



Arslan, R. Ş. & Üçok-Atasoy, M. (2020). An investigation into EFL teachers' assessment of young learners of English: Does practice match the policy? *International Online Journal of Education and Teaching (IOJET)*, 7(2), 468-484.

<http://iojet.org/index.php/IOJET/article/view/818>

Received: 20.01.2020
Received in revised form: 25.03.2020
Accepted: 28.03.2020

AN INVESTIGATION INTO EFL TEACHERS' ASSESSMENT OF YOUNG LEARNERS OF ENGLISH: DOES PRACTICE MATCH THE POLICY?

Research article

Recep Şahin Arslan 

Pamukkale University

rsarslan@pau.edu.tr

Pamukkale University

Meral Üçok-Atasoy 

meralucok@gmail.com

Ministry of National Education, Turkey

Recep Şahin Arslan is an associate professor at the English Language Teaching Department of Faculty of Education at Pamukkale University, Denizli, Turkey.

Meral Üçok-Atasoy is an English instructor at Necla-Ergun Abalıoğlu Vocational and Technical Anatolian High School, Denizli, Turkey.

Copyright by Informascope. Material published and so copyrighted may not be published elsewhere without the written permission of IOJET.

AN INVESTIGATION INTO EFL TEACHERS' ASSESSMENT OF YOUNG LEARNERS OF ENGLISH: DOES PRACTICE MATCH THE POLICY?¹

Recep Şahin Arslan

rsarslan@pau.edu.tr

Meral Üçok-Atasoy

meralucok@gmail.com

Abstract

On the grounds that assessment stands for a mirror of teaching and learning practices, its value cannot be ignored in teaching English as a Foreign Language (EFL) programmes as all those involved in foreign language teaching in non-native settings need constant feedback about the effectiveness of their ventures. Assessment of young learners of English has been also receiving rising attention as this group of language learners at the preliminary stages of learning a foreign language differ from adult learners in nature and thereby their assessment requires great care. While there exist continuous amendments in foreign language teaching policies nationally to improve the quality of EFL teaching and its assessment, it is significant to look inside the classrooms to realize whether the actual assessment practices reflect the performance outcomes expected by the policy documents. This paper, therefore, attempts to investigate the consistency between the ELT policy and EFL teachers' in-class practices of assessment of young learners in middle schools in the Turkish context. The study was conducted at the end of the spring term of 2017-2018 academic year with 152 EFL teachers working in middle schools in the central districts of Denizli province. The study employed both quantitative and qualitative research methods: the quantitative method provided information about EFL teachers' preferences of item types in terms of traditional and alternative assessment types with the help of a questionnaire while the qualitative method provided information about how frequently EFL teachers assessed four skills through exam papers they used in their classrooms. Results showed inconsistency between the policy and assessment practices of EFL teachers in the study: EFL teachers tended to design traditional paper and pencil tests based on language structures and vocabulary rather than the assessment of learners' communicative competence or language skills through alternative assessment methods.

Keywords: English as a foreign language (EFL), young learners, assessment, policy, practice.

1. Introduction

The booming technology and all lines of business, economics, politics and education require acquisition of English language for communication. With the aim of teaching English for communicative purposes, revisions in the educational policies of countries are therefore continuous. In Turkey, the Ministry of National Education (MoNE), responsible for the

¹ This study is based on the second author's M.A. dissertation submitted to the Graduate School of Educational Sciences, Pamukkale University, Denizli, Turkey in 2019.

supervision of public education under a national curriculum, designs English language teaching (ELT) programmes for all levels based on the Common European Framework of References for Languages (CEFR) (MoNE, 2018). Accordingly, Ministry of National Education policies emphasize English language teaching communicatively and suggest tasks and materials promoting learners' communicative competence. In such classrooms where the communicative competence is the main objective, assessment should also be in accordance with such objectives. In foreign language education, teaching and learning practices and assessment practices should go hand-in-hand as language assessment and teaching programme should be consistent with each other in terms of learning objectives, the kinds of tasks which the children are expected to perform, and the assessment types (Hughes, 2003). It is therefore important for teachers to understand the reasons and theoretical considerations behind such policy changes since they have the responsibility to transfer these changes to the classrooms. It can then be argued that tests can help to see what actually happens in the classroom since language assessment techniques and tools preferred by language teachers are assumed to mirror their teaching practices as well as perceptions about language teaching and learning (Alderson & Wall, 1993).

