and organizations in a myriad of activities to improve assessment practices and products in Europe, the AEA- Europe strives for developing an understanding for the impact of these practices in any educational environment.

To accomplish above mentioned purposes, the AEA- Europe has developed the 'European Framework of Standards for Educational Assessment' (AEA- Europe 2012). In a word, this framework offers standards to foster transparency for both users and educational authorities by benchmarking the on-going system of standards for the enhancement of further assessment processes. With the intention of providing an instrument for educational authorities, test providers and score users to compare their assessment practices, the European Framework of Standards for Educational Assessment has flourished as an evidence for its audience through guiding principles and an instrument. In this vein, the guiding principles are constituted by the overall evaluation of the total program by the testing procedures conducted, innovative assessment techniques in use, the European perspective adopted, the standards established to disseminate quality in assessment, the support given for variety of cultural and educational contexts, the definition of the test takers' place in the assessment process, some ethical considerations, the cornerstones of the assessment, the use of the assessment results for other educational settings, the rationale behind the assessment, the alignment of the test results to the Common European Framework of Reference for Languages (hereafter CEFR) (CoE 2001), the dissemination of the results for further use, and possible evidences put forward as the standard requirements of the tests administered. On the other hand, the instrument nestles the nature of evidence, tasks and test types in use. As the Framework goes at educational assessment, it, therefore, goes hand in hand with the European standards. It also highlights ethics in order to ensure individual's rights through fairness. It focuses on practicality, validity and impact on stakeholders as the essential quality concerns. Yet, it supports not only learning, but also decision-making and test development processes in order to enhance viewpoints towards educational assessment.

However, studies conducted in relation to the AEA- Europe's Framework seem to be limited to the field of formal education. Therefore, examining a wider range of curricula, namely not school-based and non-formal educational environments, will surely broaden the viewpoints. In this vein, it is suggested that assessment practices should be molded in reply to globalization around the world; therefore, assessment for a digital world is to be revised and re-arranged in accordance with the Framework (Halbherr-Schlienger et al. 2014: 248). Similarly, as educational assessment has some essential quality concerns not only for learning but also for decision-making and test development processes, teacher assessment literacy is supposed to be enhanced consequently. Therefore, DeLuca, LaPointe-McEwan

and Luhanga (2015: 252) have made a review of international standards and measures, in which they have touched upon the guiding principles of the AEA- Europe, as well. As one of the core professional requirements across all educational systems, the standards for assessment literacy adopted in five countries, namely Australia, Canada, New Zealand, the UK and USA have been probed with special interest on the measures developed after 1990. Henceforth, they have drawn a general frame of changes in the assessment practices over time and across different countries, which are all English-speaking ones. Correlatively, Wools (2015: 133) has developed an evaluation system of validity in order to enhance the quality of educational assessment by means of the results of a design-based project. Within, the theoretical principles and designing tenets are correlated with the guiding principles of the AEA- Europe in order to develop a prototype for validity. Additionally, the Annual Conferences of the AEA- Europe are embellished with various studies on the enhancement of educational assessment practices. Amidst the recent ones, Van Nijlen and Janssen (2014) have touched upon national assessments to measure the 21st century skills, with special reference to that of information processing. Besides, Zumbo (2015) has explored the consequences and side effects of an ecological model of testing (Hubley-Zumbo 2011: 221), in which the assessment is considered something *in vivo* rather than *in vitro*. Herein, Jones and Saville (2009: 53) have suggested the Framework as a model for learning, and as an instrument of harmonization in order to create opportunities for language assessment, and, herewith, to improve the quality of language assessment. Not to mention, Jones and Saville (2014) have highlighted the importance of Learning Oriented Assessment (LOA) with a systemic view. LOA is actually grounded upon the socio-cognitive model of language learning propounded by the Framework. It is noted that such an approach has either been "explicitly or implicitly defined in opposition to traditional externally set and assessed large scale formal examinations" (Davison-Leung 2009: 395).

Basically, if well-devised, any assessment procedure will surely enhance learning. However, learning is effected negatively if this procedure is designed haphazardly and/or poorly. Thus, providing feedback is essential for both decision-makers and program reviewers in order to enhance the quality of educational assessment, and to evaluate programs. In doing so, the Framework follows the assessment development cycle, which is basically composed of standard requirements clarified within core elements, methods of implementation and possible evidences.

4. EDUCATIONAL ASSESSMENT

Educational assessment is an integral part of determining learning outcomes. Thus, it provides feedback for different types of audiences: educators, learners, parents, policy makers and public regarding the effectiveness of the educational services rendered (National Research Council 2001: 261). Therefore, it is ensured that assessment is actually designed to enhance learner's performance, albeit not solely to audit it (Wiggins 1998: 21). In doing this, some newfound perspectives on assessment should be taken into account. Comprehensive assessment systems are to be implemented in order to cater learners with a more rigorous and ubiquitous measurement of the learning experiences (Shute-Leighton et al. 2016: 36). Henceforth, the utilization of the AEA- Europe's Framework as a baseline in assessing the assessment system is of utmost importance to improve educational assessment.

