MDPI

Article

A Toolkit for the Investigation of Greek EFL Teachers' Assessment Literacy

Eirini Gkogkou * and Ifigenia Kofou *

Department of Teaching English as a Foreign/International Language, Hellenic Open University, 26335 Patras, Greece
* Correspondence: eirinigkogkou@yahoo.com (E.G.); ikofou@gmail.com (I.K.)

Abstract: The role of assessment in a learner-centred environment is considered to be significant for both learners and teachers. Most of the time, however, it is used in traditional ways and ignores learners' individual needs. Based on the results of a survey conducted in 2019, in which a questionnaire was administered to a hundred and twenty EFL teachers, the present study aims to investigate Greek EFL teachers' responses to communicative testing techniques and their awareness of assessment methods and principles. The aforementioned survey revealed that the majority of EFL teachers in the Greek educational context use traditional tests to assess their students and, although they are aware of alternative assessment methods and the benefits they offer, they fail to employ them. Thus, a 106-item tool was created in order to help teachers design, develop, and critically evaluate tests, as well as reflect on their assessment techniques to promote the use of alternative assessment and supplement the teachers' theoretical knowledge and experience. Ninety-three EFL teachers evaluated themselves and rated their practices through the toolkit to find out the type of assessors that they are. The findings revealed that a lot of the participants are aware of the key principles of assessment and try to assess the four skills in a communicative and authentic way to a great extent, but most of them are mainstream assessors. The findings can be used to help design samples of authentic tasks for all skills and assessment-related teacher training material.

Keywords: alternative assessment; communicative testing; authenticity; third-generation testing; assessment toolkit

1. Introduction

Fundamentally, assessment is an integral part of the learning process. It is interwoven with teaching and learning, and involves making judgments about learners (Nunan 1990) and monitoring their development (Hedge 2000) in order to assess their needs and tailor instruction to optimize learning. As McNamara and Roever (2006) assert, language testing in education dictates what is to be taught, what is to be valued in instruction, and what becomes the focus of activity (Swan et al. 2006). This shows that assessment plays an important role in many people's lives (McNamara 2000), and teachers therefore need to be "competent in the principles and practice of language assessment" (Harding and Kremmel 2016).

Traditional summative assessment or standardized testing that aims to measure the students' ability and knowledge (Brown 2003), using product-oriented techniques, seems to be inadequate at measuring ongoing student development. In many cases, "teaching to the test" (Bowers 1989) in order to allocate a mark overlooks other parameters, such as the learner profile, individual needs and preferences (Tsagari 2004), and lacks authenticity and contextualization with a negative washback effect on learning. Moreover, it has a negative washforward effect, since product-oriented teaching does not prepare learners for real-life situations (Widdowson 1976). In particular, more and more EFL learners in Greece are interested in acquiring a language certificate (Papageorgiou 2009), which leads to exam-oriented teaching based on the final product and not the process of learning.

Considering this, assessment in public and private language schools in Greece seems problematic and action needs to be taken to promote teachers' professional development in alternative assessment and communicative testing. Dissatisfaction caused by the limitations of traditional testing has paved the way for alternative assessments that encourage metacognition, reflection, and self-directed learning, and which can be integrated with instruction, emphasizing both the product and the process of language learning (Chirimbu 2013). Contrary to traditional practices, alternative assessment and communicative testing can be used as a means of reflection and portraying advancement or lack thereof (Baker 2016). Communicative, process-oriented curricula, in conjunction with alternative methods for collecting information and student-centered ways of assessment, emphasize the importance of integrating assessment with instruction. There are a variety of alternative methods to assess continuous student progress and address the problems with standardized tests (Griva and Kofou 2017). Alternative/formative assessments can be used as aids in the learning process or as decision springboards for the steps that follow instruction, together with self- and peer-assessment (Bøhn and Tsagari 2021), to describe ongoing student-related information, as well as to make evaluative decisions (Brown 2003; Ioannou-Georgiou and Pavlou 2003; Tsagari 2011).

"Assessment literacy", a relatively new term coined by Stiggins (1991), refers to how literate teachers are in regarding what, why, and how they assess in order to generate "good examples of student performance" (p. 240). Alternatively, Popham (2018) describes this concept as "an individual's understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (p. 2). Clearly, assessment literacy can empower teachers (Grabowski and Dakin 2014) who need to be aware of the assessment purpose and tools they use, the testing conditions, and the utility of the learners' results, as well as the importance of their decision making (Inbar-Lourie 2008). The research conducted by López and Bernal (2009) indicates that trained language teachers use different practices of assessment to improve teaching and learning, whereas teachers with no training in language assessment used assessment as a way to solely obtain grades. Thus, teachers need to be literate in assessment and understand their critical role in the assessment process. Added to this, similar studies that have been conducted to find out the participants' training in assessment (Vogt and Tsagari 2014) point out that teachers are not equipped with sufficient knowledge on testing and assessment and commonly regard assessment as an activity separate from teaching, equal to allocating a grade or score. As Herrera and Macías (2015) claim, teachers need to "have a working knowledge of all aspects of assessment to support their instruction and to effectively respond to the needs and expectations of students, parents, and the school community" (p. 303). Therefore, appropriate pre-service and in-service training needs to take place to offer sufficient education in language assessment that will help teachers employ more effective means of assessment, taking into account the fact that alternative assessment methods have been included in the evaluation process of public schools (Government Gazette 140 2021).

With this in mind, and based on the results of a survey (Gkogkou 2019) conducted in 2019 as part of one of the authors' master's degree in the Hellenic Open University (https://apothesis.eap.gr/bitstream/repo/42987/1/103390_GKOGKOU_EIRINI.pdf, accessed on 22 October 2021), by means of a questionnaire administered to a hundred and twenty EFL teachers, in conjunction with structured interviews, the present study aims to investigate Greek EFL teachers' responses to communicative testing techniques and their awareness of assessment methods and principles. Specifically, it was revealed that the majority of EFL teachers in the Greek educational context use traditional tests to assess their students and, despite the fact that they are aware of alternative assessment methods and the benefits they offer, they fail to employ them (Gkogkou 2019). It was also shown that most teachers resort to discrete-point testing items, which test language in a rather fragmentary way and focus on language competence and usage rather than use, and not in the context of authentic real-life tasks, which require a full, authentic task environment and promote integrative language (Gkogkou and Kofou 2020). Thus, there was a tool created

in order to help the abovementioned teachers design, develop, and critically evaluate tests, as well as reflect on their assessment techniques to promote assessment literacy and supplement the teachers' theoretical knowledge and experience. The purpose of this study is to implement the tool and measure the teachers' awareness of language assessment. The toolkit is constructed based on specific criteria, according to relevant literature (Alderson and Banerjee 2001; Bachman and Palmer 1996; West 2004), and aims to urge teachers to reflect on their assessment methods. Moreover, the toolkit serves as a guide to promote the use of alternative assessment and authenticity in teaching, learning, and assessing in a foreign language classroom, since "educational communities lack empirical evidence about the value of many influential assessment instruments" (Alderson and Banerjee 2001). The tool can be used by pre-service and in-service teachers as a guide and as a self-evaluation instrument to encourage teachers to rethink their roles and develop professionally.

2. Materials and Methods

When the respondents to the aforementioned survey (Gkogkou 2019) were asked if they would be open to using an assessment instrument with specific criteria for evaluating and designing tests, more than 80% were positive, and more than 90% were willing to apply alternative assessment forms if they were supplied with appropriate materials and guidance. Thus, the researchers' consequent aim was to develop an assessment tool to supplement the teachers' theoretical knowledge and experience and enhance the self-assessment procedure. The toolkit was based on the research conducted in 2019, in which a questionnaire was administered to a hundred and twenty EFL teachers, in conjunction with structured interviews, which investigated the beliefs, perceptions, and practices of EFL teachers regarding assessment, as mentioned previously. This present paper aims to implement the previously constructed toolkit in order to explore the beliefs and practices of EFL teachers regarding assessment, focusing on the purpose, forms, and processes used; that is, whether the teachers' assessment methods actually involve the use of authentic and alternative forms of assessment, or whether more traditional practices are preferred. Similarly, Greek EFL teachers' responses to communicative testing techniques and their awareness of assessment methods and principles are also investigated. At the same time, the toolkit that we created serves as a measurement tool and a research instrument designed to measure the language assessment knowledge of the EFL teachers who have been participating in the survey. It aspires to help teachers design, develop, and critically evaluate tests, as well as help them reflect on their assessment techniques. Subsequently, it can be used by pre-service and in-service EFL teachers to raise their awareness of language assessment and can support them to make informed decisions when assessing students' learning. It can also be regarded as a teachers' guide and self-evaluation tool, which can be used several times during the school year in order to observe and reflect on their development and growth.