EFL teachers may be willing to create an appealing atmosphere and inspire students to be engaged in meaningful language activities keenly; however, the requirements of national structural examinations may put language teachers under pressure to complete the syllabus in a limited time, prepare students for examinations (Carless, 2003), and ignore the proposed assessment methods. One of the possible reasons that may cause language teachers to avoid alternative assessment tools may be the High School Placement Test, a standardized test offered by the MoNE involving only multiple choice items in contrast with the communicative objectives of ELT curriculum (Basok, 2017). Another possible reason for teachers' inconsistent assessment practices may be the limited time allocated to ELT courses in the curriculum, which may make it difficult to include four skills in the tests they apply. Regardless of what the underlying reasons may be, the mismatch between the policy and the practices of assessment is likely to bring out negative backwash effect to students' language learning and lead to undesirable consequences such as failure in the acquisition of communicative competence and ignorance of language skills (Paker, 2013). The reasons behind this mismatch need to be investigated thoroughly, and solutions be produced accordingly.

This study, therefore, aims to identify assessment practices of EFL teachers working in state middle schools seeking to find out whether there is a consistency or not between the assessment practices of EFL teachers of young learners in state middle school 5th, 6th, 7th, and 8th grades and the ELT Curriculum proposed by the MoNE for 5th, 6th, 7th, and 8th grades in middle schools. The study may therefore help English language teachers, teacher trainers, and curriculum developers better understand and improve the assessment of young learners in EFL classrooms. The study that attempts to shed light on EFL teachers' assessment practices at state middle schools seeks to answer the following questions:

1. What are the assessment practices of EFL teachers working in state middle schools?
2. To what extent are the assessment practices of EFL teachers consistent with the Ministry of National Education policy for the 5th, 6th, 7th and 8th grades' English Language Teaching Programmes?

2. Theoretical Background

Assessment has become one of the prevalent issues of today's language teaching and learning (Brown, 2007; Bachman & Palmer, 2010; Cheng & Fox, 2017) and can be realised in a number of varying applications. Cheng and Fox (2017) point out that in real classroom environments, teachers can apply both *assessment for learning* which is of formative assessment and *assessment of learning* which is of summative assessment. When the assessment provides “immediate feedback” for “ongoing teaching and learning”, this type of assessment is formative (Cameron, 2001, p.222), including informal quizzes and tests as well as observations and portfolios (Hughes, 2003). With the help of feedback by formative assessment, both teachers and students may make changes in their teaching and learning (Bachman & Palmer, 2010) since the purpose lying behind formative assessment involves more high-stake decisions instead of low-stake decisions (Mckay, 2009). On the other hand, summative assessment that can take place at the end of a unit, a term, a school year or any type of study period may be based on the teacher's summative observations of the students or the results of tests formalizing their achievement and focusing on the mastery of linguistic accuracy (Brown, 2004; Shaaban, 2005; Bachman & Palmer, 2010), emphasizing the linguistic competence rather than communicative competence (Shaaban, 2005). Additionally, Cheng and Fox (2017, p.188) argue that “teachers use assessment in their classrooms as something that is done *with* learners not *to* them” in order to stress the distinction between traditional and alternative types of assessment. Brown (2007) uses the term *alternatives in assessment* that include portfolios, projects, self-assessment, peer assessment, journals, formal/informal observations, presentations, informal questioning, and teacher-student conferences, and self- and peer assessments. As a matter of fact, traditional assessment that focuses on the accurate production of structures through such common item types as multiple-choice items, true/false items, matching items, and fill-in-the-blank items (Simonson, Smaldino, Albright, & Zvacek, 2000) fails to address the complex uniformity of language for communicative purposes (Clark, 1972; Oller, 1976). This type of assessment can be more practical than alternative assessment and can be preferred more. However, alternative assessment that requires more time, more subjective evaluation, more individualization, and more interaction in the process of providing feedback (Brown, 2004, 2007; Shaaban, 2005, Cameron, 2001, Mckay, 2009; Bachman & Palmer, 2010) indicates successful performance, highlights positive traits, provides formative rather than summative evaluation and takes into account students' needs, interests, and learning styles (Shaaban, 2005).

