Dissemination and Exploitation of the European Benchmarks for Enabling Quality Assurance in Foreign Language Testing and Assessment[1]

Nurdan KAVAKLI[2], Sezen ARSLAN[3]

ABSTRACT

The descriptors labelled by the Common European Framework of Reference for Languages: Learning, Teaching and Assessment was announced to be used across Europe together with the European Language Portfolio as a self-assessment tool (CoE, 2000). As one of the members of the Council of Europe, Turkey has been using the Common European Framework of Reference descriptors along with the European Language Portfolio to provide a basis for learning, teaching, testing and assessment. The current study aims to frame some European guidelines in order to maintain good practice in language testing and assessment practices. Therefore, the European guidelines set by the European Association for Language Testing and Assessment, the Association of Language Testers in Europe, International Language Testing Association, and the Association for Educational Assessment- Europe, all of which comply with the CEFR (CoE, 2001) are elaborated in detail. In this vein, the benchmarks for goodness in language testing and assessment by the European Association for Language Testing and Assessment (EALTA, 2006), the code of practice by the Association of Language Testers in Europe (ALTE, 1994), the guidelines for practice by the International Language Testing Association (ILTA, 2007), and the European framework of standards for educational assessment by the AEA- Europe (AEA- Europe, 2012) are taken into thematic analysis to define the core tenets of how quality profiles for language tests are labelled. Accordingly, it is yielded that the European guidelines embrace the core tenets of test construction, test analysis, administration and logistics, test production, communication with stakeholders, marking and grading, and item writing. Besides, connectives as the integrative elements of each tenet are also probed into. Correlatively, some practical recommendations to pursue quality assurance in language testing and assessment practices through the dissemination and exploitation of these European standards are listed, as well.

Keywords: Language Testing, Assessment, Dissemination, EALTA, ALTE.

EXTENDED ABSTRACT

Turkey, one of the members of Council of Europe (CoE), uses the descriptors in Common European Framework of Reference for Languages (CEFR) and European Language Portfolio (ELP) in order to provide a well-established basis for the learning, teaching, testing, and assessment. Thus, the procedures of designing and implementing language tests are facilitated; accordingly, a sound basis is provided for reporting and evaluating the learning outcomes. The alignment of the tests has been set and built up in accordance with the Framework. Apart from the Framework, the European Association for Language Testing and Assessment (EALTA) provides guidelines for understanding the theoretical principles for language testing and assessment (EALTA, 2006). Pre-service teachers, in-service teachers and test developers are the main audiences that the association addresses. In doing so, a collaboration could be established among all of the parties, considering the test takers, validity, reliability and fairness. In addition, Association of Language Testers in Europe (ALTE) enables common standards for language testing and assessment through maintaining linguistic and cultural diversity. By incorporating standards and diversity, it enables ALTE Q-mark which shows how well the standards are accomplished. In addition to Q-mark, Code of Practice is also

employed to describe language testing and assessment standards (ALTE, 1994). ALTE includes ELP and European Association for Quality Language Services (EAQUALS)- ALTE portfolio. ELP is useful for documenting the formal and informal learning practices, thereby promoting life-long learning. Likewise, EAQUALS-ALTE encourages learners to take control of their own learning; that is to say, independent learning could be fostered. As for International Language Testing Association (ILTA), it aims to describe the characteristics of language testers in order to improve testing implementations. Therefore, two main sources 1) Code of Ethics, and 2) Guidelines for Practice are employed. Code of Ethics is mainly based on ethical principles and if any failure happens in accomplishing those principles, ILTA Ethics Committee could decide on the withdrawal of membership. Guidelines for Practice include two sections regarding the liabilities of test developers and users, and those of test takers (ILTA, 2007). That is to say, a deeper understanding of the test specifications, test objectives, and procedures regarding validity and reliability should be gained by the test users and developers.

Furthermore, The AEA- Europe offers a platform to discuss the novelties in educational assessment in Europe with a solidarity between individuals and organizations. In doing this, individuals' professional development is promoted by means of conferences and joint projects held across Europe, and cooperation between related organizations is cherished. In order to improve assessment products and practices in the field of education and develop a new understanding into the probable effects of these products, the AEA- Europe applies for the 'European Framework of Standards for Educational Assessment' (AEA- Europe, 2012).