Bearing these in mind, the current study was conducted to scrutinize whether the Framework might replenish a fundamental basis for the reconsideration of educational assessment, laying the emphasis on the English language schools serving as non-formal educational settings in Turkey. Accordingly, the perceived gap in the literature is postulated to be filled with the answers to the research questions that come into picture as below:

1. What are the teacher-testers attributes from non-formal private institutions to the utilization of the Framework set by the AEA- Europe?
2. What is the effect of director's control in the utilization of the European standards for defining educational assessment?

5. METHODOLOGY

This study aims to provide insights from a mixed-methods research design-based exploration of the appropriateness of the current testing and assessment practices of English language schools rendering non-formal education in Turkey to the European Framework of Standards for Educational Assessment set by the AEA- Europe. Therefore, both qualitative and quantitative data were collected in order to arrive at an understanding of the on-going testing and assessment practices of three institutionalized private English language schools offering education in their branches in all of the major cities in Turkey.

5.1. Participants and Setting

Three major non-formal private institutions serving as English language schools in Turkey as the source of subjects were selected for this study. For the selection process, the primary concern was to cooperate with the most prominent courses which were renowned for quality in learning English in Turkey with the highest course attendee capacity and highest number of branches in Turkey in order to enable the generalizability of the results. For the selection process, 'convenience sampling' (Dörnyei 2007: 129; Nunan 1992: 142) was also adopted as a technique concerning the fact that participants could be more convenient for accessibility by the researcher.

In the light of these, the data were collected in the fall term of the academic year 2016-2017 with the participation of 40 English language teachers (12 male and 28 female participants) recruited from aforementioned 3 English language schools, whose name were kept anonymous for the confidentiality of the results; therefore, labelled as A, B and C. The English teachers participated in the study were each counted as 11, 19 and 10 from the above labelled English language schools respectively. The participants' age range ranked from 18-25 (N= 27) and 26-35 (N= 12) to 36-45 (N= 1). When their years of experience were considered, teachers mostly had the experience of less than five years (N= 32) which was followed by 5 to 9 years (N= 6) and more than 14 years (N= 2) respectively. One more to note, all of the participants were both English language teachers and test (-item) developers at the private institutions they were working. The table given below summarizes the demographic information about the participants:

Table 1: Overall Demographic Information of the Participants

		N	Percentage %
Institution	A	11	27.5%
	B	19	47.5%
	C	10	25.0%
Gender	Male	12	30.0%
	Female	28	70.0%
Age	18-25	27	67.5%
	26-35	12	30.0%
	36-45	1	2.5%
Years of Experience	less than 5	32	80.0%
	5-9	6	15.0%
	more than 14	2	5.0%
Occupational Field	teacher	40	100.0%
	test (-item) developer	40	100.0%
Total N		40	100.0%

To elaborate, the institution A was composed of 11 English language teachers who were also working as test (-item) developers. Of those, 7 were female (63.6%), and 4 were male (36.4%) with the age range of 18-25 (N= 7; P= 63.6%) and 26-35 (N= 4; P= 36.4%). Additionally, they had the years of teaching experience ranging from less than five years (N= 8; P= 72.7%) and from five to nine years (N= 2; P= 18.2%) to fourteen years and above (N=1; P= 9.1%) respectively. On the other hand, the institution B was composed of 19 English language teachers who were also working as test (-item) developers. Of those, 16 were female (84.2%), and 3 were male (15.8%) with the age range of 18-25 (N= 17; P= 89.5%) and 26-35 (N= 2; P= 10.5%). Besides, they all had less than five years of teaching experience (N= 19; P= 100%). One more to note, the institution C was composed of 10 English language teachers who were also working as test (-item) developers. Of those, 5 were female (50%), and 5 were male (50%) with the age range of 18-25 (N= 3; P= 30%), 26-35 (N= 6; P= 60%) and 36-45 (N=1; P= 10%). Additionally, they had the years of teaching experience ranging from

less than five years (N= 5; P= 50%) and from five to nine years (N= 4; P= 40%) to fourteen years and above (N=1; P= 10%).