Until the revision of the Ministry of National Education 1997 foreign language teaching curriculum in 2005, assessment in EFL classrooms had been based on traditional structure-based paper-and-pencil tests and after the revision, performance-based assessment was proposed in parallel with the principles of CLT (Kırkgöz, 2007). Along with the 2012 reform in the Turkish educational system and the ELT Programme, MoNE (2013, p. XV) suggested four types of assessment: “project and portfolio evaluation, pen and paper tests, self and peer evaluation, and teacher observation and evaluation. The revised version of the ELT Programme for Primary and Middle Schools published in 2018 stresses the unity of teaching, learning and assessment in order to create *beneficial backwash effect* on the whole teaching and learning process (MoNE, 2018). With the revision in 2018, MoNE made dramatic changes in the suggested assessment types and techniques. When compared to the previous types of assessment, this revised version has a broader scope and supports a mixture of all assessment types instead of overuse of certain assessment techniques. In the 2018 ELT Programme, *learner autonomy* and *communicative competence* in language teaching have certain emphasis and accordingly *self-assessment*, *alternative*, and *process-oriented assessment* are within the main suggested assessment tools (MoNE, 2018). The suggested testing techniques for the

assessment of four skills for the ELT programme of MoNE (2018, pp. 7-8) include testing techniques for four language skills as well as integrated skills and alternative assessment such as “Portfolio Assessment, Project Assessment, Performance Assessment, Creative Drama Tasks, Class Newspaper/Social Media Projects, Journal Performance” in line with CEFR assessment (CoE, 2001). On the other hand, formal assessment tools such as written and oral exams, quizzes, homework and projects are within the suggested assessment tools in the 2018 ELT Programme (MoNE, 2018). It is stated in the ELT Programme (MoNE, 2018) that 2nd and 3rd graders are suggested to be assessed with the help of formative procedures. However, the young learners in the 4th, 5th, 6th, 7th, and 8th grades are advised to be assessed via both summative and formative assessment tools and techniques in both product and process-oriented procedures. It is obvious that MoNE (2018) suggests EFL teachers of young learners to utilize all the possible assessment techniques and tools in regard to young learners’ developmental features since young learners are different in nature and their assessment needs to reflect such peculiar characteristics (Cameron, 2001; McKay, 2006; Nikolov, 2016; Cheng & Fox, 2017).

Limited research investigating studies on the EFL teachers’ assessment practices of young learners of English may offer insights into understanding how assessment is implemented in EFL contexts. Yildirim and Orsdemir’s (2013) study to find out the reality in classrooms of young learners in the Turkish context in terms of availability of performance tasks in line with the policy proposals revealed that teachers utilized performance tasks effectively compatible with the curriculum; however, the document analysis showed a different application as listening, reading and speaking skills were totally ignored while writing and grammar were slightly fostered. Thus, the researchers concluded that rather than a match, a mismatch showed up between EFL teachers’ assessment practices of performance assessment in young learners’ classrooms and the policy proposals. The results of Brumen, Cagran, and Rixon’s (2009) study with EFL teachers of young learners in three Eastern European countries showed that Croatian teachers were prone to assessing listening and speaking more frequently, Czech teachers mostly went in for assessing the literacy skills (reading and writing) rather than oral skills (listening and speaking) and Slovenian teachers tended to use more grammar and vocabulary-oriented tests. In the Turkish context Han and Kaya’s (2014) survey to find out primary and secondary state Turkish EFL teachers’ perceptions and in-class practices of assessment of four skills revealed that reading and writing skills were mostly assessed by EFL teachers while listening and speaking were assessed less frequently. Similarly, Basok (2017) investigated the consistency between policy and implementations of the curriculum by the EFL teachers in Turkey and the study results showed that teachers could not implement what the policy suggested and instead they preferred to prepare the students for the examinations by using grammar-based teaching and assessment practices. Sarıgöz and Fişne’s (2018) investigation of the consistency between the policy and the actual language assessment practices in the 4th grade classrooms also revealed that English language assessment and evaluation fit formative purposes and written exams and assignments mainly tested learners’ writing and vocabulary. In addition, this particular study aims to find out the extent to which EFL teachers of young learners working in state middle schools follow the suggested procedures while assessing young learners and also to what extent they assess language skills.