The checklist was developed over three stages. First, we collected an initial set of items based on West's (2004) existing recommendations for constructing and administering tests. Based on these recommendations, we developed the assessment criteria that were afterwards grouped into categories and divided into two parts. Finally, clear language descriptions were used to provide objective evaluations of the three different types of assessors.

The toolkit (Appendix A) is divided into two parts and can be regarded as a form of self-assessment for the teachers. Determining what teachers do or do not know with regard to language assessment was the starting point of the study, with the aim of encouraging them to become more assessment literate. Therefore, the first part of the toolkit aims to provide both novice and experienced teachers with input that concerns the basic principles of language assessment in order to help them understand how tests are constructed. Teachers can review or familiarize themselves with fundamental principles of language testing for describing, categorizing, and evaluating published tests, as well as designing their own tests. Specifically, the first part consists of 130 randomly ordered criteria, which

are provided in the form of a checklist and can be used to evaluate and rate the teachers' practices. Furthermore, key assessment principles, such as authenticity, reliability, validity, practicality, washback and washforward effects, feedback, and reflection, are included and investigated in this part, based on the relevant literature, as mentioned above. For example, according to Bachman and Palmer (1996, p. 18), a model of test usefulness should include qualities such as reliability, construct validity, authenticity, interactiveness, impact, and practicality, which can also apply to alternative assessment techniques.

The second part deals with testing each of the four communicative skills (writing, speaking, reading, and listening), each by itself or in combination with others (OECD 2021), using a checklist that shows teachers what to take into account when assessing the students' receptive and productive skills in a communicative and integrated way (Fulcher 2000, 2012). It is divided into four sections, one for each skill, and includes specific criteria that can help language teachers evaluate tests. The distinction of the four skills that is usually drawn in large-scale standardized testing and textbooks, however, does not invalidate the integration of skills that is desirable both in classroom and testing settings, which is also evident from the fact that many of these criteria may overlap. These criteria refer to text and task authenticity, the types of tasks and processes, and rating. Teachers can take these criteria into consideration when exploring the testing of all four skills to understand how tests are constructed, which will also aid them to design their own assessment tasks. Thus, teachers will be equipped with useful insights into test evaluation and design that will help them design their own tests for assessing learners in more authentic and communicative ways.

To that end, the toolkit (Appendix A) was practised by ninety-three EFL teachers out of the one hundred and twenty who participated in the original survey, corresponding to 77.5% of them and very close to the 80% of the EFL teachers who had stated their willingness to apply the instrument. Actually, random sampling was used to ensure the generalizability of the findings (Marshall 1996) by minimizing the possibility for bias and increasing the credibility of the results (Patton 2002). Thus, the sample included English language teachers who either work in the private or public sector and who are university degree or simply C2 language certificate holders (93.5% female, 21–50 years old, 60% with a master's degree, 20% with a bachelor's degree). The aim was to understand the criteria that should be assigned to testing in order to become better prepared for assessing learners, and therefore the present study focuses on the investigation of their attitude towards the integration of the particular assessment toolkit. It aims to indicate the methods, techniques, and types of traditional and alternative assessment that teachers use and outline the criteria that they take into account when assessing learners. The criteria aim to unfold their beliefs about assessment, concerning the purposes, reasons, and types of assessment they employ, as well as the feedback they provide to students. By using the aforementioned toolkit, the survey participants evaluate themselves and respond on the basis of a three-point scale (yes/to some extent/no). After responding, the answers are analyzed by using a simple formula as follows:

$$P = f/n \times 100\%$$

P = Percentage
f = Frequency
n = Number of questions

By using the aforementioned toolkit, with no psychometrics involved, the teachers evaluated themselves via the three-point checklist, reflected on their assessment methods and rated their practices to find out the type of assessors that they are. The final aim was to encourage and invite teachers to use the specific criteria in their future assessment practices.

3. Results

The population of both the initial survey and this present study was important for the reliability and validity of the research results in order to draw robust conclusions. For this reason, the group of the participants was carefully defined. We used purposeful sampling so that we could identify and select "information-rich cases", in line with the

research purpose (Patton 2002). On the one hand, the participants were knowledgeable and experienced with the subject of interest, while on the other hand, random sampling was used to ensure the generalizability of the findings (Marshall 1996).

In the present study, the assessment toolkit was dispensed to ninety-three EFL teachers, who either work in the private or public sector. The data collected after evaluating themselves and rating their practices through the toolkit offered insights into the type of assessors that the majority of teachers are. The findings revealed that a lot of the participants are aware of the key principles of assessment and try to assess the four skills in a communicative and authentic way to a great extent. However, the majority of them are mainstream assessors who feel that while they should be using authentic assessment more often, there are reasons that keep them back.

The findings can be used to help design samples of authentic tasks for all skills, but also for teacher training in assessment modes and teacher professional development programmes related to language assessment.

3.1. Principles of Testing

The first part of the questionnaire dealt with the principles of testing applied to the respondents' tests. In particular, the majority of the participants believe that the testing content is similar to the teaching content (85%). Specific teachers assume that testing promotes autonomous and self-directed learning and learner-centred assessment with a clear purpose, including tasks that suit the learners' abilities. However, skills are not highly integrated, and students' motivation and involvement are not triggered. It also seems that assessment is not regarded as a shared responsibility, and students cannot express their opinion on how they will be assessed, nor are they given a choice of assessment tasks (Table 1). Significantly, in Greek public schools, the introduction of the item bank will likely lead to rather predictable close-ended assessment tasks (Government Gazette 111 2020).

Table 1. Testing.

Similar testing and teaching context	Frequency	Percent
No	1	1.1
To some extent	13	14.0
Yes	79	84.9
Total	93	100.0
Integrated skills	Frequency	Percent
No	17	18.3
To some extent	35	37.6
Yes	41	44.1
Total	93	100.0
Autonomous and self-directed learning	Frequency	Percent
No	12	12.9
To some extent	56	60.2
Yes	25	26.9
Total	93	100.0
Learner-centred assessment	Frequency	Percent
No	7	7.5
To some extent	37	39.8
Yes	49	52.7
Total	93	100.0
Students' motivation and engagement	Frequency	Percent
No	6	6.5
To some extent	49	52.7
Yes	38	40.9
Total	93	100.0

Table 1. *Cont.*

Students' opinion of assessment	Frequency	Percent
No	40	43.0
To some extent	39	41.9
Yes	14	15.1
Total	93	100.0
Assessment as a shared responsibility	Frequency	Percent
No	33	35.5
To some extent	32	34.4
Yes	28	30.1
Total	93	100.0
Students' choice of assessment tasks	Frequency	Percent
No	49	52.7
To some extent	25	26.9
Yes	19	20.4
Total	93	100.0
Assessment tasks suitable to students' abilities	Frequency	Percent
No	7	7.5
To some extent	35	37.6
Yes	51	54.8
Total	93	100.0

Regarding authenticity (Table 2), almost all of the participants, to some or great extent, put emphasis on communicative language, believing that the language in the tests is natural, the test items are contextualized and emphasize the communicative view of the language, and that the topics and situations are interesting, enjoyable, humorous, and relevant to the learners' age and level. About four out of ten participating teachers think that the student's ability to apply knowledge to real-life problems is tested and that real-world situations and processes, useful for everyday life, are replicated. The task rubrics are contextualized to offer a more realistic and communicative view of the language, but genuine materials, found in the real world and not for testing purposes, or some thematic outline, are not used to a great extent. It is evident, therefore, that assessment practices do not follow communicative teaching practices to a great extent as EFL teachers allege, probably because testing is seen as summative assessment and related to assigning a grade and not for spotting learners' weaknesses and needs.

Table 2. Authenticity.

Emphasis on communicative language	Frequency	Percent
No	1	1.1
To some extent	47	50.5
Yes	45	48.4
Total	93	100.0
Natural language	Frequency	Percent
No	2	2.2
To some extent	37	39.8
Yes	54	58.1
Total	93	100.0
Contextualized test items	Frequency	Percent
No	4	4.3
To some extent	36	38.7
Yes	53	57.0
Total	93	100.0
Contextualized rubrics	Frequency	Percent
No	3	3.2
To some extent	53	57.0
Yes	37	39.8
Total	93	100.0

Table 2. *Cont.*

Genuine materials	Frequency	Percent
No	16	17.2
To some extent	46	49.5
Yes	31	33.3
Total	93	100.0
Interesting and relevant test topics	Frequency	Percent
No	4	4.3
To some extent	38	40.9
Yes	51	54.8
Total	93	100.0
Thematic organization	Frequency	Percent
No	30	32.3
To some extent	41	44.1
Yes	22	23.7
Total	93	100.0
Real-world assessment tasks	Frequency	Percent
No	6	6.5
To some extent	53	57.0
Yes	34	36.6
Total	93	100.0
Assessment close to real-life problems	Frequency	Percent
No	6	6.5
To some extent	49	52.7
Yes	38	40.9
Total	93	100.0
Replication of real-world situations	Frequency	Percent
No	8	8.6
To some extent	49	52.7
Yes	36	38.7
Total	93	100.0

Regarding validity (Table 3), more than 90% of the participants believe that the test guidelines are clear, the timing of the test is appropriate, students are notified in advance, and appropriate review and preparation for the test is offered. About eight out of ten teachers think that, before the test, beneficial strategies are frequently suggested, the learning objectives are identified in the assessment practices and appropriately represented in the test, and that the test's difficulty level is appropriately pitched. According to the two-thirds of the participants, assessment is tied significantly to curricular practices and supports the goals and objectives of the syllabus, using methods that measure what needs to be measured, with assessment types being explained, and with no surprises. A percentage of 66.7% believe that the structure of the test is challenging enough to match students' performance and therefore motivate them. More than half of the respondents allege that regarding test specifications (e.g., time allocated to each skill) (53.8%) were considered and advice was offered by a colleague on improvement of the test (57%—no, 19.4%—to some extent). Thus, piloting a test or sharing it with a colleague would probably increase the validity of a test. This is an action followed in the item bank tasks (Government Gazette 111 2020), which are reviewed by two assessors before they are uploaded onto the platform (http://www.iep.edu.gr/el/anazitisi-thematon, accessed on 22 October 2021).