To conclude, European standards provide a basis for language testing and assessment through guiding principles, specifications and resources. In the current study, the pillars of establishing good practice in foreign language testing are highlighted. Accordingly, general principles to be applied can be listed as collaboration amidst the parties involved, fairness, reliability, respect for the candidates, responsibility, and validity. Those who are involved in the language testing and assessment process are expected to clarify themselves: Are they curriculum developers, practicing teachers, and/or trainees? Besides, the purpose of the assessment, its relation to the curriculum, and test specifications are to be known and discussed to detect how well the curriculum is covered. Besides, test review and revision of the test items are to be followed in order to initiate change(s) in the current practice. To add, 21st century skills including information-processing are the components that language testing and assessment are expected to include (Van Nijlen & Janssen, 2014). That is why, the CEFR is recommended as a guide to create opportunities for language testing and assessment, and, thereto, to enhance the quality of related practices (Finch; 2009; Jones, 2007; Kavaklı & Arslan, 2017; Saville, 2005). Also, language assessment courses to be administered meticulously in order to develop a sound basis for language assessment literacy (Inbar-Lourie, 2008; Mirici & Kavaklı, 2017).

Yabancı Dilde Ölçme ve Değerlendirmede Kalite Güvencesinin Sağlanması için Avrupa Ölçütlerinin Kullanımı ve Yaygınlaştırılması[1]

Nurdan KAVAKLI[2], Sezen ARSLAN[3]

ÖZET

Avrupa Dilleri Öğretimi Ortak Çerçeve Programı: Öğrenim, Öğretim ve Değerlendirme ile belirlenen göstergelerin ve beraberinde öz değerlendirme aracı olarak kabul edilen Avrupa Dil Portfolyosu'nun Avrupa'da ve ötesinde kullanılması üzerine ortak karar alınmıştır (CoE, 2000). Avrupa Konseyi'ne üye devletlerden biri olan Türkiye, çerçeve programın göstergelerini (CoE, 2001) ve Avrupa Dil Portfolyosu'nu öğrenme, öğretme, ölçme ve değerlendirme uygulamalarının temeline yerleştirmiş ve kullanmaya başlamıştır. Ek olarak, çerçeve programın öğrenim, öğretim ve değerlendirme alanlarında kullanımı üzerine kalite standartlarının arttırılması amacıyla çeşitli Avrupa ölçütlerini önerdiği de aşikardır. Bu noktada, bu çalışma ile yabancı dilde ölçme ve değerlendirmede iyi uygulamaların sürekliliğini sağlamak için bazı Avrupa ölçütlerinin kullanımını göstermek amaçlanmıştır. Bu nedenle, çerçeve program ile uyum gösteren Avrupa Dil Becerileri Ölçme ve Değerlendirme Derneği (EALTA, 2006), Avrupa Dil Sınav Sunucuları Birliği (ALTE, 1994), Uluslararası Dil Ölçme Birliği (ILTA, 2007) ve Eğitsel Değerlendirme Derneği- Avrupa (AEA- Europe, 2012)'nın belirlediği ölçütler detaylıca irdelenmiştir. Bu bağlamda, Avrupa Dil Becerileri Ölçme ve Değerlendirme Derneği'nin yabancı dilde ölçme ve değerlendirme ana esasları, Avrupa Dil Sınav Sunucuları Birliği'nin uygulama ilkeleri, Uluslararası Dil Ölçme Birliği'nin uygulama kılavuzu ve Eğitsel Değerlendirme Derneği- Avrupa'nın çerçeve ölçütleri tematik analize alınarak, yabancı dil testlerinde kalitenin sağlanması için gerekli olan ana ilkeler belirlenmiştir. Buna göre, Avrupa ölçütlerine göre belirlenen bu ana ilkelerin sınav yapısı, sınav üretimi, idare ve lojistik, puanlama ve derecelendirme, sınav analizi, ilgili taraflar arası iletişim ve madde yazımı etrafında döndüğü saptanmıştır. Ayrıca, her bir ilkeye ait birleştirici alt unsurlar belirlenmiş ve detaylıca açıklanmıştır. İlgili olarak, yabancı dilde ölçme ve değerlendirme çalışmalarında kalitenin ve başarılı uygulamaların sürdürülebilirliğinin sağlanmasında ilgili Avrupa ölçütlerinin kullanımı ve yaygınlaşması için bazı tatbiki tavsiyelerde de bulunulmuştur. Tüm bunların temelinde, bu çalışmada, Türkiye'de yabancı dil olarak İngilizce öğretiminde benimsenen ölçme ve değerlendirme uygulamalarının belirlenen Avrupa ölçütleri kapsamında gelişimi tartışılmış, bundan sonraki çalışmalara da bir ışık tutması öngörülmüştür.

Anahtar Kelimeler: Yabancı Dil, Ölçme ve Değerlendirme, Avrupa Dilleri Öğretimi Ortak Çerçeve Programı, Kalite Güvencesi.