3. Method of Study

This study, which seeks to examine EFL teachers’ testing and assessment practices and their consistency with the ELT Programme suggested by the Turkish MoNE, was designed as a mixed-methods research. Since single application of questionnaire as a means of quantitative data would result in insufficient information about how the EFL teachers assess middle school students’ English, assessment documents used by the EFL teachers in real classrooms would enhance the results of quantitative data. In this study, *parallel databases design* under the

Convergent Parallel Approach of mixed-methods research was applied through “the collection of different but complementary data on the same phenomena” and “for the converging and subsequent interpretation of quantitative and qualitative data” (Edmonds & Kennedy, 2017, p.181). Parallel databases design involves the simultaneous but separate collection of the quantitative and qualitative data and “allows researchers to validate data by converging the QUAN results with the QUAL findings” (Edmonds & Kennedy, 2017, p.182). In this study, the quantitative method provides information about EFL teachers’ preferences of item types in terms of traditional and alternative assessment types with the help of a questionnaire. Qualitative method provides information about how frequently EFL teachers assess four skills through exam papers teachers use in their classrooms. The quantitative and qualitative data are collected at the same time but with the help of different collection tools in conformity with parallel databases design.

3.1. Data Collection Procedures

3.1.1. Setting & participants

This study aims to find out the assessment practices of EFL teachers working with young learners at 5th, 6th, 7th and 8th grades in state middle schools in two central districts (Merkezefendi and Pamukkale) in Denizli. In the target districts of Denizli there were 286 EFL teachers working in 70 state middle schools in the spring term of 2017-2018 academic year. On a voluntary basis, 152 EFL teachers out of 286 accepted to participate in the study. In the first part of the questionnaire a consent part was presented to the teachers in order to formally ensure the teachers’ willingness to participate in the study and also to share their assessment documents. However, in the qualitative data collection procedure only 41 out of 152 teachers voluntarily shared their documents they used in assessing their students. The documents were only formal achievement tests which were administered after a few units were completed during and at the end of the semester. 56 achievement tests were collected in total. In this study, the tests are mentioned as *assessment documents* or *exam papers* interchangeably. 69.08 % of participants were female EFL teachers while 30.92 % of participants were male EFL teachers. The vast majority (75.7 %) of teachers graduated from the ELT Departments while 17 teachers (11.2 %) graduated from English Language and Literature Departments and 11 teachers graduated from other teaching branches such as Maths Teaching, Turkish Teaching and Primary School Teaching. The highest percentages belong to teachers whose experiences were between 6-10 years (n=48) and 11-15 years (n=58). The lowest percentage belongs to the teachers who had experience for over 20 years (n=4). 94.1% of teachers (n=143) held BA degree while 5.3% of teachers (n=8) held Master of Arts (MA) and only one teacher was with a PhD degree.

3.1.2. Instrumentation

In the quantitative part of this study, a Likert-scale questionnaire was adapted and developed by utilizing the studies of Anderson (1998) and Çalışkan and Kaşıkçı (2010). When preparing the questionnaire, the researchers consulted two field experts and some English language teachers in order to enhance its validity. Of 25 five-point Likert-scale items in the questionnaire the first two items addressed teachers’ general attitudes of assessment in terms of accuracy and communicative competence. The other 23 items were composed of several traditional and alternative assessment types. Table 1 presents the items in the questionnaire.

Table 1. *ELT teachers' assessment practices*

Assessment 1: I design my tests in order to assess accuracy

Assessment 2: I design my tests in order to assess communicative competence

I apply the following assessment types:

- 1-Multiple-choice questions (students select the answer from a set of options).
- 2-True/False questions (students select one of two choices, true or false).
- 3-Matching questions (students select the answers in one list that match the ones in the other list).
- 4-Fill-in-the-blank questions (students fill in a word or a phrase in a blank).
- 5-Wh- questions (students write content information depending on the question word)
- 6-Yes/No questions (students scrutinize a question or statement and construct a short response starting with Yes or No).
- 7-Translation questions (students translate the given words or sentence/s into the requested language).
- 8-Unscramble (students places the given letters or words in order to construct the requested word/s or sentence/s).
- 9-Informal question-answer (you ask students questions during the teaching and learning process).
- 10-Oral exams (you rate students with interviews).
- 11-Teacher-student conferences (you engage in a focused discussion with students about their work without giving marks).
- 12-Informal observations (you rate students' performance without pre-set criteria).
- 13-Formal observations (you rate students' performance with pre-set criteria).
- 14-Role-playing (an improvised conversation performed by students when given a situation).
- 15-Musical presentation (students sing songs or rhymes).
- 16-Presentations (students-created report/demonstration).
- 17-Portfolios (students' compilations of selected work with rating/reflection)
- 18-Creative writing (students-created poetry, short stories)
- 19-Journals (students' personal writing on self-chosen or assigned topics)
- 20-Projects (assignments given to students which involve the use of more time and resources than available during the normal class period)
- 21-Products (student-created graphs, tables, crafts, maps, web pages)
- 22-Self-assessment (students evaluate their own work)
- 23-Peer assessment (students evaluate other students' work)

In addition, EFL teachers' assessment documents formed the qualitative part of the study. 41 of 152 teachers shared whatever they had used as assessment tools in their exams already administered to their students in middle schools.

3.1.3. Data analysis

Quantitative data obtained from the questionnaires were analysed by the Statistical Package for Social Sciences (SPSS) 24. Cronbach's Alpha value of study was .85 and the number of items was 25. Results of Kolmogorov-Smirnov and Shapiro-Wilk normality tests showed the quantitative data as non-parametric ($p < 0.05$). For this reason, Kruskal Wallis Test and Mann Whitney U Test as non-parametric tests were applied. Qualitative data were gathered from teachers' exam papers. As *document analysis* brings the elements of *content analysis and thematic analysis* (Bowen, 2009), *document analysis* (Bowen, 2009) was applied to find out the type of items teachers used in their exams and also whether or to what extent teachers

assessed four skills of EFL students. The exam papers of EFL teachers were therefore specifically analysed through *content analysis* (Cohen, Manion, & Morrison, 2018, p. 674). In that sense *superficial examination* was applied in this study to provide evidence to the information about the assessment types of teachers gathered from the questionnaires. Exam papers were firstly examined by the researchers in order to detect the existence and frequency of four skills' assessment both on grade basis (5th, 6th, 7th, 8th grades separately) and in total. In order to enhance the rater-reliability, the other two coders checked the exam papers one after another. By this way, the analysis of the questionnaire and the analysis of the assessment documents complemented each other.

4. Findings

4.1. EFL Teachers' Assessment Practices: Traditional or Alternative Assessment?

152 EFL teachers completed a questionnaire as to whether they applied traditional paper-pencil tests or alternative ways of assessment. Table 2 displays descriptive statistics regarding EFL teachers' traditional and alternative assessment practices.

Table 2. *Assessment types: traditional and alternative assessment*

Item Types	Never	Rarely	Sometimes	Usually	Always	Mean \bar{x}	Std. Deviation
Assessment 1 accuracy	0	17	32	84	19	3.69	.83
Assessment 2 communicative competence	4	24	51	58	15	3.36	.95
1. Multiple choice	0	15	36	52	49	3.88	.97
2. True-false	0	4	22	68	58	4.18	.77
3. Matching	0	6	17	56	73	4.28	.81
4. Fill in the blanks	0	7	28	50	67	4.16	.88
5. Wh-question (open-ended)	0	16	55	45	36	3.66	.95
6. Yes/No (closed)	19	28	77	20	8	2.80	.99
7. Translation	48	46	37	17	4	2.23	1.09
8. Unscramble (words/sentences)	8	24	52	47	21	3.32	1.06
9. Informal question/answer	3	17	47	49	36	3.64	1.02
10. Oral exams	28	40	64	12	8	2.55	1.04
11. Teacher student conferences	30	33	55	22	12	2.69	1.17
12. Informal observations	10	28	58	43	13	3.13	1.02
13. Formal observations	13	27	54	41	17	3.14	1.10
14. Role playing	1	14	48	52	37	3.72	.95
15. Musical presentation	21	17	46	39	29	3.25	1.27
16. Presentations	7	20	69	40	16	3.25	.97
17. Portfolios	7	43	48	32	22	3.12	1.11
18. Creative writing	26	45	55	18	8	2.58	1.07
19. Journals	46	47	38	19	2	2.23	1.05
20. Projects	7	15	31	60	39	3.71	1.09
21. Products	11	21	43	49	28	3.40	1.15
22. Self-assessment	30	30	57	25	10	2.70	1.15
23. Peer assessment	17	36	55	37	7	2.87	1.05