Reliability (Table 4) seems to be taken into account, since all of the participants make sure that every student has a cleanly photocopied test sheet (100%), sound amplification is clearly audible to everyone in the room (95.7%), and objective scoring procedures are used (90.3%). However, writing and speaking rating scales and assessment criteria to reduce subjectivity are used by half of the respondents (51.6%), although teachers seem to be quite well trained and competent (see also Vogt and Tsagari 2014) and a lot of them should be acquainted with rating scales when preparing learners for certification. Since rating scales and rubrics are also part of alternative and descriptive assessment (Griva and Kofou 2017),

they could be the focus of attention in teacher training seminars and workshops, as they are important for increasing reliability.

Table 3. Validity.

Assessment tied to curriculum	Frequency	Percent
No	1	1.1
To some extent	22	23.7
Yes	70	75.3
Total	93	100.0
Supports syllabus goals and objectives	Frequency	Percent
No	1	1.1
To some extent	21	22.6
Yes	71	76.3
Total	93	100.0
Objectives represented in the test	Frequency	Percent
No	1	1.1
To some extent	15	16.1
Yes	77	82.8
Total	93	100.0
Test checked by a colleague	Frequency	Percent
No	53	57.0
To some extent	22	23.7
Yes	18	19.4
Total	93	100.0
Test specifications given	Frequency	Percent
No	18	19.4
To some extent	25	26.9
Yes	50	53.8
Total	93	100.0
Explanations provided	Frequency	Percent
No	6	6.5
To some extent	15	16.1
Yes	72	77.4
Total	93	100.0
Clear directions	Frequency	Percent
No	1	1.1
To some extent	2	2.2
Yes	90	96.8
Total	93	100.0
Appropriately pitched difficulty level	Frequency	Percent
To some extent	18	19.4
Yes	75	80.6
Total	93	100.0
No surprises	Frequency	Percent
No	4	4.3
To some extent	18	19.4
Yes	71	76.3
Total	93	100.0
Prompt notification	Frequency	Percent
No	1	1.1
To some extent	4	4.3
Yes	88	94.6
Total	93	100.0
Appropriate timing	Frequency	Percent
To some extent	7	7.5
Yes	86	92.5
Total	93	100.0

Table 3. *Cont.*

Appropriate review and preparation	Frequency	Percent
No	1	1.1
To some extent	7	7.5
Yes	85	91.4
Total	93	100.0
Suggestion of strategies	Frequency	Percent
No	2	2.2
To some extent	13	14.0
Yes	78	83.9
Total	93	100.0
Challenging structure	Frequency	Percent
No	2	2.2
To some extent	29	31.2
Yes	62	66.7
Total	93	100.0

Table 4. Reliability.

Cleanly photocopied test	Frequency	Percent
Yes	93	100.0
Sound amplification audible	Frequency	Percent
No	1	1.1
To some extent	3	3.2
Yes	89	95.7
Total	93	100.0
Objective scoring	Frequency	Percent
To some extent	9	9.7
Yes	84	90.3
Total	93	100.0
Writing and speaking rating scales	Frequency	Percent
No	9	9.7
To some extent	36	38.7
Yes	48	51.6
Total	93	100.0

Practicality in testing (Table 5) also seems to be taken into consideration by the majority of the participating teachers. Specifically, more than 90% assert that they know how the test will be marked, that the materials and equipment are ready in advance, that the cost of the test is within budgeted limits, and that students can complete the test within the set time frame. Furthermore, more than three-quarters think that they know how the results will be reported (87.1%), that students are informed about task marking (73.1%) and assessment criteria (76.3%), that the scoring/evaluation system of the test is feasible in the teacher's time frame (84.9%), and that any administrative details are established before the test (74.2%).

Regarding the washback effect (Table 6), 86% of the participants believe that the test tasks are related to teaching and learning, but that the results are not very promising for the washforward effect. In particular, 65.6% of the teachers ask their students to use the test results as a guide for setting goals for their future effort, half of them (49.5%) assume that learners acquire strategies and necessary life-long learning skills in tasks that emphasize communication, and only one-third believe that the test is forward looking (33.3%) and satisfies the learner's communicative needs with tasks that have real-world applications that might be encountered in real life. That means that the test is basically assigned for allocating a grade and not for preparing students for real-life problem-solving settings, depriving them of skills that are highly requested in the 21st century.

Table 5. Practicality.

Establishment of administrative details	Frequency	Percent
No	10	10.8
To some extent	14	15.1
Yes	69	74.2
Total	93	100.0
Reasonable time frame	**Frequency**	**Percent**
To some extent	7	7.5
Yes	86	92.5
Total	93	100.0
Care of materials and equipment	**Frequency**	**Percent**
To some extent	4	4.3
Yes	88	94.6
Total	93	100.0
Budgeted limits	**Frequency**	**Percent**
No	1	1.1
To some extent	5	5.4
Yes	87	93.5
Total	93	100.0
Feasible scoring/evaluation	**Frequency**	**Percent**
To some extent	14	15.1
Yes	79	84.9
Total	93	100.0
Awareness of marking	**Frequency**	**Percent**
No	1	1.1
To some extent	4	4.3
Yes	88	94.6
Total	93	100.0
Report of results	**Frequency**	**Percent**
To some extent	12	12.9
Yes	81	87.1
Total	93	100.0
Students' awareness of marking	**Frequency**	**Percent**
No	5	5.4
To some extent	20	21.5
Yes	68	73.1
Total	93	100.0
Students' awareness of evaluation criteria	**Frequency**	**Percent**
No	5	5.4
To some extent	17	18.3
Yes	71	76.3
Total	93	100.0

Table 6. Washback.

Test results as a guide to setting goals		Frequency	Percent
No	6	6.5	6.5
To some extent	26	28.0	28.0
Yes	61	65.6	65.6
Total	93	100.0	100.0
Forward-looking test		**Frequency**	**Percent**
No	11	11.8	11.8
To some extent	51	54.8	54.8
Yes	31	33.3	33.3
Total	93	100.0	100.0
Test tasks related to teaching and learning		**Frequency**	**Percent**
To some extent	13	14.0	14.0
Yes	80	86.0	86.0
Total	93	100.0	100.0

Table 6. *Cont*.

Acquisition of strategies and skills		Frequency	Percent
No	2	2.2	2.2
To some extent	45	48.4	48.4
Yes	46	49.5	49.5
Total	93	100.0	100.0

As far as feedback is concerned (Table 7), nine out of ten participants (90.3%) encourage students to improve their learning processes, give them guidance and assistance in their learning, and discuss the assignments with them in order to help them understand the content better (87.1%). About seven to eight out of ten teachers inform students about their strong points concerning learning (80.6%) and discuss with them how to utilize their strengths to move forward (73.1%), discuss the progress that they have made (81.7%), and comment on the students' test performances (69.9%). They also make a list of the weak points (71%) and discuss them in a class conference, and, after the assessment, they inform students on their weak points concerning learning and consider ways on how to improve them together (73.1%). However, only six out of ten teachers give students a chance to report on their own feedback and seek clarification of any issues that are fuzzy (68.8%) and to set new and appropriate goals for themselves in the future, and less than half of them give more than a number, grade, or phrase as feedback when returning students' tests (49.5%) or discuss the answers given with each student (45.2%). Teachers' competence in providing feedback seems rather high, at least on a class level, although in Vogt and Tsagari's (2014) study, training in this field does not appear to be as high and there is an expressed demand for more training.

Table 7. Feedback.