1. Introduction

The development of the Framework dates back to the 1970s albeit officially launched in 2001 by the Council of Europe (CoE). Within a historicist perspective, the wreck after World War II and Cold War bequeathed European nations with a ruin of economy, society and international relations. Herein, post-war effects led European nations to unite against the (un)expected challenges of globalization by means of a robust entity in case such a loss might reiterate (Valax, 2011). Correlatively, a myriad of organizations was established (i.e. Council of Europe, European Cultural Convention, etc.) in order to create a much stronger entity under an umbrella term, paving the way towards the constitution of the Framework. It is also ensured by the CoE (2001) itself to accomplish a unity amidst its members "by the adoption of common action in the cultural field".

Trim (2005) defines this post-war situation as that "under such conditions, language teachers became quite out of touch with the up-to-date realities of the languages and cultures they were teaching and concentrated their attention on puristic formal correctness and the heritage of national literature". Therefore, the CoE introduced the Common European Framework of Reference for Languages (CEFR) in order to determine and pursue goodness in modern language teaching, learning and assessment together with the appraisal of cultural development. The Framework has purported non-language-specific linguistic descriptors together with some (sub) competences gathered through the levels of specifications

on a basis of a six-level scale (A1, A2, B1, B2, C1, and C2). Revealing a trajectory for language learning and teaching, the Framework also enshrines a novel way for language testing and assessment.

In this context, the Framework has been used as a point of origin to devise the existing language tests, design new ones, and make comparisons amidst them by setting standards (Cephe & Toprak, 2014). To elaborate, the European Language Portfolio (ELP) is introduced as a self-assessment tool that provides learners with the results gained, qualification gathered and competences accomplished through the use of language biography, language passport, and the dossier. Besides, as the Framework has a remarkable effect on language testing and assessment, the alignment of the language tests to the Framework is of utmost importance. Accordingly, the 'Manual for Relating Language Examinations to the CEFR' (CoE, 2009) together with the 'Reference Supplement to the Manual for Relating Examinations to the CEFR' has been introduced to ensure standardization in the alignment process. Correlatively, the Association of Language Testers in Europe (ALTE) has produced the 'Manual for Language Test Development and Examining' to be exploited together with the Framework for the alignment process (CoE, 2011). As a revised version of the 'Users' Guide for Examiners' (CoE, 1996), 'Manual for Language Test Development and Examining' has addressed a more comprehensible approach to set standards for the alignment of the language tests to the Framework.

In the great scheme of this article, basic tenets to establish goodness in foreign language testing and assessment are scrutinized in the light of the Framework together with some practical considerations and pedagogical implications. In doing this, the standards set by the European Association for Language Testing and Assessment (EALTA), ALTE, International Language Testing Association (ILTA) and Association for Educational Assessment (AEA)- Europe are highlighted with an attempt to frame some European benchmarks for language testing and assessment in order to maintain good practice in related practices. Correlatively, this study aims to fill the gap in the literature, and contribute to the field of foreign language testing and assessment by suggesting some European standards that are to be considered; henceforth; the shareholders can be guided and encouraged to address important concepts of language testing and assessment such as test construction, test production, administration and logistics, marking and grading, test analysis, communication with the stakeholders, item writing and the like.

2. The European Benchmarks for Language Testing and Assessment

The current study, hereby, aims to frame some European guidelines in order to maintain good practice in language testing and assessment practices. Therefore, the European guidelines set by the European Association for Language Testing and Assessment (EALTA), the Association of Language Testers in Europe (ALTE), International Language Testing Association (ILTA), and the Association for Educational Assessment- Europe (AEA- Europe), all of which comply with the Framework (CoE, 2001) are elaborated below in detail. In this vein, the guidelines for good practice in language testing and assessment by the EALTA (EALTA, 2006), the code of practice by the ALTE (ALTE, 1994), the guidelines for practice by the ILTA (ILTA, 2007), and the European framework of standards for educational assessment by the AEA- Europe (AEA- Europe, 2012) are probed in order to define how quality profiles for language tests are defined.

2.1. The Manual of the Framework

As one of the members of the CoE, Turkey has been using the CEFR descriptors along with the ELP to form a basis for learning, teaching, testing and assessment. Correlatively, the alignment of the qualifications to the standards defined by the Framework is supervened by the process of devising language tests as described in the Manual. Thus, learning outcomes could be reported and interpreted in a symbolic format by means of levels proposed by the Framework itself. Herein, the Manual caters examination providers with the actual processes of test development, implementation and reporting to situate the tests in use with regard to the Framework. Exploited as a guide for linking a test to the Framework, the Manual also provides the validity of the linking process, and other related claims. In this vein, relating a test to the Framework is basically building an argument on the rationale of validity. Accordingly, validity is ensured by the key concepts of rating, benchmarking, specification, standardization, and familiarization (CoE, 2009).