* Traditional Test: \bar{x} =3.87 (items 1-8)

*Alternative Assessment: \bar{x} =2.98 (items 9-23)

The first two items in the questionnaire were related to assessing accuracy and communicative competence: the mean scores of them were very similar while the teachers

preferred assessing accuracy ($\bar{x}=3.69$) more frequently than the assessment of communicative competence ($\bar{x}=3.36$). As it is also demonstrated in Table 2, *traditional assessment* and *alternative assessment* item types were compared in terms of teachers' frequency of preference: traditional assessment ($\bar{x}=3.87$) was more frequently preferred by the teachers than alternative assessment ($\bar{x}=2.98$). The highest mean scores belonged to *matching* ($\bar{x}=4.28$), *true-false* ($\bar{x}=4.18$) and *fill-in-the blank* items ($\bar{x}=4.16$) with slight differences. The lowest mean scores within traditional assessment belonged to *translation* ($\bar{x}=2.23$), *journals* ($\bar{x}=2.23$), *oral exams* ($\bar{x}=2.55$) and *creative writing* ($\bar{x}=2.58$). The highest mean scores within alternative assessment types belonged to *role-plays* ($\bar{x}=3.72$) and *projects* ($\bar{x}=3.71$) while the lowest mean scores belonged to *journals* ($\bar{x}=2.23$) and *oral exams* ($\bar{x}=2.55$). It could be concluded that most of the teachers preferred assessing accuracy rather than communicative competence in their exams or traditional assessment was more preferable for teachers than alternative assessment. Results supported that the mean scores of traditional assessment were higher than those of the alternatives. It meant EFL teachers mostly applied traditional pen and paper tests while assessing young learners in middle schools.

In addition to the questionnaire as a source of data about teachers' assessment practices, 41 out of 152 teachers shared 56 exam papers they administered during a semester. There were 12 papers for 5th grade, 13 papers for the 6th grade, 16 papers for the 7th grade and 15 papers for the 8th grade in this study. Table 3 demonstrates the item types used in the exams of 5th grade EFL students and their frequencies within all the 5th grade exam papers. 56 exam papers collected from EFL teachers were also analysed in order to find out the item types used and the skills assessed in the exams.

Table 3. *Item types: 5th grade exam papers*

Item Type	Related Linguistic Components	f	Total papers	%
Matching	Grammar- Vocabulary	12	12	100
Fill-in-the blank	Grammar-Vocabulary	12	12	100
Multiple choice	Grammar-Vocabulary- Reading	6	12	50
Wh- items	Grammar- Reading	4	12	33.3
Translation	Grammar	3	12	25
Unscrambling (word/sentence)	Grammar	2	12	16.6
Odd-one out	Grammar- Vocabulary	1	12	8.3
Restricted response essay (paragraph writing)	Writing	1	12	8.3

As can be seen in Table 3, eight different item types were detected concerning the EFL exams of 5th grade students. The most preferred item types were *matching* and *fill-in-the blanks* items of all the exams of 5th grade students. However, both *odd-one out* and *restricted response* items were used in only one exam. In half of the exam papers there were *multiple choice* items. *Wh- items*, *translation* and *unscrambling* items were within the item types but less frequently used in the 5th grade EFL exams. As for the linguistic components to be assessed in 5th grade exam papers, *matching*, *fill-in-the blank* and *odd-one-out* items were prepared to assess grammar and vocabulary components. *Multiple choice* items were prepared to assess reading skill in addition to grammar and vocabulary. *Wh- items* were prepared to assess grammar and reading skill. *Translation* and *unscrambling* items were prepared to assess only grammar.

Finally, restricted response essays were prepared to assess writing skill. In addition, Table 4 demonstrates the item types used in the 6th grade EFL exams.