Information about students' strong points	Frequency	Percent
To some extent	18	19.4
Yes	75	80.6
Total	93	100.0
List of weak points	Frequency	Percent
No	7	7.5
To some extent	20	21.5
Yes	66	71.0
Total	93	100.0
Information about improving the weak points	Frequency	Percent
To some extent	25	26.9
Yes	68	73.1
Total	93	100.0
Encouraging the learning process	Frequency	Percent
To some extent	9	9.7
Yes	84	90.3
Total	93	100.0
Guiding and assisting learning	Frequency	Percent
To some extent	9	9.7
Yes	84	90.3
Total	93	100.0
Discussing assignments	Frequency	Percent
No	1	1.1
To some extent	11	11.8
Yes	81	87.1
Total	93	100.0
Discussing progress	Frequency	Percent
No	1	1.1
To some extent	16	17.2
Yes	76	81.7
Total	93	100.0

Table 7. *Cont.*

Discussing answers with each student	Frequency	Percent
No	13	14.0
To some extent	38	40.9
Yes	42	45.2
Total	93	100.0
Comments on test performance	Frequency	Percent
No	1	1.1
To some extent	27	29.0
Yes	65	69.9
Total	93	100.0
Providing elaborate feedback	Frequency	Percent
No	12	12.9
To some extent	35	37.6
Yes	46	49.5
Total	93	100.0
Clarifying fuzzy issues	Frequency	Percent
No	8	8.6
To some extent	21	22.6
Yes	64	68.8
Total	93	100.0

The last principle of testing considered is reflection (Table 8), which is rather under-valued. About 65% of the participants ensure that students know what they can learn from their assessment (67.7%) and encourage students to reflect on their learning processes (67.7%) and how they can improve their performance, or ask students to indicate what went well and what went badly concerning their assessment (63.4%). Only 41.9% ask students how they think they are doing while working on their assignments. The above data make one consider that the implementation of alternative assessment practices can develop learners' ability to reflect, especially by using a diary (Kofou 2017, p. 357).

Table 8. Reflection.

Monitoring students	Frequency	Percent
No	19	20.4
To some extent	35	37.6
Yes	39	41.9
Total	93	100.0
Encouraging reflection	Frequency	Percent
No	4	4.3
To some extent	26	28.0
Yes	63	67.7
Total	93	100.0
Indicating assessment	Frequency	Percent
No	7	7.5
To some extent	27	29.0
Yes	59	63.4
Total	93	100.0
Checking learning	Frequency	Percent
No	2	2.2
To some extent	28	30.1
Yes	63	67.7
Total	93	100.0

3.2. Assessing Skills

The second part of the questionnaire concerned the assessment of skills. Regarding reading (Table 9), it seems that the focus is on examining the students' reading skills, integrated with grammar and vocabulary use (71%), integrating higher- and lower-order skills, taking into account the interactive nature of reading (62%), and incorporating

objective-integrative techniques that pay attention to processes of reading, such as inference, completion, and construction (61.3%). Only half of the respondents affirm that they use third-generation communicative tasks, such as information transfer, multiple matching, and modified cloze tasks (55.9%), or simulations, real-life-based activities, and problem-solving activities (50.5%), which test global comprehension and are authentic in purpose, or learner-friendly and learner-centred activities (50.5%) that involve the reader in a re-encoding process. About four out of ten are positive to text authenticity (44.1%), i.e., real-life text, written for a real-world purpose, with the source and text-type identified to the reader, or tasks that focus on the process and the 'how', rather than a final product and the 'what' of the tasks (49.5%). Thus, in order to promote authenticity in testing and prepare learners for real-life tasks, teachers have to be trained, obtain support through communities of practice, and share materials that could be used for testing practices.

Table 9. Reading.

Identified text authenticity	Frequency	Percent
No	4	4.3
To some extent	48	51.6
Yes	41	44.1
Total	93	100.0
Task authenticity	Frequency	Percent
No	5	5.4
To some extent	42	45.2
Yes	46	49.5
Total	93	100.0
Simulations, real-life and problem-solving activities	Frequency	Percent
No	9	9.7
To some extent	37	39.8
Yes	47	50.5
Total	93	100.0
Process-focused tasks	Frequency	Percent
No	9	9.7
To some extent	47	50.5
Yes	37	39.8
Total	93	100.0
Testing global comprehension	Frequency	Percent
No	6	6.5
To some extent	39	41.9
Yes	48	51.6
Total	93	100.0
Integrating higher and lower order skills	Frequency	Percent
No	2	2.2
To some extent	33	35.5
Yes	58	62.4
Total	93	100.0
Integrating reading with grammar and vocabulary	Frequency	Percent
No	1	1.1
To some extent	26	28.0
Yes	66	71.0
Total	93	100.0
Communicative activities	Frequency	Percent
No	3	3.2
To some extent	38	40.9
Yes	52	55.9
Total	93	100.0

Table 9. *Cont*.

Learner-centred activities	Frequency	Percent
No	6	6.5
To some extent	40	43.0
Yes	47	50.5
Total	93	100.0
Incorporating objective integrative techniques	Frequency	Percent
No	5	5.4
To some extent	31	33.3
Yes	57	61.3
Total	93	100.0

Regarding authenticity in listening testing items (Table 10), half of the respondents use simulations, real-life-based activities, and problem-solving activities (55.9%), as well as top-down and bottom-up processing (52.7%) in order for the students to understand both the overall and specific meanings, or use an authentic source (48.4%) to avoid unauthentic, contrived language. Just over half include productive tasks, with students tested objectively using reliable testing techniques (51.6%), as they focus on the process and the 'how' (47.3%), rather than on the final product and the 'what' of the tasks. Third-generation testing tasks, such as information transfer, multiple matching, and modified listening cloze tasks, which are learner-friendly and learner-centred and involve the reader in a re-encoding process, are used by 45.2%, while hard-focus, extended listening activities, which require selective listening to gather specific information and listen with a purpose in mind, are included by 30.1% only, and tasks which entail interpretation rather than asking students to identify points and extract specific information are used by 20.4%. Thus, authenticity in testing listening is included only to some extent and needs to be enhanced for learners to be able to cope with a variety of authentic communicative situations and feel more confident when encountering different types of listening tasks.

Table 10. Listening.

Text authenticity	Frequency	Percent
No	10	10.8
To some extent	38	40.9
Yes	45	48.4
Total	93	100.0
Task authenticity	Frequency	Percent
No	7	7.5
To some extent	39	41.9
Yes	47	50.5
Total	93	100.0
Simulations and real-life tasks	Frequency	Percent
No	8	8.6
To some extent	33	35.5
Yes	52	55.9
Total	93	100.0
Process-focused tasks	Frequency	Percent
No	11	11.8
To some extent	38	40.9
Yes	44	47.3
Total	93	100.0
Third-generation communicative tasks	Frequency	Percent
No	7	7.5
To some extent	44	47.3
Yes	42	45.2
Total	93	100.0

Table 10. *Cont.*

Top-down and bottom-up processes	Frequency	Percent
No	2	2.2
To some extent	42	45.2
Yes	49	52.7
Total	93	100.0
Tasks entail interpretation	Frequency	Percent
No	11	11.8
To some extent	63	67.7
Yes	19	20.4
Total	93	100.0
Hard-focus, extended listening activities	Frequency	Percent
No	11	11.8
To some extent	54	58.1
Yes	28	30.1
Total	93	100.0
Productive tasks tested objectively	Frequency	Percent
No	7	7.5
To some extent	38	40.9
Yes	48	51.6
Total	93	100.0

On the contrary, writing (Table 11) is tested in a more authentic way, since the writer is aware of the writing purpose, the register to be used, and the audience that they are addressing (90.3%). About 65% to 75% of the participants believe that when assessing writing, the context is pre-defined (73.1%), the task is communicative, involving the learner in meaningful, forward-looking communicative situations (75.3%), and that the learner needs to exhibit useful language skills that may be needed in a real-world context (69.9%). Guidance is also provided (65.6%), with notes given to the learner in order to guide the content and the lexical elements of the language, and a full task environment is given (66.7%). About 60% of teachers stated that the writer is involved in a purposeful situation (60.2%) because they adopt a realistic role with a real-life outcome and a realistic output text (giving the product authenticity), and that, in this way, the tasks are authentic in purpose. In contrast, a lower percentage of teachers used global and analytic rating scales that avoid impressionistically scoring essays (impressionistic scoring has low reliability), reduce subjectivity, and increase reliability (47.3%). A similar number used tasks that focus on the process and the 'how', rather than on the final product and the 'what' of the tasks (45.2%), while fewer ensured text authenticity with the use of real-world sources and genuine input texts, not written for language teaching (41.9%). Fewer still included mediation, relaying information from an authentic Greek text into English (33.3%). It seems that writing is tested in a more authentic way than other skills, but the need for training teachers on how to use rating scales also emerges.