5.2. Instruments

The instruments to collect data were remarked as a questionnaire composed of the guiding principles and instrument of the 'European Framework of Standards for Educational Assessment' (AEA- Europe 2012) for establishing quality profiles in educational assessment, and a form of semi-structured interview sessions conducted with the directors of selected private institutions. Accordingly, a questionnaire composed of 24 items on a 5-point Likert-type response basis was administered for this study. The first section of the questionnaire aimed to collect demographic information about the sample group such as gender, age, years of teaching experience and occupational field. The second section of the questionnaire was composed of 24 standards for establishing a quality profile in educational assessment. These standards were aligned with the guidelines and instrument set by the AEA- Europe, and were arranged in the format of a 5-point Likert type scale, in which 'Strongly Disagree' was the lowest possible rating and 'Strongly Agree' was that of highest. The test items were all molded into a table adjacent to the cells next to each test item. During the arrangement process, the wording of the questionnaire was slightly modified. More precisely, instead of 'The tests should require …' pattern, 'The tests in use require …' pattern was employed in the wording of each test item. Herewith, the participants were asked to read each statement carefully and circle the number in the cells (from 1 to 5) which was the best descriptor of their own opinions, ensuring that there was not any correct or false answer, and all of the information that could identify them would remain confidential. The minimum standards were set in liaison with the Framework; however, they were not gathered together, evaluated and exploited by researchers all at once. Therefore, in order to check the internal consistency of the scale in use, a reliability analysis was conducted. As a prior step, negatively worded items were checked and noted as none. Then, overall Cronbach's Alpha level for the instrument was evaluated for the context in which the present study was conducted. The Alpha reliability co-efficient of the data collection instrument was estimated as .894, indicating that the reliability of the data collection instrument was considered to be strong. Because the Pearson Correlation Coefficient is categorized as 1, perfect; .70- .90, strong; .40- .60, moderate; .10- .30, weak; and 0, zero (Dancey-Reidy 2004). Additionally, the internal consistency was also checked by split-half reliability. It was yielded by the split-half reliability analysis that Cronbach's Alpha for the first part was .820 (r^1 for 12 items) whereas that of the second part was calculated as .818 (r^2 for 12 items). As noted, there was a high internal consistency within items and no problematic data entry was identified. An outline of these standards could be seen in table given below:

Table 2: An Outline of the Questionnaire Items

Section	Sub-section(s)	Number of Items
The AEA-EUROPE's Standards for Educational Assessment	1. Guiding Principles	19 (Item No. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19)
	2. Instrument/ Identifying the Nature of Evidence, Tasks and Test Types	5 (Item No. 20, 21, 22, 23, 24)
TOTAL	2 sub-sections	24 items

Besides, the data gathered by the questionnaire from the teachers and test (-item) developers were laced with semi-structured interviews with the directors of the institutions assigned. The semi-structured interview sessions were led by the researcher within the scope of the running of the on-going testing and assessment practices, difficulties and problems encountered in the implementation of the testing and assessment practices together with the recommendations for further improvement in educational assessment.

5.3. Data Analysis

Following data collection process by convenience sampling, the raw data were taken to analysis by aparting quantitative data at one side and that of qualitative one at the other side. For quantitative data, statistical procedures were employed via SPSS Version 23.0 after entering all the valid data in. On the other hand, the data gathered qualitatively, which were noted as the directors' reports both from the institutions were analyzed through constant-comparison analysis method. The selection of each statistical technique primarily depended upon accuracy and precision in essence. The data gathered was taken to analysis with the identification of the demographic information first. Besides, the mean scores were ranked from the highest to the lowest in order to distinguish the most positive and more negative items assessed. Through descriptive statistics, each of the items were summarized enabling comparisons across the institutions selected, enabling researcher to compare the relative weightings of the exploitation of the Framework by the selected private institutions.

On the other hand, the semi-structured interview sessions with the directors of selected private institutions were conducted in the first language of the director, which was Turkish. Therefore, after the sessions, the researcher translated the original version into the target language, which was English. With the help of back-translation method, two independent raters translated this version into the original language with no prior knowledge of the original content, enabling the researcher to consult with the translators to detect any discrepancies (Marín-Marín 1991). In order to prevent the translated instrument skewed one-way and to reduce "human factor as each inquirer had his/her own unique final destination just like a scientific two-edged sword" (Patton 2015: 433), those independent raters were selected concerning the fact that they had different background of knowledge, expertise and world view, albeit proficient in the target language.

As a procedure, the analysis of the semi-structured interview reports of the directors followed a constant-comparison analysis method (Bogdan-Biklen 2003: 66). The constant comparison analysis method pursues a very similar way to the grounded theory approach, in which researchers come up with an emergent fit; therefore, they adjust the category to fit the data, albeit do not go for data to link with a pre-determined category (Taber 2000: 473). In this sense, the constant-comparison analysis method encompasses a process of reducing the data gathered by means of constant recoding (Glaser-Strauss 1967: 102). Thus, the procedure is broken down into steps starting with the comparison between the already existing incidents, which is further pursued by the comparisons between concepts and incidents. Elliott and Jordan (2010: 34-35) states that "… it is through the process of comparing concept to incident that the researcher can check to see if further incidents fit with the newly developed concepts and, in so doing, ensure that the concepts are capable of accounting for all related incidents in the data".

Based on this, the researcher designated codes to each line directly in the margins of the interview reports, associating entries with codes with similar meanings into a new category. This process continued for each of the remaining reports of the directors. Following a reiterative angle, codes from the first report were transferred to the second one, and those of the second report were carried over to the third one. This procedure made it possible to create thematic trends across the institutions, and the self-reports of their directors through reunification.