Table 4. *Item types: 6th grade exam papers*

	Related Linguistic Components	f	Total papers	%
Matching	Grammar- Vocabulary	13	13	100
Fill-in-the-blanks	Grammar- Vocabulary- Listening	13	13	100
Multiple choice	Grammar- Vocabulary- Reading	8	13	61.5
Wh- items	Grammar- Reading	5	13	38.4
True/False	Reading	3	13	23
Translation	Grammar- Vocabulary	2	13	15.3

Table 4 shows six different types of items in the 6th grade exam papers. The most preferred item types by the EFL teachers were *matching* and *fill-in-the blanks* items in the 6th grade EFL exams the same as the 5th grade. Teachers used these item types in all the exams they administered. Another mostly used item type was *multiple choice* items. Eight out of 13 papers included multiple choice items. The other item types used in the 6th grade papers were *wh-*, *true/false* and *translation* items. Language components assessed in the 6th grade exam papers were similar to the ones in the 5th grade exam papers. For instance, *matching* and *translation* items were prepared to assess grammar and vocabulary; *fill-in-the blank* items were prepared to assess listening skill in addition to grammar and vocabulary; *multiple choice* items were prepared to assess reading skill in addition to grammar and vocabulary; and *true/false* items were prepared to assess only reading skill. Table 5 demonstrates the item types included in the 7th grade EFL exam papers.

Table 5. *Item types: 7th grade exam papers*

	Related Linguistic Components	f	Total papers	%
Matching	Grammar- Vocabulary	16	16	100
Fill-in-the-blank	Grammar- Vocabulary	16	16	100
Multiple choice	Grammar- Vocabulary- Reading	10	16	62.5
Wh- items	Grammar- Reading	7	16	43.7
Unscrambling(word/sentence)	Grammar- Vocabulary	5	16	31.2
Translation	Grammar- Vocabulary	3	16	18.7
Yes/No	Reading	2	16	12.5

According to Table 5 seven types of items were used in the 7th grade EFL exam papers. Similar to 5th and 6th grades, *matching* and *fill-in-the blanks* items were used in all the 7th grade exam papers. *Multiple choice* items, one of the most preferred items, were available in 10 out of 16 7th grade exam papers. The other item types used in the 7th grade EFL exams were *wh-*, *unscrambling*, *translation* and *Yes/No* items. As for the linguistic components to be assessed

in the 7th grade exam papers, it is clear that they were prepared with similar purposes to the items prepared in the 5th and 6th grade exam papers. For instance, *matching*, *fill-in-the blank*, *unscrambling* and *translation* items were prepared to assess grammar and vocabulary. Likewise, *multiple choice* items were prepared to assess reading skill in addition to grammar and vocabulary. *Wh- items* were prepared to assess grammar and reading. Finally, *Yes/No items* were prepared to assess reading skill. Table 6 demonstrates the item types used in the 8th grade EFL exams.

Table 6. *Item types: 8th grade exam papers*

	Related Linguistic Components	f	Total papers	%
Matching	Grammar- Vocabulary	15	15	100
Fill-in-the blanks	Grammar- Vocabulary	15	15	100
Multiple choice	Grammar- Vocabulary- Reading	13	15	86.6
Wh- items	Grammar- Vocabulary- Reading	8	15	53.3
True/False	Reading	6	15	40
Yes/No items	Grammar- Reading	5	15	33.3
Restricted response (paragraph writing)	Writing	3	15	20
Error-correcting	Grammar	2	15	13.3
Odd-one-out	Vocabulary	1	15	6.6

According to Table 6 there were nine types of items in the 8th grade EFL exams. Not surprisingly, *matching* and *fill-in-the blanks* items were available in all the 8th grade exam papers. More frequently than in the other grade exam papers, *multiple choice* items were within the mostly preferred item types in the 8th grade exam papers. Differently from the item types in the exam papers of 5th, 6th, and 7th grades, in two 8th grade exam papers there were *error correcting* items. Linguistic components assessed in the 8th grade exam papers were similar to the ones in the 5th, 6th and 7th grade exam papers. For instance, *matching* and *fill-in-the blank* items were prepared to assess grammar and vocabulary; *multiple choice* and *wh- items* were prepared to assess reading skill in addition to grammar and vocabulary; *true/false* items were prepared to assess only the reading skill; *Yes/No items* were prepared to assess grammar and reading skill; *restricted response essays* were prepared to assess writing skill; *error-correcting* items were prepared to assess grammar; and *odd-one-out* items were prepared to assess vocabulary.