Finally, speaking (Table 12) seems to be tested in a partially authentic way. About six to seven out of ten teachers acknowledge that the learner is involved in a purposeful conversation within a given context (73.1%). Interactive, guided tasks that give weight to communication are used by 73.1%, and the stated aim of the tasks is to elicit authentic language (73.1%) that can be used in non-test situations and real-world tasks. Moreover, during the task, the speakers exchange information and communicate ideas for normal purposes, using spontaneous and unplanned language to negotiate meaning (68.8%), and simulations, real-life-based activities, problem-solving activities, as well as information-gap techniques, are included (68.8%), in which the interactive nature and unpredictability of the spoken language is ensured (59.1%). About half of them believe that the tasks they use focus on the process and the 'how', rather than on the final product and the 'what' (52.7%). Real-world sources and authentic visual input are used by 50.5%, guidance is provided in the form of given notes by 49.5%, and global and analytic performance scales are employed to reduce subjectivity and increase reliability by 53.8%. With regard to text authenticity, only 43% say that the input and prompts are not simplified so that tasks are authentic in purpose and context.

Table 11. Writing.

Task authenticity	Frequency	Percent
No	4	4.3
To some extent	19	20.4
Yes	70	75.3
Total	93	100.0
Guidance provided	Frequency	Percent
No	3	3.2
To some extent	29	31.2
Yes	61	65.6
Total	93	100.0
Full task environment given	Frequency	Percent
No	4	4.3
To some extent	27	29.0
Yes	62	66.7
Total	93	100.0
Pre-defined context	Frequency	Percent
No	4	4.3
To some extent	21	22.6
Yes	68	73.1
Total	93	100.0
Text authenticity ensured	Frequency	Percent
No	11	11.8
To some extent	43	46.2
Yes	39	41.9
Total	93	100.0
Communicative tasks	Frequency	Percent
No	2	2.2
To some extent	26	28.0
Yes	65	69.9
Total	93	100.0
Tasks authentic in purpose	Frequency	Percent
No	6	6.5
To some extent	34	36.6
Yes	53	57.0
Total	93	100.0
Process-focused tasks	Frequency	Percent
No	7	7.5
To some extent	44	47.3
Yes	42	45.2
Total	93	100.0
Purposeful situations	Frequency	Percent
No	3	3.2
To some extent	34	36.6
Yes	56	60.2
Total	93	100.0
Awareness of the writing purpose	Frequency	Percent
No	1	1.1
To some extent	8	8.6
Yes	84	90.3
Total	93	100.0
Mediation included	Frequency	Percent
No	23	24.7
To some extent	39	41.9
Yes	31	33.3
Total	93	100.0
Global and analytic rating scales	Frequency	Percent
No	8	8.6
To some extent	41	44.1
Yes	44	47.3
Total	93	100.0

Table 12. Speaking.

Communicative authenticity and interactive guided tasks	Frequency	Percent
No	3	3.2
To some extent	22	23.7
Yes	68	73.1
Total	93	100.0
Guidance provided	Frequency	Percent
No	6	6.5
To some extent	41	44.1
Yes	46	49.5
Total	93	100.0
Texts authentic in purpose and context	Frequency	Percent
No	5	5.4
To some extent	48	51.6
Yes	40	43.0
Total	93	100.0
Simulations, real-life problem-solving activities	Frequency	Percent
No	2	2.2
To some extent	27	29.0
Yes	64	68.8
Total	93	100.0
Process authenticity	Frequency	Percent
No	4	4.3
To some extent	40	43.0
Yes	49	52.7
Total	93	100.0
Purposeful conversation	Frequency	Percent
No	2	2.2
To some extent	23	24.7
Yes	68	73.1
Total	93	100.0
Eliciting authentic language	Frequency	Percent
No	2	2.2
To some extent	23	24.7
Yes	68	73.1
Total	93	100.0
Exchange of information and communication	Frequency	Percent
No	2	2.2
To some extent	27	29.0
Yes	64	68.8
Total	93	100.0
Interactive nature and unpredictability	Frequency	Percent
No	1	1.1
To some extent	37	39.8
Yes	55	59.1
Total	93	100.0
Spontaneous and unplanned language	Frequency	Percent
No	3	3.2
To some extent	28	30.1
Yes	62	66.7
Total	93	100.0
Authenticity ensured with real-world sources	Frequency	Percent
No	4	4.3
To some extent	42	45.2
Yes	47	50.5
Total	93	100.0
Global and analytic performance scales	Frequency	Percent
No	10	10.8
To some extent	33	35.5
Yes	50	53.8
Total	93	100.0

3.3. Categorization of Teachers

The application of the toolkit (see Appendix A) gave the participating teachers the opportunity to test how authentic an assessor they are. In particular, they evaluated themselves and responded on the basis of a three-point scale (2 points for each YES, 1 point for each TO SOME EXTENT, 0 points for each NO), and then converted their score to a

percentage (Part I: score/130 × 100 = ... %; Part II score/86 × 100 = ... %). The categorization more or less followed the categorization patterns of EFL exam certification, taking into account that 60% is the passing score, and a score of 90% is a distinction.

The three different profiles of assessors that were created were based on the assessment activities and methods preferred by the teachers. The teachers' assessment processes, actions, and activities used determine the kind of assessors that they are. In other words, the characteristics they share and the differences they have help establish the different profiles. Added to this, their beliefs about assessment and how they relate to the practical side of assessing is what led the researchers to distinguish and establish the three different models of assessors, ranging from alternative to non-enthusiastic.

In Part I, those that score 91% and above are considered to be Alternative Assessors (TEACHER A). Most teachers of this profile use different types of assessment. Teachers who belong to this category devise their own tasks and tests to help students develop their independence. They use assessment as a way for learning and as guidance for the next steps. They may also feel it is very important to use alternative, authentic, and third-generation assessment as much as possible by adhering to the key principles of assessment. They probably prefer the sort of language assessment that offers a more realistic and communicative view of the language, for example replicating real-world processes using genuine unaltered materials. This is often the sort of language assessment that they do in class and may be able to increase its benefits with continuous reflection and by promoting active student involvement.

Those that score 61–90% are considered to be Mainstream Assessors (TEACHER B). Teachers who score close to average belong to this profile. They may find that they do not fall exactly into either of the other categories (alternative and non-enthusiastic), which makes them a mixture. They tend to combine different ways at different times depending on the situation and what they are doing. As a result, they use both test-based and alternative assessments for different purposes. They may sometimes feel, however, that they should be using more authentic ways of assessment, but there are probably reasons that keep them back. Mainstream Assessors need to find more time to learn and be more self-critical. If they become more aware of the reasons why they avoid using alternative, authentic, and third-generation assessment more often, they may find it easier to do something about them. Following the principles of testing will also help them increase the benefits of their assessment.

Finally, those that score up to 60% are Non-enthusiastic Assessors (TEACHER C). Teachers in this category use assessment in a less diverse way. They mainly use the information that they get from assessment to grade students' performance using more traditional methods. This can be identified as collecting information and providing feedback through grading. Assessment for Non-enthusiastic Assessors relies a lot on reproduction and memorization by focusing on form and accuracy. Perhaps they should be less preoccupied with assessing student knowledge, and focus more on learning as well. This score does not mean that they are not good language assessors, but there is always room for improvement. Perhaps this is the first time that they have thought about the way they assess learners. A good starting point would be to adopt the principles of language assessment and offer a choice of assessment methods. Knowing more about this and receiving adequate training can be very useful in helping them become more effective language assessors.

In Part II, those that score 71% and above are considered to be Alternative Assessors (TEACHER A). Teachers of this profile try to involve learners in communicative, real-world, guided tasks or simulated authentic situations that reflect genuine communication. Teachers who belong to this type of assessor take into account the interactive nature of learning by combining authentic, third-generation techniques that focus both on the product and process, which aim to help learners exhibit language skills that may be useful in a real-world context. As a result, they seek to involve learners in meaningful, forward-looking communicative situations that extend to real-life language use.

Those that score 41–70% are considered to be Mainstream Assessors (TEACHER B). Teachers in this category move beyond traditionally constructed tests by trying to adopt a communicative approach that extends to real-life language use with the use of a combination of third-generation testing techniques. Mainstream assessors may find that they use a mixture of different testing techniques, combining all three generations of testing. They might feel, however, that they should be using more authentic, third-generation tasks to assess the students' skills. Taking into account the criteria for assessing the receptive and productive skills will make it easier for them to create authentic tasks to trigger motivation and increase the students' involvement. The more the communicative criteria are met, the more the learners' needs are satisfied.

Finally, those that score up to 40% are Non-enthusiastic Assessors (TEACHER C). Teachers in this category tend to test skills objectively using non-authentic, disembodied techniques that are not related to the students' real-world needs. The teacher who belongs to this category should try modifying tasks to achieve authenticity. The first step is to embody third-generation tasks and real-world activities, using authentic, unaltered materials. Similarly, the tasks could be significantly improved by providing an authentic context to engage students in meaningful activities that are actually needed in the real world. As a result, priority will be given to the issue of authenticity with the aim of boosting the students' communicative competency and satisfying their needs.