6. FINDINGS AND RESULTS

6.1. The Attributes of Teacher-Testers from Non-Formal Private Institutions to the Utilization of The Framework Set by the AEA- Europe

In order to define the teacher-tester attributes, two main considerations of the AEA-Europe were taken into consideration. The first main consideration of the AEA- Europe, namely guiding principles, was composed of 19 core items. Guiding principles were constituted by the overall evaluation of the total program by the testing procedures conducted, innovative assessment techniques in use, the European perspective adopted, the standards established to disseminate quality in assessment, the support given for variety of cultural and educational contexts, the definition of the test takers' place in the assessment process, some ethical considerations, the cornerstones of the assessment, the use of the assessment results for other educational settings, the rationale behind the assessment, the alignment of the test results to the CEFR, the dissemination of the results for further use, and possible evidences put forward as the standard requirements of the tests administered. Secondarily, the instrument nestling the nature of evidence, tasks and test types were checked by means of 5 items. Whence, it was asked what kind of tests were applied in practice by the previously selected private institutions, such as summative assessment, formative assessment, performance assessment, standardized tests and/or competency tests.

In the light of these, the highest mean score of the guiding principles was detected with the item asserting that assessment types were laced with feedback on the on-going educational system's overall performance (M= 4.27; SD= .65). Following that, the participants of this study stated that the test results could be appropriately used as one of the essentials of the quality as they were meaningful (M= 4.09; SD= .54). Likewise, it was stipulated by the results of this study that the testing procedures were adorned with the overall evaluation of

the total program together with the assessment of the on-going educational system (M= 4.09; SD= .54). It was followed by the item claiming that the test results could be valid for various types of educational contexts for further use (M= 4.00; SD= .63). Alike, the tests in use were supposed to indorse the dissemination of the core principles of the actual testing and assessment practices (M= 3.91; SD= .54). At the very same, it was asserted by the participants of this study that decision makers had the opportunity to reckon with the programs through the exploitation of the test results (M= 3.91; SD= .54). Relatively, the assessment process was stipulated to have a basis on a rationale for the proposed learning of the predetermined educational process (M= 3.91; SD= .54). Moreover, the tests in use were supposed to cover various cultural and educational contexts (M= 3.91; SD= .70). In the same vein, the tests in use were assumed to be embellished with the elements of test development cycle of the CEFR (M= 3.91; SD= .83). Correlatively, the assessment procedures were alleged to consider some ethical concerns (M= 3.91; SD= .83), paying regard not only to the rights of the test administrators, albeit to those of the test takers, as well (M= 3.91; SD= .83).

With regard to test design followed by the guiding principles, it was stipulated that innovative assessment techniques were considered in designing tests (M= 3.82; SD= .60). Correlatively, the assessment process was stipulated to cover the test takers' place within (M= 3.82; SD= .75). Besides, it was asserted that the tests in use were adorned with a European perspective to the assessment practices in a widespread interest (M= 3.82; SD= .98). Based on this, the purpose of the assessment was supposed to promote the overall education of the test takers (M= 3.73; SD= .79). To some extent, the anchors of the assessment process were assumed to be addressed delicately (M= 3.64; SD= 1.03). Furthermore, the assessment procedures were estimated to follow the aims set by the CEFR to some degree (M= 3.64; SD= 1.03). Accordingly, the tests in use were stipulated to cover the essentials of the assessment process by means of some possible evidences (M= 3.64; SD= .51). One more to note on the guiding principles of the AEA- Europe, it was stated that the rights of the test takers complied with the regulations of the United Nations Convention on the Rights of the Child at the lowest mean score of all (M= 3.45; SD= .69).

The overall estimations regarding the exploitation of the Framework by selected private institutions was reported by means, standard deviations and standard errors of mean for each of them elaborately. In this context, the replies of the English language teachers to the questionnaire were noted at one hand, and the results of each private institution were reported separately. Accordingly, it was yielded by the findings of this study that for the utilization of the Framework, the highest mean score was estimated by the private institution B (M= 4.07; SD= .07), which was followed by that of institution A (M= 3.84; SD= .10) and that of institution C (M= 3.53; SD= .11) respectively.

6.2. The Effect of Director's Control in the Utilization of the European Standards for Defining Educational Assessment

The general paradigm of a sample of leading professionals from a range of non-formal English language schools in Turkey on the implementation of testing and assessment procedures as defined by the European guidelines was drawn taking the views of the decision-makers from the selected private institutions. The viewpoints of the directors from the selected private institutions on the utilization of the Framework in testing and assessment practices were highlighted by means of semi-structured interview sessions conducted face-to-face. The answers were noted pursuant to the directors' standpoints on

the current implementations in testing and assessment, and analyzed through constant-comparison analysis method. As a result, it was underscored that the development of a more practical curriculum, the need for more qualified language teachers, the need for a validation process for language certificate examinations, and thereby the need for a standardization process in language teaching and assessment mushroomed as the standpoints.