In that, there has been a tendency towards the alignment of the tests in use to the Framework. However, the Framework is a 'concertina-like tool' (North, Martyniuk, & Panthier, 2010) that acts as a touchstone for teaching, learning and assessment practices (Kavaklı, 2018), which means that it is for users to choose the stepping-stones that are appropriate to the context in use. In doing this, the Manual of the Framework is applied as a reference tool in order to report the learning outcomes within a symbolic format by means of levels. Even so, there may be false interpretations of the linking process. In this context, Alderson (2007) suggests that even the Validation Committee of the CoE has refused to put forward an equivalent mechanism, except for the publicly available reports on the validation of the tests in use to the CEFR; however, the EALTA may act as an organization to deal with such kind of problems with regard to validation and accreditation.

2.2. The Guidelines of the EALTA

In addition to the Manual, the EALTA has set some guidelines for understanding of the theoretical principles of language testing and assessment, and sharing of the aforementioned practices throughout Europe, which is labelled as the EALTA Guidelines for Good Practice in Language Testing and Assessment (EALTA, 2006). For the dissemination of the Framework together with the ELP, the EALTA serves for the improvement of the quality of the language testing and assessment practices without any diminution of cultural identity. Therefore, it is targeted to develop a public understanding for developing links between those who are interested in language testing and assessment, and taking joint part in activities for the enhancement of the related practices. With these in mind, the EALTA addresses three types of audience: (1) pre-service teachers, (2) in-service teachers, and (3) test developers in national and/or institutional testing units or centers. In doing these, the EALTA strives for the respect for the students and/or examinees, validity, reliability, fairness, responsibility, and collaboration amidst the allied parties.

Juxtaposing the tenets of inclusiveness, accountability and transparency, the EALTA also introduces some main considerations for the pre-defined types of audience in order to keep goodness in language testing and assessment. Accordingly, the considerations for teacher pre-service and in-service training in language testing and assessment cover the nature of the assessment, trainees' awareness in the principles of language testing and assessment, training in language and assessment to fill the void between theory and practice, the essentials of marking and grading, the appropriate interpretation of the assessment results, the accuracy of the assessment procedure, and the overall evaluation of the trainees' knowledge on language testing and assessment.

On the other hand, the considerations for classroom testing and assessment involve the clarification of the in-class applications within a triangle of pupils, parents and test developers. Hence, classroom testing and assessment is framed by the specification of the assessment purposes, assessment procedures, and possible consequences (EALTA, 2006). Herein, assessment purposes are defined with their relations to the curriculum in practice, and test specifications. Correlatively, assessment procedures embrace the procedural design, the eligibility of the designed procedures to the learners' expectations and levels, the information retrieval and storage process, fairness and accuracy in assessment procedure, accreditation between cross-over testing and assessment applications, and learners' viewpoints on the overall procedure. One more to note, possible consequences involve the subsequent use of the assessment results, the maneuvers taken to enhance learning, the kinds of learner-based feedback, and the ways to improve classroom practice for the following assessment procedures.

Additionally, the EALTA has also purported some basic considerations for test development in national and/ or institutional testing units or centers. In this context, answers are sought to the questions aligned under the major topics of "(1) test purpose and specification, (2) test design and item writing, (3) quality control and test analyses, (4) test administration, (5) review, (6) washback, and (7) linkage to the CEFR" (EALTA, 2006, p. 4). Since the EALTA members who are involved in the test development process are to clarify themselves to the related stakeholders, answers to the above-listed headings are assumed to be provided; therefore, test developers could be bolstered to employ themselves in a collaboration with decision-makers in order to foster the quality of the current language testing and assessment practices, and/or systems.

2.3. The Code of Practice, Minimum Standards and Portfolios of the ALTE

In addition to the EALTA, the ALTE endeavors for putting common standards for language testing and assessment by creating an environment where multilingualism is supported for the goodness of cultural and linguistic diversity within and beyond Europe. In this sense, 'setting standards' and 'sustaining diversity' are the primary objectives of the ALTE, who has a bunch of members together with the institutional and individual affiliates to represent the testing of a great number of different languages (ALTE, 2012). In terms of setting standards, the ALTE has its own objectives which are enounced for the preservation of the cultural and linguistic enrichment to support multilingualism (Kavaklı, 2018). Besides, the ALTE applies for the Code of Practice, Minimum Standards, and Portfolios which are all developed by the ALTE itself in order to subsidize the pre-set standards (ALTE, 2007).