4. Discussion

The purpose of this research was two-fold. Our aims were, firstly, to investigate Greek EFL teachers' awareness of communicative testing techniques and alternative assessment methods and principles, and, secondly, to promote assessment literacy based on the use of the aforementioned assessment tool.

As for the type of assessors the research indicated that the majority of teachers belong to Mainstream Assessors, since alternative methods of assessment fail to be employed on a regular basis and third-generation assessment techniques are rarely used owing to lack of familiarity. Teachers seem to be positively disposed towards the benefits of alternative assessment, but at the same time are indecisive due to the practicality that traditional assessment offers (Zarali and Kofou 2020, p. 187). Given this, the need to achieve solid assessment literacy is great. Assessment literacy and training needs to be promoted to help teachers modify their assessment practices for the benefit of teaching and learning. This highlights the importance of continuing professional development and training in assessment practices, which is also evident in Vogt and Tsagari's (2014) study across Europe, in which the participating teachers "seemed to perceive a need for training across the whole spectrum" (p. 376) of language assessment literacy. Promoting changes in educational policies to help teachers employ new methods can prompt them to experiment and increase the use of alternative sources to keep up with the changes in the field of assessment.

All in all, this particular study offered the chance to help teachers understand the criteria that should be assigned to testing to become better prepared for assessing learners and to pull away from traditional testing with written tests that include second-generation activities, such as multiple-choice and true/false questions and summative assessment, usually conducted at the end of each unit or at the end of the semester. The findings of this study may have various implications for the development of assessment literacy, as they may help teachers, policymakers, stakeholders, and researchers proceed with their work in the field. More specifically, as Giraldo (2018) asserts (and which the present study reinforces), language teachers need to have the necessary knowledge (theoretical considerations regarding validity and reliability, for example), instructional skills that they can apply to assessment practices, and the ability to design testing items for the four language skills and apply language assessment principles to be considered assessment literate. In this context, the interpretation of the data implies that appropriate pre-service and in-service training should take place to help teachers employ more effective means of assessment that can lead to student empowerment. As a result, teachers should reflect

more on the concept of third-generation communicative testing along with alternative methods to achieve an effective, more communicative, lifelike assessment of the students' performance.

Thus, the main contribution of the toolkit presented lies in the fact that it supports and guides teachers on how to use assessments in order to have a positive effect on student learning. It also points out the inextricable relationship between language assessment and teaching, as well as the importance of reflection on the practices and methods used by the teachers. This toolkit can remain an open-access tool that will be available for reflective practice or used as a guide to draw teachers' attention to ways that they could improve their assessment practices. Ideally, it can contribute to the shaping of language assessment literacy in general. To that end, together with the results of the study, it can be the basis of a language assessment training programme.

5. Conclusions

In summary, assessment literacy and training can encourage teachers to modify their assessment practices for the benefit of teaching and learning. This also highlights the importance of giving professional development opportunities to teachers. Clearly, promoting changes in educational policies to help teachers employ new methods will give 'traditional' teachers the opportunity to incorporate new methods when assessing. As Scarino (2013) claims, assessment-literate teachers can explore and evaluate their preconceptions and become aware of their own framework of knowledge and practices. The ideal result would be to motivate EFL teachers to redefine their roles and equip them with techniques that enrich their practice of assessment. To that end, the Tale project (Tsagari et al. 2018) can further help them reflect on their assessment practices and raise their awareness and levels of LAL through an online, self-study training course. Shaping teacher assessment literacy can improve and enhance the quality of language education by preparing autonomous and independent learners, as well as competent teachers who can understand the nature of assessment.

Author Contributions: Conceptualization, E.G. and I.K.; methodology, E.G. and I.K.; software, E.G. and I.K.; validation, E.G. and I.K.; formal analysis, E.G. and I.K.; investigation, E.G. and I.K.; resources, E.G. and I.K.; data curation, E.G. and I.K.; writing—original draft preparation, E.G. and I.K.; writing—review and editing, E.G. and I.K.; visualization, E.G. and I.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available. (https://docs.google.com/spreadsheets/d/1Ba0WF_-azuZMNDhNnHUuBbCTt8pQyZ6BRGrIlHVUOBc/edit?usp=sharing).

Appendix A

LANGUAGE ASSESSMENT FOR TEACHERS

TEACHER SELF-EVALUATION CHECKLIST ON ASSESSMENT
A TOOLKIT FOR TEACHER DEVELOPMENT

This toolkit is based on the results of a research project which investigated the beliefs, perceptions, and practices of EFL teachers regarding assessment. It aims to help teachers design, develop, and critically evaluate tests as well as help them reflect on their assessment techniques. It can be used by pre- and in-service EFL teachers to raise their awareness of language assessment and support them to make informed decisions when assessing students' learning.

The toolkit is divided into two parts. The first part outlines the principles of language assessment, while the second part deals with the assessment of the four skills. Teachers are provided with criteria in the form of checklists with which they evaluate and rate their practices. It can also be regarded as a guide and self-evaluation tool for teachers that can be used several times during the school year to observe and reflect on their development and growth. Thus, teachers will be equipped with useful insights into test evaluation and design so as to be able to design their own tests for assessing learners.

PART I

The first part aims to provide both novices and experienced teachers with input that concerns the basic assessment principles to help them understand how tests are constructed. Teachers can review or familiarize themselves with fundamental principles of language testing for describing, categorizing, and evaluating published tests, as well as designing their own tests.

Teacher Self-Evaluation Checklist on Assessment

Directions:
Indicate the degree to which you incorporate each criterion below into your current practices.

2—YES, 1—TO SOME EXTENT, 0—NO

PRINCIPLES OF LANGUAGE ASSESSMENT			
	CRITERIA		**2 1 0**
Authenticity	1.	Tasks emphasize the communicative view of the language.	
	2.	The language in the test is as natural as possible.	
	3.	The test items are as contextualized as possible rather than isolated.	
	4.	The rubrics used for the exercises have been contextualized to offer a more realistic and communicative view of the language.	
	5.	Genuine unaltered materials that can be actually found in the real world, not written for a language teaching purpose, have been used.	
	6.	The test topics and situations are interesting, enjoyable, and humorous for the learners, as well as relevant to their age and level.	
	7.	Some thematic organization is provided in the test, such as through a storyline or episode.	
	8.	Assessment tasks represent, or closely approximate, real-world processes with effective tasks which are relevant to real-life situations.	
	9.	The assessment tests the student's ability to apply knowledge to real-life problems.	
	10.	The assessment tasks replicate real-world situations which are useful for everyday life.	
Testing	1.	The testing content is similar to the teaching one.	
	2.	All skills are integrated.	
	3.	Promotes autonomous and self-directed learning.	
	4.	Promotes learner-centred assessment with a clear purpose.	
	5.	Triggers student's motivation, engages and involves him/her in the activities	
	6.	Students can express their opinion on how they will be assessed.	
	7.	Assessment is regarded as a shared responsibility.	
	8.	Students are given a choice of assessment tasks.	
	9.	Students are given assessment tasks that suit their abilities.	

		PRINCIPLES OF LANGUAGE ASSESSMENT			
		CRITERIA	**2**	**1**	**0**
Validity	1.	The assessment is tied to curricular practices.			
	2.	The assessment supports the goals and objectives of the syllabus followed, using methods that measure what needs to be measured.			
	3.	The classroom/lesson objectives are identified in the assessment practices and are appropriately represented in the test.			
	4.	The test is given to a colleague to be checked and examined for its clarity to detect any possible problems and offer suggestions for improvement before being administered.			
	5.	Test specifications are given. For example, if you have to test all four of the students' skills in one teaching period, you specify how many minutes your students should spend for each skill.			
	6.	I have explained to students how each type of assessment is to be used.			
	7.	The directions of the designed test are clear.			
	8.	The test's difficulty level is appropriately pitched.			
	9.	The test has no "surprises".			
	10.	Students are told in advance when and on what they are being assessed.			
	11.	The timing of the test is appropriate.			
	12.	I offer students appropriate review and preparation for the test.			
	13.	I suggest strategies that will be beneficial for students before the test.			
	14.	I structure the test so that the best students will be modestly challenged and the weaker students will not be overwhelmed.			
Reliability	1.	I make sure that every student has a cleanly photocopied test sheet.			
	2.	I make sure that sound amplification is clearly audible to everyone in the room.			
	3.	I use objective scoring procedures that leave little debate about the correctness of an answer.			
	4.	I use writing and speaking rating scales and assessment criteria to reduce subjectivity.			
Practicality	1.	I establish administrative details clearly before the test, such as the rubric of scoring and specification of the test.			
	2.	Students can complete the test I give, reasonably within the set time frame.			
	3.	All materials and equipment of the test are ready, for example, listening tape for listening section and answer sheets for students.			
	4.	The cost of the test is within budgeted limits.			
	5.	The scoring/evaluation system of the test is feasible in the teacher's time frame.			
	6.	I am aware of how my assessment will be marked.			
	7.	I know in advance how the results will be reported (i.e., feedback given)			
	8.	Students know how a particular assessment task will be marked.			
	9.	My students know what the evaluation criteria for their assessments are.			
Washback and Washforward Effect	1.	I ask students to use the test results as a guide to setting goals for their future effort.			
	2.	The test is forward-looking and satisfies the learner's communicative needs with tasks that have real-world applications and which might be encountered in real life.			
	3.	The test tasks are related to teaching and learning.			
	4.	Learners acquire strategies and necessary life-long learning skills in tasks that emphasize communication.			