Accordingly, it was reported by the director of A that the private institution(s) appeared as a trading house which was merchandizing education. Therefore, the student(s) enrolled in such kind of private institution(s) were well aware of the fact that it was the identity of the institution(s) which was protected, albeit not that of student(s). To set an example, the director of A stated that if there was a vacancy in A2-level proficiency class, a student who was marked as proficient at B1 level via placement test was also sent to that class due to the fact that B1-level proficiency class was full. Moreover, the tests were conducted in multiple-choice-item format within the scope of vocabulary, grammar, listening and reading. Besides, each English language teacher prepared his/her own speaking and writing examinations, and conducted these examinations at his/her convenience. Thereafter, a mean value was calculated to get a final score for the placement test. As there were no standards in testing and assessment of speaking and writing, the director of A reported that some a priori problems might mushroom as a result of misapplications.

In other respects, for the enhancement of the on-going testing and assessment practices within the institution A, its director recommended that performance assessment was to be placed more importance than paper-and-pencil tests. Postulated as the fundamentals of language teaching by the director of A, the productive skills were suggested to be given more prominence by even creating and adopting a new form of placement test based on an oral proficiency examination, as well. For the improvement of the on-going testing and assessment practices across the country, the director of A stated that a skills-based approach was to be employed by all education centers; henceforth, the students enrolled in any of those centers could internalize the English language better.

On the other hand, the director of the private institution B asserted that the students enrolled in the institution B were taken to a diagnostic test in order to determine the level of language proficiency at the outset. Particularly, this diagnostic test was done on students' speaking skill, and the results gathered made it possible to know where the students were academically so as to bring them to where they were actually in need to be. With respect to the recommendations for the enhancement of on-going testing and assessment practices within the institution B, its director stated that the students were to be given freedom so that they could quiet their minds, and feel free to speak when they did feel truly ready. For the improvement of the on-going testing and assessment practices across the country, the director of the private institution B recommended that Turkish system of English language teaching led by the MoNE was to be revised and modernized so as not to be out-of-date. To set an example for this, the director of B addressed that English language teaching could be a part of early childhood education and/or pre-school education, and be a prerequisite for further education. In the same context, it was marked out by the director of B that the ELT curriculum was to be reviewed as the newly graduates of the ELT departments in Turkey had some problems in conducting skills-based testing and assessment procedures. To add more, the director of B suggested that there was to be a standardization in testing and

assessment practices across the country. Because someone with a proficiency level of B1 might be regarded as proficient at the level of A2 by another institution.

One more to note, it was reported by the director of C that the most difficult part was the teachers' internalization of the new applications as it was marked as rather hard to persuade the teachers on the use of them. Correlatively, for the enhancement of on-going testing and assessment practices within the institution C and across the country, its director recommended that language testing and assessment was to be linked to a more standardized system. In addition, its director suggested that skills-based teaching was to be highlighted more, and put into use.

7. DISCUSSION

An in-depth analysis of the results was applied in a two-way alternate: (1) data collected from a 5-point-Likert type scale, and demographic information gathered from the English language teachers (40 in total, 28 female and 12 male, working at the selected private institutions also as the test-item developers); (2) data gathered from the semi-structured interview sessions conducted with the directors of the selected private institutions (3 directors in sum).

The data regarding the utilization of the European Framework of Standards for Educational Assessment by selected private institutions have yielded that even the most prominent English language schools which are renowned for quality in learning English in Turkey with the highest course attendee capacity and highest number of branches across the country have not embraced these guidelines in language testing and assessment thoroughly. Even more, they assert that they have adopted the Framework as the fundamental basis for defining education assessment and conducting language testing and assessment procedures, albeit inefficiently. Besides, it is reported by the findings of this study that the English language teachers, who are also test (-item) developers at those private institutions, are well aware of the importance of the Framework, yet have not implemented it effectively.

To note more, it is reported that the guiding principles are applied more (M= 3.86) than the instrument (M= 3.76) although overall estimates regarding the adoption of the AEA-Europe's Framework is cumulatively low (M= 3.81). Therefore, it is to be noted that the lowest mean score of the guiding principles is estimated for the item numbered 7 (M= 3.45), which is about the goodness of the test takers as the individuals who are taking the tests if aligned with the United Nations Convention on the Rights of the Child (UN 1990: 2). Accordingly, individual's place in the assessment procedure is expected to be guaranteed by the declaration of the United Nations, confirming that everyone is entitled to all rights asserted without any distinction of any kind, such as race, ethnicity, language, gender, or any other status. However, it is stipulated by the findings of this study that a big majority of the English language teacher-testers as the participants of the study are not well-aware of what it is actually about as they have noted themselves as mostly 'not sure' in reply to the aforementioned test item (P= 42.5%).