While canalizing into two majors (i.e. setting standards and sustaining diversity), the ALTE exploits a compile of common standards of language testing and assessment, namely test development, marking and grading, item writing, reporting the test results, and test administration and analysis. In doing this, the ALTE appeals to its own quality indicator, the ALTE Q-mark, in order to seek for the accessibility of the quality standards. Correlatively, seventeen minimum standards are defined by the Association itself within the areas of test construction, test analysis, marking and grading, administration and logistics, and communication with the stakeholders. After a rigorous auditing process, the findings are reported, and announced whether the test is verified by Q-mark. A test with a Q-mark provides test takers and/or users with the assurance of validity, reliability and appropriateness. Apart from Q-mark, the ALTE has purported some guidelines with the 'Code of Practice', some criteria to enable effectiveness in language testing and assessment with the 'Minimum Standards', and some self-assessment tools for the encouragement of the independent and autonomous learning environment with the 'Portfolios'.

To elaborate, the Code of Practice is introduced to depict the standards in language testing and assessment for the ALTE members (ALTE, 1994). Within, the responsibilities of the language test providers, takers and users are highlighted for the comparability of the quality standards. Herein, the ALTE offers two types of responsibilities around the core concepts of fairness, appropriateness, informativeness and availability: (1) responsibilities of the members, and (2) responsibilities of the test users. The responsibilities of the members are defined as to label the features of the tests, such as test purpose, test specifications, target population, test development, test measurement, and test administration; henceforth, the appropriateness of the test to the test takers could be ensured. This also leads the path towards providing information to the test users, who are at the position to decide on the test that fits best to their needs. Moreover, the members are also expected to render swift responses for the interpretation of the test results. Herein, it is important to determine the criteria for passing grades according to the test takers' performances so that test users could interpret the test results appropriately and accurately.

Another key concept of the Code of Practice is fairness. Accordingly, test fairness is to be ensured with no discrimination due to ethnicity, gender, race, or any handicapping situation. A misunderstanding may arise from the test materials in use; therefore, these materials are to be reviewed, and revised, if needed. Additionally, the language in use together with the test content are to be one-size-fits-all in order to preclude test sensitivity, albeit appealing to different backgrounds. Besides, the test takers are to be informed equally about the rubrics in use, the release of the results, and copyright issues, as well.

In other respects, the ALTE makes use of two different types of portfolios, namely the ELP, and European Association for Quality Language Services (EAQUALS)-ALTE Portfolio. The ELP has been introduced as a result of a project to cater learners with their own recordings of formal and informal learning practices. Ensuring integration and mobility throughout Europe, and supporting life-long learning, the ELP helps learners to monitor their learning experiences both inside and outside of the classroom. In doing these, the ELP encourages learners to convey up-to-date information at regular intervals through the use of language passport, language biography, and the dossier. On the other hand, the EAQUALS-ALTE Portfolio is a version which is developed by the ALTE members, and originated from the ELP and EAQUALS. Functioning like the ELP, the EAQUALS-ALTE Portfolio also promotes learner autonomy, responsibility and awareness. Establishing a fair link between the Framework and tests accredited by the ALTE, the ALTE Portfolio is now available in seven languages in both paper-based and electronic formats.

2.4. The Code of Ethics and Guidelines for Practice of the ILTA

ILTA is constituted as a group of well-recognized scholars together with the practitioners from the field of language testing and assessment. The primary goal is to identify the characteristics of a language tester in order to enhance the on-going language testing and assessment implementations. In doing this, the dissemination of information is enabled amidst its members through two types of resources: (1) the Code of Ethics, and (2) the Guidelines for Practice.

Accordingly, the Code of Ethics is framed by the ethical principles proposed for language testers. These principles are molded with justice, beneficence and non-maleficence, and respect for autonomy and civil society (ILTA, 2000). Herein, ILTA pinpoints nine fundamentals in order to accomplish the aforementioned principles through the use of 'ought-to-do'es and 'ought-to-not-do'es. Failure to fall in line with these principles leads to the withdrawal of the ILTA membership by the decision of the ILTA Ethics Committee.