		PRINCIPLES OF LANGUAGE ASSESSMENT	
		CRITERIA	**2 1 0**
Feedback	1.	I inform my students about their strong points concerning learning and discuss with them how to utilize their strengths to move forward.	
	2.	I make a list of the weak points and discuss them in a class conference.	
	3.	After the assessment, I inform my students on their weak points concerning learning and consider ways on how to improve them together with my students.	
	4.	I encourage my students to improve on their learning processes.	
	5.	I give students guidance and assistance in their learning.	
	6.	I discuss assignments with my students to help them understand the content better.	
	7.	I discuss with my students the progress they have made.	
	8.	After a test, I discuss the answers given with each student.	
	9.	I give a comment generously and on the students' test performance.	
	10.	I give more than a number, grade, or phrase as my feedback when returning students' tests.	
	11.	I give students a chance to feedback on my feedback to seek clarification of any issues that are fuzzy and to set new and appropriate goals for themselves in the future.	
Reflection	1.	While working on their assignments, I ask my students how they think they are doing.	
	2.	I encourage my students to reflect on their learning processes and how they can improve their performance.	
	3.	I ask my students to indicate what went well and what went badly concerning their assessment.	
	4.	I ensure that my students know what they can learn from their assessment.	

How to calculate your score:
Score: 2 points for each YES
1 point for each TO SOME EXTENT
0 points for each NO
TOTAL SCORE:
Convert your score to a percentage: /130 = ... %

RESULTS
SCORE: 91% and above
TEACHER A—Alternative Assessors
Most teachers in this profile use different types of assessment. If you belong to this category, you devise your own tasks and tests to help students develop their independence. You use assessing as a way for learning and as guidance for the next steps. You may also feel that it is very important to use alternative, authentic, and third-generation assessments as much as possible by adhering to the key principles of assessment. You probably prefer the sort of language assessment where you need to offer a more realistic and communicative view of the language, for example, by replicating real-world processes using genuine unaltered materials. This is often the sort of language assessment you do in class, and you may be able to increase its benefits with continuous reflection and by promoting active student involvement.

SCORE 61–90%
TEACHER B—Mainstream Assessors
Teachers who score close to average belong to this profile. You may find that you do not fall exactly into either of the alternative and non-enthusiastic categories, which

makes you a mixture as you combine different ways at different times depending on the situation and what you are doing. As a result, you use both test-based and alternative assessments for different purposes. You may sometimes feel, however, that you should be using more authentic ways of assessment but there are probably reasons that keep you back. Try finding more time to learn and be more self-critical. If you become more aware of the reasons you avoid using alternative, authentic, and third-generation assessments more often, you may find it easier to do something about them. Following the principles of testing will also help you increase the benefits of your assessment.

SCORE Up to 60%

TEACHER C—Non-enthusiastic Assessors

Teachers in this category use assessment in a less diverse way. You may find that you mainly use the information you get from assessment to grade your students' performance using more traditional methods. This can be identified as collecting information and providing feedback through grading. Assessment for you relies a lot on reproduction and memorization by focusing on form and accuracy. Perhaps you should be less preoccupied with assessing only student knowledge, and focus more on learning as well. This score does not mean that you are not a good language assessor, but there is always room for improvement. Perhaps this is the first time that you have thought about the way you assess learners. Try to offer a choice of assessment methods. A good starting point would be to adopt the principles of language assessment. Knowing more about this and receiving adequate training can be very useful in helping you to become a more effective language assessor.

PART II

The second part deals with the assessment of the four skills. The following checklist will show teachers what to take into account when assessing the students' receptive and productive skills. It is divided into four sections, one for each skill, and includes specific criteria that can help evaluate tests and also help teachers design their own assessment tasks. Teachers can take these criteria into consideration when exploring the testing of all four skills.

Teacher Self-Evaluation Checklist on Assessing Skills

Directions:

Indicate the degree to which you incorporate each criterion below into your current practices.

2—YES, 1—TO SOME EXTENT, 0—NO

ASSESSING SKILLS			
		CRITERIA	2 1 0
Reading	1.	Text authenticity; real-life text, written for a real-world purpose, with the source and text-type identified to the reader.	
	2.	The tasks are authentic in purpose.	
	3.	Simulations, real-life-based activities, and problem-solving activities are included.	
	4.	Tasks focus on the process and the 'how', rather than a final product and the 'what' of the tasks.	
	5.	The tasks test global comprehension.	
	6.	The questions try to integrate higher- and lower-order skills, taking into account the interactive nature of reading.	
	7.	The focus is on examining the student's reading skills, integrated with grammar and vocabulary use.	
	8.	Use of third-generation activities, such as information transfer activities, multiple matching activities, and modified cloze tasks.	
	9.	Use of learner-friendly and learner-centred activities that involve the reader in a re-encoding process.	
	10.	Incorporate objective integrative techniques that pay attention to processes of reading, such as inference, completion, and construction.	

ASSESSING SKILLS			
Listening	1.	Text authenticity; use of an authentic source to avoid unauthentic contrived language.	
	2.	The tasks are authentic in purpose.	
	3.	Simulations, real-life-based activities, and problem-solving activities are included.	
	4.	Tasks focus on the process and the 'how', rather than a final product and the 'what' of the tasks.	
	5.	Use of third-generation activities, such as information transfer activities, multiple matching activities, and modified listening cloze tasks, which are learner-friendly and learner-centred and involve the reader in a re-encoding process.	
	6.	Top-down and bottom-up processing is involved to understand both the overall and specific meanings.	
	7.	Tasks entail interpretation rather than identify points and extract specific information.	
	8.	Hard-focus, extended listening activities are included, which require selective listening to gather specific information and listen with a purpose in mind.	
	9.	Productive tasks can be included but should be tested objectively with reliable testing techniques. For example, completing information only in short answers.	
Writing	1.	The task is communicative, involving the learner in meaningful, forward-looking communicative situations.	
	2.	Guidance is provided; notes are given to the learner in order to guide the content and the lexical elements of the language.	
	3.	Full task environment specification is given.	
	4.	The context is pre-defined.	
	5.	Text authenticity is ensured with the use of real-world sources and genuine input texts, not written for language teaching.	
	6.	Task authenticity; the learner needs to exhibit useful language skills that may be needed in a real-world context.	
	7.	The tasks are authentic in purpose.	
	8.	Tasks focus on the process and the 'how', rather than a final product and the 'what' of the tasks.	
	9.	The writer is involved in a purposeful situation by adopting a realistic role with a real-life outcome and a realistic output text (product authenticity).	
	10.	The writer is aware of the writing purpose, the register to be used, and the audience to be addressed.	
	11.	Mediation can be included (requires the student to process and relay information from an authentic Greek text into English).	
	12.	Global and analytic rating scales are used to avoid impressionistically scored essay (low-reliability), reduce subjectivity and increase reliability.	
Speaking	1.	Communicative authenticity; use of interactive, guided tasks that give weight to communication.	
	2.	Guidance is provided in the form of given notes.	
	3.	The tasks are authentic in purpose and context, which are given to the learner without making use of simplified input and prompts.	
	4.	Simulations, real-life-based activities, problem-solving activities, and information-gap techniques are included.	
	5.	Process authenticity; tasks focus on the process and the 'how', rather than a final product and the 'what' of the tasks.	
	6.	The learner is involved in a purposeful conversation within a given context.	
	7.	The task's aim is to elicit authentic language that can be used in non-test situations and real-world tasks.	
	8.	During the task, the speakers exchange information and communicate ideas for normal purposes.	
	9.	The interactive nature and unpredictability of the spoken language is ensured.	
	10.	The speakers are able to use spontaneous and unplanned language to negotiate meaning.	
	11.	Authenticity is ensured with the use of real-world sources and authentic visual input.	
	12.	Global and analytic performance scales are used to reduce subjectivity and increase reliability.	

How to calculate your score:
Score: 2 points for each YES

1 point for each TO SOME EXTENT
0 points for each NO
TOTAL SCORE:
Convert your score to a percentage: /86 = ... %

RESULTS
SCORE 71% and above
TEACHER A—Alternative Assessors
Teachers in this profile try to involve learners in communicative, real-world, guided tasks or simulated authentic situations that reflect genuine communication. If you belong to this type of assessors, it is evident that you take into account the interactive nature of learning by combining authentic, third-generation techniques that focus both on the product and process, which aim to help learners exhibit language skills that may be useful in a real-world context. As a result, you seek to involve learners in meaningful, forward-looking communicative situations that extend to real-life language use.