On the other hand, it is noted by the findings regarding the instrument by the AEA-Europe that the lowest mean score is estimated on the use of standardized tests within selected private institutions (M= 3.45). In effect, a Reference Supplement to the Manual for Relating Examinations to the CEFR has been introduced (Banerjee 2004; Eckes 2009; Kaftandijeva 2004; Verhelst 2004a-b-c-d) to enable standardization in developing tests, and

aligning them to the Framework. To note more, it is reported by the findings of this study that summative assessment is the type of assessment which is most generally applied in the selected private institutions (M= 4.00). It is followed by the implementations of formative assessment (M= 3.82), and those of performance assessment (M= 3.73). At this juncture, Spinelli (2007: 103) has suggested informal assessment as an authentic solution to the need for formative assessment in order to involve individual's learning styles and personal challenges into the process; thus, teachers can track the on-going educational process more regularly, and often by taking students' snapshots throughout the process. Correlatively, the data gathered from the directors of those selected private institutions have yielded that similar types of assessment formats are in use, which are mostly summative. Test takers are provided with contemporary self-assessment tools to some extent, such as the European Language Portfolio (hereafter ELP) (CoE 2001: 5). For the private institution B, the ELP is the classroom-based assessment tool. However, for the private institution A and C, there are some restrictions in use, such as age and language proficiency level. However, the ELP is the fundamental tool for learners to keep record their own learning by themselves (CoE 2011: 6; Little 2005: 331; Mirici 2008: 28; Mirici-Kavaklı 2017: 75; Sarıçoban 2011: 400; Schäerer 2005: 5); therefore, the recognition and implementation of the ELP is a necessity of the time, albeit not a choice.

However, some problematic issues blossom as there are no standards in language testing and assessment practices of the selected private institutions. Therefore, someone with a proficiency level of B1 might be regarded as proficient at the level of A2 by another institution. According to the views of the directors from selected private institutions, it is reported that current testing and assessment practices are to be linked to a more standardized system. So, the problem is setting standards for quality. However, it is to be noted that setting standards is not the same with adopting standardization due to the fact that standardization refers to settings things in completely the same way (Sleeter-Carmona 2017: 43). Even so, such kind of standardized tests should at least be laced with some alternative assessment measures (Menken 2008). In the same vein, it is recommended that formative assessment should enhance learning by providing feedback for both teachers and learners together with the opportunity for self-evaluation (Walvoord-Anderson 2010: 61).

Besides, standard requirements, methods and samples of evidence as the sub-components of the instrument set by the AEA- Europe were stipulated not to be sufficiently addressed by means of observations and verifications. This situation reveals that the design of the assessment procedure does not properly represent the content which is covered by knowledge, skills and other attributes, and the setting in which the assessment is going to take place. For the evaluation and next iteration phase, the results are expected to embrace their further use for other educational cases; however, it was yielded by the findings of this study that the concept of next iteration was not fully understood either to develop a new form of assessment, or to improve the already existing one within the scope of the European standards touched upon above. Even in higher education, (vice-)directors of foreign language schools in Turkish universities are from the field of foreign language teaching, albeit not language assessment (Zengin-Hacıfazlıoğlu 2013). Therefore, a more robust auditing system is needed in order to enhance the quality of language testing and assessment practices in non-formal educational settings (Kavaklı 2018).

8. CONCLUSION

To sum up, in this study, the current language testing and assessment practices in non-formal educational settings, as the arteries of Turkish education economy, have been explored and discussed to improve the quality of educational assessment through the exploitation of the guidelines set by the AEA- Europe. Basically, it is concluded that adopting themselves as the main beneficiaries, the English language teacher-testers have not given due weight in order to guarantee test takers' rights since the assessment process is inscribed more to the test administrators and developers more than test takers. With special concern to the AEA- Europe's Framework, it was concluded that the guiding principles were applied merely to some extent by the English language teachers, who were also working as test (-item) developers at the selected private institutions. At this point, the English language teacher-testers admitted that they suffered from using traditional assessment techniques more than innovative ones although the Framework, itself, did focus on educational assessment supporting learning. This might indicate that new forms of assessment which fit for a European environment are not adequately placed emphasis. It also seems that disseminating quality in educational assessment for the development of quality in educational assessment with a European perspective has blossomed as a need for all of the private institutions rendering English language education in a non-formal way. In this context, the English language teachers might have experienced role models or mentors in order to grasp the gist of the Framework.

To conclude, contrary to the ordinary, the testing and assessment practices of non-formal private institutions are taken as the core instructional context for this study within teacher-testers' and directors' perspectives. Therefore, the results of this study are expected to lend assistance to different types of audiences: English language teachers, test (-item) developers, the directors of the private institutions, public enterprises and the directors of other non-governmental organizations.

9. DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

BIBLIOGRAPHY

Association for Educational Assessment in Europe (AEA- Europe). (2012). *European Framework of Standards for Educational Assessment* (Version 1.0). Rome: Edizioni Nuova Cultura.