To probe into these principles, the very first principle is holding respect to keep test takers' esteem and privacy. No issues of discrimination in terms of ideology, spiritual matters and politics could have an influence on the treatment towards test takers. Besides, testers are supposed to safeguard the test takers' rights. For the confidentiality of the results, testers are also expected to keep information on the test results. Correlatively, it is indicated that the ethical standards should be pursued in any academic inquiry since language testing and assessment regards human as the sample in a research activity, and/or experiment. Herein, it is also guaranteed by the Code of Ethics that the language testers are provided with professional support in order to follow the latest developments within the scope of language testing and assessment. Therefore, they are encouraged to read new publications in well-respected journals, and to participate in workshops, conferences, seminars, annual meetings and the like. In doing these, language testers are expected to share a common knowledge with their colleagues to provide integrity in language testing and assessment as a profession, which basically amounts to quality improvement through collaboration.

In addition to the Code of Ethics, ILTA has also proposed the Guidelines for Practice, which was first announced at the annual ILTA meeting in Ottowa in 2005. It was revised at another ILTA meeting held in Barcelona in 2007, and finally, the latest version was approved in 2010. Accordingly, the Guidelines for Practice are constituted by two major sections concerning the responsibilities of the test developers and users, and those of test takers (ILTA, 2007). To elaborate, the test developers and users are expected to develop an understanding for test constructs to ensure validity. Besides, test purpose and specifications should be clearly stated, and the test results should be accurate and comparable to elicit test reliability. Thus, inter- and intra-rater reliability estimations are gathered following the procedures of marking and grading. If there is more than one test form, it is also expected to estimate and present inter-form reliability. In order to ensure equality, test takers are to be treated in the same way, and the interpretations of the test results should be in an accurate rapport inter see.

Concomitantly, test results are to be kept safely, especially in the administration of the high stakes examinations. For the preparation of these examinations, it is advised to depend on a language testing theory. Besides, it is recommended to receive support in item writing if needed (e. g. taking help from someone with a high level of proficiency in the target language, or from a native speaker for the testing of non-native speakers of that language). The results should be announced with satisfactory information on how to use them for further purposes. Additionally, it is expected to put the test results in storage for sustainability, and progression to conduct quality control analysis. In brief, the responsibilities of the test result users are depicted as making good interpretations and opting for fair judgements on the test results, scrutinizing the limitations of the test results, putting forward decisions on the accuracy of the test results, and estimating the standard error of measurement (SEM) before reaching at final decisions on the test results.

On the other hand, the responsibilities of the test takers are defined as being informed about the overall testing process, key features of the test in use, the results of not taking the test, privacy statements, and being aware of their own responsibilities and rights regarding language testing and assessment process. Herein, they are expected to follow the directives, albeit consult to a tester in case of difficulty.

With no favor bestowed due to gender, age, ethnicity, religion and other characteristics, the test takers are given fundamental right of respect without any discrimination. Therefore, test takers are welcomed with respect to accentuate any concern and/ or dispute about the language testing and assessment process.

Apart from these, ILTA has its own resources for those who are professionally involved in language testing and assessment process. In this vein, ILTA caters its members with a regularly updated and online presented bibliography of dissertations written on language testing and assessment. ILTA also organizes research activities such as seminars, workshops, conferences and webinars attached to a calendar of events. One more to note, ILTA also provides a list of academic journals by which recent publications in the field of language testing and assessment can be pursued.

2.5. The Core Elements, Guiding Principles and Instrument of the AEA- Europe

The AEA- Europe offers a platform to discuss the novelties in educational assessment in Europe with a solidarity between individuals and organizations. In doing this, individuals' professional development is promoted by means of conferences and joint projects held across Europe, and cooperation between related organizations is cherished. Constituted by different types of committees such as scientific committee, publications committee, professional development committee, conference organizing committee and other ad hoc committees, the AEA- Europe serves as a Council to improve assessment products and practices; therefore, to develop a new understanding on the probable effects of these products and practices in the field of education by means of a myriad of activities. In this context, the AEA- Europe applies for the 'European Framework of Standards for Educational Assessment' (AEA- Europe, 2012) to accomplish the aforementioned aims.

This Framework serves as a guide to benchmark current and further assessment practices by fostering transparency for educational authorities together with the users via core elements, guiding principles, and the instrument. Briefly, the core elements are defined as the goal construct, nature of evidence of tasks, gathering evidence, capturing outcomes, decision-making, interpreting and reporting results, and evaluation and next iteration. On the other hand, the guiding principles can be listed as focusing on educational assessment, fitting for a European environment, emphasizing ethics, fairness and the rights of the individual, addressing essential quality concerns of validity, reliability and impact on stakeholders, and supporting learning, test development and review. To note more, the instrument is composed of standard requirements, methods, and samples of evidence to be in good accord with the Framework.