SCORE 41–70%
TEACHER B—Mainstream Assessors
Teachers in this category move beyond traditionally constructed tests by trying to adopt a communicative approach that extends to real-life language use with the use of a combination of third-generation testing techniques. You may find that you are a mixture of different techniques, combining all three generations of testing. You might feel, however, that you should be using more authentic, third-generation tasks to assess the students' skills. Taking into account the criteria for assessing the receptive and productive skills will make it easier for you to create authentic tasks to trigger motivation and increase the students' involvement. The more the communicative criteria are met, the better the learners' needs are satisfied.

SCORE Up to 40%
TEACHER C—Non-enthusiastic Assessors
Teachers in this category tend to test skills objectively using non-authentic, disembodied techniques that are not related to the students' real-world needs. If you belong to this category, try modifying tasks to achieve authenticity. The first step is to embody third-generation tasks and real-world activities, using authentic unaltered materials. Similarly, the tasks could be significantly improved by providing an authentic context to engage students in meaningful activities that are actually needed in the real world. As a result, priority will be given to the issue of authenticity with the aim of boosting the students' communicative competence and satisfying their needs.

References

Alderson, J. Charles, and Jayanti Banerjee. 2001. Language testing and assessment (Part I). *Language Teaching* 34: 213–36. [CrossRef]

Bachman, F. Lyle, and Adrian S. Palmer. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Baker, A. Beverly. 2016. Language Assessment Literacy as Professional Competence: The Case of Canadian Admissions Decision Makers. *The Canadian Journal of Applied Linguistics* 19: 63–83.

Bøhn, Henrik, and Dina Tsagari. 2021. Teacher Educators' Conceptions of Language Assessment Literacy in Norway. *Journal of Language Teaching and Research* 12: 222–33. [CrossRef]

Bowers, C. Bruce. 1989. *Alternatives to Standardized Educational Assessment*. ERIC Digest Series Number EA 40 (ED312773). Washington, DC: Office of Educational Research and Improvement.

Brown, H. Douglas. 2003. *Language Assessment: Principles and Classroom Practices*. New York: Longman.

Chirimbu, Sebastian. 2013. Using Alternative Assessment Methods in Foreign Language Teaching. Case Study: Alternative Assessment of Business English for University Students. *Scientific Bulletin of the Politehnica University of Timişoara: Transactions on Modern Language* 12: 91–98.

Fulcher, Glenn. 2000. The 'communicative' legacy in language testing. *System* 28: 483–97.

Fulcher, Glenn. 2012. Assessment Literacy for the Language Classroom. *Language Assessment Quarterly* 9: 113–32.

Giraldo, Frank. 2018. Language assessment literacy: Implications for language teachers. *Teachers' Professional Development* 20: 179–95. [CrossRef]

Gkogkou, Eirini. 2019. An investigation of the Teachers' Beliefs, Perceptions and Assessment Practices in English Language Teaching. Unpublished dissertation, Hellenic Open University, Patras, Greece. Available online: https://apothesis.eap.gr/bitstream/repo/42987/1/103390_GKOGKOU_EIRINI.pdf (accessed on 22 October 2021).

Gkogkou, Eirini, and Ifigenia Kofou. 2020. An investigation of the teachers' beliefs, perceptions and assessment practices in English language teaching. Paper presented at the 24th International Symposium on Theoretical and Applied Linguistics, Thessaloniki, Greece, October 2–4.

Government Gazette 111. 2020. *N. 4692/2020 (ΦΕΚ Α 111-12.06.2020) Αναβάθμιση του Σχολείου και άλλες διατάξεις*. Ελλάδα: Υπουργείο Παιδείας.

Government Gazette 140. 2021. *Εσωτερική και εξωτερική αξιολόγηση του εκπαιδευτικού έργου των σχολείων*. Ελλάδα: Υπουργείο Παιδείας.

Grabowski, C. Kirby, and Jee Wha Dakin. 2014. Test development literacy. In *Companion to Language Assessment*. Edited by Antony. J. Kunnan. Oxford: Wiley-Blackwell, pp. 869–89.

Griva, Eleni, and Ifigenia Kofou. 2017. *Alternative Assessment in Language Learning: Challenges and Practices*. Thessaloniki: Kyriakidis Editions, vol. 1.

Harding, Luke, and Benjamin Kremmel. 2016. Teacher assessment literacy and professional development. In *Handbook of Second Language Assessment*. Edited by Dina Tsagari and Jayanti Banerjee. Berlin: De Gruyter, pp. 413–28. [CrossRef]

Hedge, Tricia. 2000. *Teaching and Learning in the Language Classroom*. Oxford: Oxford University Press.

Herrera, Leonardo, and Diego Macías. 2015. A call for language assessment literacy in the education and development of Teachers of English as a foreign language. *Colombian Applied Linguistics Journal* 17: 302–12. [CrossRef]

Inbar-Lourie, Ofra. 2008. Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing* 25: 385–402. [CrossRef]

Ioannou-Georgiou, Sophie, and Pavlos Pavlou. 2003. *Assessing Young Learners*. Oxford: Oxford University Press.

Kofou, Ifigenia. 2017. Alternative assessment modes: Reading and movie diaries. Paper presented at the 3o Διεθνές Συνέδριο για την Προώθηση της Εκπαιδευτικής Καινοτομίας, Larissa, Greece, October 13–15. pp. 355–63.

López, Alexis, and Ricardo Bernal. 2009. Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers' Professional Development* 11: 55–70.

Marshall, N. Martin. 1996. Sampling for qualitative research. *Family Practice* 13: 522–25. [CrossRef] [PubMed]

McNamara, Tim. 2000. *Language Testing*. Oxford: Oxford University Press.

McNamara, Tim, and Carsten Roever. 2006. *Language Testing: The Social Dimension*. Language Learning and Monograph Series. Malden and Oxford: Blackwell Publishing.

Nunan, David. 1990. Action Research in the Classroom. In *Second Language Teacher Education*. Edited by Jack C. Richards and David Nunan. Cambridge: Cambridge University Press, pp. 62–82.

OECD. 2021. *PISA 2025 Foreign Language Assessment Framework*. Paris: PISA, OECD Publishing.

Papageorgiou, Spiros. 2009. *Setting Performance Standards in Europe: The Judges' Contribution to Relating Language Examinations to the Common European Framework of Reference*. Frankfurt: Peter Lang.

Patton, Q. Michael. 2002. *Qualitative Research and Evaluation Methods*, 3rd ed. Thousand Oaks: Sage Publications, Inc.

Popham, W. James. 2018. *Assessment Literacy for Teachers in a Hurry*. Alexandria: Association for Supervision and Curriculum Development.

Scarino, Angela. 2013. Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing* 30: 309–27. [CrossRef]

Stiggins, J. Richard. 1991. Assessment literacy. *Phi Delta Kappan* 72: 534–39.

Swan, Karen, Jia Shen, and Starr Roxanne Hiltz. 2006. Assessment and Collaboration in Online Learning. *Journal of Asynchronous Learning* 10: 45–62. [CrossRef]

Tsagari, Constance. 2004. Alternative Methods if Assessing Achievement. In *Testing and Assessment in Language Learning: Assessing Students without Tests*. Edited by Constance Tsagari and Richard West. Patras: Hellenic Open University, vol. 3, pp. 119–342.

Tsagari, Dina. 2011. Investigating the 'assessment literacy' of EFL state school teachers in Greece. In *Classroom-Based Language Assessment*. Edited by Dina Tsagari and Ildikó Csépes. Frankfurt am Main: Peter Lang, pp. 169–90.

Tsagari, Dina, Karin Vogt, Veronika Froelich, Ildikó Csépes, Adrienn Fekete, Anthony Green, Liz Hamp-Lyons, Nicos Sifakis, and Stefania Kordia. 2018. *Handbook of Assessment for Language Teachers*. Nicosia: TALE Project. ISBN 978-925-7399-0-5. (printed).

Vogt, Karin, and Dina Tsagari. 2014. Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly* 11: 374–402. [CrossRef]

West, Richard. 2004. *Testing and Assessment in Language Learning*. Patras: Hellenic Open University.

Widdowson, Henry. 1976. The authenticity of language use. In *Explorations in Applied Linguistics*. Edited by Henry Widdowson. Oxford: Oxford University Press.

Zarali, Lambrini, and Ifigenia Kofou. 2020. How aware are EFL teachers of developing and assessing learners' writing skill in an alternative way? Paper presented at the 6ο Διεθνές Συνέδριο, για την Προώθηση της Εκπαιδευτικής Καινοτομίας, Larissa, Greece, October 16–18. pp. 182–90. Available online: http://synedrio.eepek.gr (accessed on 31 August 2021).