Banerjee, Jay (2004). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the CEF: Section D: Qualitative analysis methods*. Strasbourg: Language Policy Division.

Bogdan, Robert C.- Biklen, Sari Knopp (2003). *Qualitative research of education: An introductive to theories and methods* (4th edition). Boston: Allyn and Bacon.

Colardyn, Danielle (ed.) (2002). *Lifelong learning: Which ways forward?* Utrecht: Lemma.

Council of Europe (CoE). (1999). "A report on non-formal education". The Parliamentary Assembly of the Committee on Culture and Education. assembly.coe.int/nw/xml/XRef/X2H-Xref-ViewHTML.asp?FileID=8807&lang=en. [20.06.2017.]

Council of Europe (CoE). (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe (CoE). (2011). *Manual for language test development and examining: For use with the CEFR*. Strasbourg: Language Policy Division.

Council of Higher Education (CoHE). (2007). "Türkiye'nin yu¨kseköğretim stratejisi" [Higher education strategy of Turkey]. Ankara: Council of Higher Education. www.yok.gov.tr/documents/10279/30217/yok_strateji_kitabi/27077070-cb13-4870-aba1-6742db37696b [25.06.2017.]

Dancey, Christine P. - Reidy, John. (2004). *Statistics without Maths for psychology: Using SPSS for windows*. London, UK: Prentice Hall.

Davison, Chris - Leung, Constant (2009). "Current issues in English language teacher-based assessment". *TESOL Quarterly, 43*(3), 393-415.

De Jong, John Hal (2004). "Comparing the psycholinguistic and the communicative paradigm of language proficiency". *International Workshop Psycholinguistic and Psychometric Aspects of Language Assessment in the Common European Framework of Reference for Languages*. University of Amsterdam, The Netherlands.

Deluca, Christopher, Lapointe-Mcewan, Danielle et al. (2015). "Teacher assessment literacy: A review of international standards and measures". *Educational Assessment, Evaluation and Accountability, 28*(3), 251-272.

Demirer, Derya Keskin (2015). "Reproduction of inequality through private out-of-school education". *Education Applications and Development: Advances in Education and Educational Trends*. ed. Mafalda Carmo. World Institute for Advanced Research and Science (WIARS), Lisbon: The Science Press. 259-269.

Dolgunsöz, Emrah (2016). "A sudden change in Turkish education system: Public attitude towards dershane debates in Turkey". *E-International Journal of Educational Research (E-IJER), 7*(2), 56-75.

Dörnyei, Zoltán (2007). Research methods in applied linguistics. Oxford: Oxford University Press.

Eckes, Thomas (2009). *Reference Supplement to the preliminary pilot version of the Manual for Relating Language examinations to the CEF: Section H: Many-Facet Rasch Measurement*. Strasbourg: Language Policy Division.

Elliott, Naomi- Jordan, Joanne (2010). "Practical strategies to avoid the pitfalls in grounded theory research". *Nurse Researcher, 17*(4), 29-40.

Fordham, Paul E. (1993). *Informal, non-formal and formal education programmes in YMCA George Williams College ICE301 Lifelong Learning Unit 2*. London, UK: YMCA George Williams College.

Glaser, Barney G.- Strauss, Anselm L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine De Gruyter.

Graddol, David (2006). *English Next: Why global English may mean the end of "English as a foreign language*". The United Kingdom: The British Council.

Halbherr, Tobias-Schlienger, Claudia, et al. (2014). "Assessments for a digital world". *The Annual AEA- Europe Tallinn Conference: Assessment of students in a 21st century world*. Tallinn, Estonia.

Hubley, Anita M.- Zumbo, Bruno D. (2011). "Validity and the consequences of test interpretation and use". *Social Indicators Research, 103*(2), 219-230.

Hulstijn, Jan H. (2007). "The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency". *The Modern Language Journal (MLJ), 91*(4), 663-667.

Jones, Neil- Saville, Nick (2014). *Learning oriented assessment: A systemic approach (Studies in Language Testing)*. Cambridge: Cambridge University Press.

Jones, Neil-Saville, Nick (2009). "European language policy: Assessment, learning, and the CEFR". *Annual Review of Applied Linguistics, 29*, 51-63.

Kaftandjieva, Felianka (2004). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the CEF: Section B: Standard setting.* Strasbourg: Language Policy Division.

Kavaklı, Nurdan (2018). *CEFR oriented testing and assessment practices in non-formal English language schools in Turkey*. Unpublished PhD Thesis. Ankara: Hacettepe University.

La Belle, Thomas J. (1982). "Formal, non-formal and informal education: A holistic perspective on lifelong learning". *International Review of Education, 28*(2), 159-175.

Little, David (2005). "The common European framework and the European language portfolio: Involving learners and their judgements in the assessment process". *Language Testing, 22*(3), 321-336.