To probe into the core elements, the test development cycle is composed of seven standard requirements (AEA- Europe, 2012): "(1) defining the goal, (2), identifying the nature of evidence and of tasks, (3) gathering evidence, (4) capturing outcomes, (5) decision-making, (6) interpreting and reporting results, and (7) evaluation and next iteration" (p. 9). Accordingly, what is intended to be measured by the test should be clearly stated. For further amendments, strengths and weaknesses experienced during the assessment process should be highlighted. The test content should be laced with clear-cut instructions with a sufficient number of representations covering knowledge, skills, and aptitudes. Thus, the construct-relevant variance should be eliminated from discrimination against test takers to make valid inferences. However, comparisons of the test results should be made in order to review and revise the tasks in use, and other related test materials.

Within a European perspective of worldwide interest, the Framework also integrates new approaches of language testing and assessment into current traditions under the theme of fitness-for-purpose. Disseminating aspects of quality in educational assessment, the Framework addresses the cornerstones of impact on stakeholders, reliability, practicality, and validity. Besides, the Framework embowers the rights of the individual through ethics and fairness. Therefore, it is assumed that the test results reflect a certain extent of credibility. It is also advised that providing feedback is of paramount importance in order to evaluate, and enhance the programs in use. In doing this, the Framework exploits the instrument that addresses the levels of standard requirements, methods, and samples of evidence on a generic-based and illustrative way, albeit not on a prescriptive basis (Kavaklı, 2018).

3. The Core Components of Language Testing and Assessment Procedure

The European guidelines above explained in detail embody the core tenets of test construction, administration and logistics, test production, communication with stakeholders, marking and grading, and item writing. To note more, unifiers as the integrative elements of each tenet are elaborated below together with some practical recommendations to pursue goodness in language testing and assessment practices through the dissemination and exploitation of these European standards.

Test construction entails the basic premise that language testing and assessment process should nestle a theoretical construct, in which a clear-cut description of the test purpose and context of use is scrutinized. Test construction also addresses the issues of reliability, validity, stability and consistency. Enabling this, language testing and assessment process should also refer to an external reference system through the linkage to the CEFR. On the other hand, administration and logistics should be laced with clear, established and transparent procedures and/or regulations. In order to provide security for the handling of testing documents, appropriate support systems should be used in line with the current data protection legislation. Besides, candidates with special needs and care should be given support during the testing and assessment process.

For marking and grading, accuracy is regarded as one of the essentials of the process. Besides, accurate methods should be applied for data collection and analysis procedures. For instance, assessing speaking embodies more distinctive features than assessing the achievement of candidates in writing. To elaborate test analysis, item-level data should be collected from an adequate sample of candidates for computing item difficulty, item discrimination, standard errors of measurement, and test reliability. Besides, the continuum of measurement should be purified from any influence rooted in age, ethnic origin, gender, native language and the like. One more to note, a clear and prompt communication system amidst the allies is desired regarding the concept of communication with stakeholders. In doing this, information provided to the stakeholders should be appropriate, reliable, and laced with the issues of test purpose and context of use.

4. Conclusion and Suggestions

As highlighted, the European standards form a basis for language testing and assessment with their guiding principles, main considerations, specifications and resources. A panorama to those European standards is drawn in depth above. Basic tenets to establish good practice in foreign language testing and assessment are also elaborated in detail. Correlatively, some suggestions to establish and maintain good practice in related implementations are given below in order to draw a frame of practical recommendations.

In this vein, general principles to be applied can be listed as collaboration amidst the parties involved, fairness, reliability, respect for the candidates, responsibility, and validity. Those who are involved in the language testing and assessment process are expected to clarify themselves: Are they curriculum developers, practicing teachers, and/or trainees? Besides, the purpose of the assessment, its relation to the curriculum, and test specifications are to be known and discussed to detect how well the curriculum is covered. Besides, test review and revision of the test items are to be followed in order to initiate change(s) in the current practice. With the aim of boosting professionalism and maintaining quality in implementations, the washback effect is to be specified; in that, learning is not solely bounded to the final score. Thus, possible consequences of language testing and assessment process can be figured out through the ecological model (Hubley & Zumbo, 2011), in which language testing and assessment process is regarded as something *in vivo*, albeit not *in vitro*.

Additionally, language testing and assessment practices are expected to cover the 21st century skills, such as information-processing (Van Nijlen & Janssen, 2014). In doing these, the CEFR is recommended as a guide to create opportunities for language testing and assessment, and, thereto, to enhance the quality of related practices (Finch; 2009; Jones, 2007; Kavaklı & Arslan, 2017; Saville, 2005). Besides, relating tests to the CEFR is the *sine qua non* before delving into the depths of the language testing and assessment process. Therefore, building a core knowledge in language testing and assessment has mushroomed as a must with the arrival of the CEFR. Herewith, language assessment courses are expected to be conducted cautiously and efficiently, albeit not haphazardly, in order to build a ground for language assessment literacy (Inbar-Lourie, 2008; Mirici & Kavaklı, 2017). If needed, more professional training can be

provided for those who want to develop a better understanding for related practices (Malone, 2017). Consequently, assessment is at the center of language learning; therefore, the Framework can be initiated as a model, and as a bridge of harmonization between the current and future implementations for quality assurance in foreign language testing and assessment.

References

Alderson, J. C. (2007), The CEFR and the need for more research. *The Modern Language Journal (MLJ)*, *91*(4), 659-663.

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

Association of Language Testers in Europe (ALTE). (1994). *Code of Practice*. Retrieved from https://www.alte.org/resources/Documents/code_practice_en.pdf

Association of Language Testers in Europe (ALTE). (2007). *Minimum Standards*. Retrieved from https://www.alte.org/resources/Documents/minimum_standards_en.pdf

Association of Language Testers in Europe (ALTE). (2012). *Constitution for the Association of Language Testers in Europe*. Retrieved from http://www.alte.org/docs/constitution-2012.pdf

Cephe, P. T. & Toprak, T. E. (2014). The Common European Framework of Reference for Languages: Insights for language testing. *International Online Journal of Education and Teaching (IOJET)*, *10*(1), 79-88.

Council of Europe (CoE). (1996). *Users' Guide for Examiners*. Strasbourg: Language Policy Division.

Council of Europe (CoE). (2000). *Resolution on the European Language Portfolio*. Retrieved from http://culture.coe.int/portfolio

Council of Europe (CoE). (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge University Press: Cambridge.

Council of Europe (CoE). (2009). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual, Strasbourg: Language Policy Division.

Council of Europe (CoE). (2011). Manual for Language Test Development and Examining: For Use with the CEFR. Strasbourg: Language Policy Division.

European Association for Language Testing and Assessment (EALTA). (2006). *EALTA Guidelines for Good Practice in Language Testing and Assessment*. Retrieved from http://www.ealta.eu.org/documents/archive/guidelines/English.pdf

Finch A. E. (2009). Europass and the CEFR: Implications for language teaching in Korea. *English Language and Literature Teaching*, *15*(2), 1-23.

Hubley, A. M. & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. Social Indicators Research, 103(2), 219-230.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25, 385-402.

International Language Testing Association (ILTA). (2000). *Code of Ethics*. Retrieved from http://c.ymcdn.com/sites/www.iltaonline.com/resource/resmgr/docs/ilta_code_english.pdf

International Language Testing Association (ILTA). (2007). *Guidelines for practice in English*. Retrieved from http://c.ymcdn.com/sites/iltaonline.siteym.com/resource/resmgr/docs/ilta_guidelines.pdf

Jones, N. (2007). Assessment and the national languages strategy. *Cambridge Journal of Education*, 37, 17-33.

Kavaklı, N. & Arslan, S. (2017). Applying EALTA guidelines as baseline for the foreign language proficiency test in Turkey: The case of YDS. *International Journal of Curriculum and Instruction (IJCI)*, *9*(1), 104-118.

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

Malone, M. E. (2017). Training in language assessment. In E. Shohamy, L. G. Or & S. May (Eds.), *Language testing and assessment. Encyclopedia of Language and Education* (3rd Ed.). Cham: Springer, 225-239.

Mirici, İ.H. & Kavaklı, N. (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.

North B., Martyniuk, W. & Panthier, J. (2010). Introduction: The manual for relating language examinations to the Common European Framework of Reference for Languages in the context of the Council of Europe's work on language education. In M. Martyniuk (Ed.), *Aligning tests with the CEFR: Reflections on using the Council of Europe's draft manual* (Cambridge: Cambridge University Press, 1–17.

Saville, N. (2005). An interview with John Trim at 80. *Language Assessment Quarterly*, *2*(4), 263-288.

Trim, J. L. M. (2005). *The role of the Common European Framework of Reference for Languages in teacher training*. Retrieved from https://www.ecml.at/Portals/1/resources/John%20Trim%20collection/trim_CECR_20050916.pdf

Valax, P. (2011). *The Common European Framework of Reference for Languages: A critical analysis of its impact on a sample of teachers and curricula within and beyond Europe* (Unpublished doctoral thesis). University of Waikato: Hamilton.

Van Nijlen, D. & Janssen, R. (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. Retrieved from https://lirias.kuleuven.be/handle/123456789/471002