Marín, Gerardo-Marín, Barbara VanOss (1991). *Research with Hispanic populations*. Newbury Park, CA: Sage.

Menken, Kate (2008). *English learners left behind: Standardized testing as language policy.* Clevedon: Multilingual Matters.

Mirici, İsmail Hakkı (2008). "Development and validation process of a European language portfolio model for young learners". *Turkish Online Journal of Distance Education (TOJDE), 9*(2), 26-34.

Mirici, İsmail Hakkı-Kavaklı, Nurdan (2017). "Teaching the CEFR-oriented practices effectively in the MA program of an ELT department in Turkey". *International Online Journal of Education and Teaching (IOJET), 4*(1), 74-85.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

Nunan, David (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.

Özel Öğretim Kurslar, Dershaneler ve Etüt Eğitim Merkezleri Birliği Derneği. (ÖZ-KUR-DER). (2011). "Kamuoyuna açıklama". [Declaration to the Public]. The Association of

Private Educational Establishments and Study Centers of Turkey. www.ozkurder.com/bilgilendirme/kamuya_bilgi.htm. [02.05.2017.]

Patton, Michael Quinn (2015). *Qualitative research and evaluation methods: Integrating theory and practice*. Thousand Oaks, CA: Sage. 4th edition.

Pearson English. (2014). "English: The world's language (infographic)". Pearson. www.english.com/english_learning_infographic [17.09.2016.]

Romi, Shlomo-Schmida, Mirjam (2009). "Non-formal education: A major educational force in the postmodern era". *Cambridge Journal of Education, 39*(2), 257-273.

Sarıçoban, Arif (2011). "A Study on the English language teachers' preparation of tests". *Hacettepe University Journal of Education, 41*, 398-410.

Schäerer, Rolf (2005). *European language portfolio: Interim report 2005 with executive summary*. Strasbourg: Language Policy Division.

Shute, Valerie J.-Leighton, Jacqueline P. et al. (2016). "Advances in the science of assessment". *Educational Assessment, 21*(1), 34-59.

Silova, Iveta-Budiene, Virginija, et al. (eds.). (2006). *Education in a hidden marketplace: Monitoring of private tutoring*. New York, NY: Open Society Institute.

Sleeter, Christin E.-Carmona, Judith Flores (2017). *Un-standardizing curriculum: Multicultural teaching in the standards-based classroom.* New York, NY: Teachers College Press. 2nd edition.

Southgate, Darby (2009). *Determinants of shadow education: A cross-national analysis*. Unpublished PhD Thesis. The USA: The Ohio State University.

Spinelli, Cathleen G. (2007). "Addressing the issue of cultural and linguistic diversity and assessment: Informal evaluation measures for English language learners". *Reading and Writing Quarterly, 24*(1), 101-118.

Taber, Keith S. (2000). Case studies and generalizability: "Grounded theory and research in science education". *International Journal of Science Education, 22*, 469-487.

United Nations (UN). (1990). "Convention on the Rights of the Child". Human Rights Office of the High Commissioner, the United Nations. www.ohchr.org/EN/ProfessionalInterest/Pages/CRC.aspx [20.11.2016.]

United Nations Educational, Scientific And Cultural Organization (UNESCO). (1972). *Learning to be: The world of education today and tomorrow*. Paris: UNESCO.

Van Nijlen, Daniël- Janssen, Rianne (2014). "Measuring 21st century skills through national assessments: The case of information processing skills". *AEA- Europe Tallinn Conference: Assessment of students in a 21st century world*. Tallinn, Estonia.

Verhelst, Norman (2004a). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the CEF: section C: Classical test theory.* Strasbourg: Language Policy Division.

Verhelst, Norman (2004b). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the CEF: section E: Generalizability theory.* Strasbourg: Language Policy Division.

Verhelst, Norman (2004c). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the CEF: section F: Factor analysis.* Strasbourg: Language Policy Division.

Verhelst, Norman (2004d). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the CEF: Section G: Item response theory.* Strasbourg: Language Policy Division.

Walvoord, Barbara E.-Anderson, Virginia Johnson (2010). *Effective grading: A tool for learning and assessment in college*. San Francisco: Jossey-Bass Inc. 2nd edition.

Wiggins, Grant (1998). *Educative assessment: Designing assessment to inform and improve student performance*. San Francisco: Jossey-Bass Inc.

Wools, Saskia (2015). *All about validity: An evaluation system for the quality of educational assessment*. Enschede: University of Twente.

Zengin, Buğra-Hacıfazlıoğlu, Özge (2013). "Profile of preparatory school administrators at universities". *Cypriot Journal of Educational Sciences, 8*(3), 351-360.

Zumbo, Bruno D. (2015). "Consequences, side effects and the ecology of testing: Keys to considering assessment in 'In Vivo'". *The annual meeting of the Association for Educational Assessment - Europe (AEA-Europe).* Glasgow, Scotland.