4.2. EFL Teachers' Assessment of Language Skills

Furthermore, analysis of exam papers prepared by EFL teachers for young learners of English at state middle schools in this particular study shows that no exam papers were designed specifically for the assessment of four skills. All the papers included formal questions for the assessment of a few skills together. Table 7 demonstrates the frequencies and percentages of assessed skills. *Frequency (f)* refers to the existence of the skills per exam paper.

Table 7. *Frequencies of assessed skills*

Skills	5 th Grade			6 th Grade			7 th Grade			8 th Grade		
	f	Total	%	f	Total	%	f	Total	%	f	Total	%
Listening	0	12	0	1	13	7.69	0	16	0	0	15	0
Reading	6	12	50	6	13	46.15	9	16	56.2	15	15	100
Writing	1	12	8.33	3	13	23.07	0	16	0	3	15	20
Grammar	12	12	100	13	13	100	16	16	100	15	15	100
Vocabulary	12	12	100	13	13	100	16	16	100	15	15	100

Table 7 displays that none of the 5th grade papers included the assessment of listening skill. In half (n=6) of the 5th grade papers there were parts assigned to reading questions. In all the 5th grade exam papers (n=12) there were parts assigned to the assessment of grammar and vocabulary. Finally, in only one of the 5th grade exam papers there was a part in which the students were requested to write a paragraph on a given topic which was intended to assess writing skill. Likewise, in the 6th grade exam papers (n=13) there were questions prepared for both grammar and vocabulary assessment. In nearly half (n=6) of the 6th grade papers there were parts involving questions for reading assessment. In three of the 6th grade exam papers there was a part assigned to writing a paragraph. Surprisingly in only one of the 6th grade exam papers there was a part which involved questions for listening assessment about an audio-record. The percentages of the assessed skills at 7th grade were similar to the percentages at 5th and 6th grades. For example, in all the exam papers at 7th grade there were parts which involved questions for both grammar and vocabulary assessment. Similar to 5th and 6th grade papers, there were no single questions assigned to listening and writing skills in any of the 7th grade exam papers. In nine of the 7th grade papers there were parts assigned to the assessment of reading skill. The results of 8th grades were very similar to the results of 5th, 6th and 7th grade exam paper analyses. For example, again grammar and vocabulary were assessed in the entire (n=15) 8th grade EFL exam papers. In 8th grade papers, there was a remarkable difference from the other grades in the percentage of reading assessment: the entire exam papers involved reading assessment. In the 8th grade EFL exam papers it was exactly the same as the 5th, 6th and 7th grades since there was no inclusion of listening skill at all.

With regard to the percentages of assessed skills in the 56 exam papers in total, it is clear that EFL teachers tended to assess grammar and vocabulary in all exams while they did not assess listening and speaking except for one 6th grade exam paper involving a listening part. As to the assessment of speaking skill assessment teachers did not share any separate assessment documents, so it can be inferred that EFL teachers did not assess speaking skill at all. As for the reading skill, it was assessed in all exam papers just like grammar and vocabulary at the 8th grade. However, at the other grades percentage of reading assessment decreased below 50 %. On the other hand, it could also be inferred that teachers did not prefer assessing writing skill regularly at middle schools since its percentage was also pretty low.

5. Discussion

Both quantitative and qualitative data were analysed and it was found out that despite the proposals of the policy which insistently emphasize communicative language testing and alternative ways of assessment in harmony with other possible assessment tools, teachers utilized merely the traditional paper and pencil exams. Although the teachers reported that they used alternative ways of assessment together with the traditional types in the questionnaires, the only assessment tools shared by the teachers were exam papers rather than materials of

alternative assessment such as portfolios and projects. On this basis, it could be inferred that EFL teachers working in state middle schools assessed their students by applying achievement tests in certain periods during the semester. Even though they might have used alternative types of assessment to some extent, they did not share them with the researchers.

The findings gathered from the document analysis of the teachers' exam papers were also parallel to the findings of the descriptive statistics. The exam papers were analysed in order to detect the item types used by the teachers. Accordingly, it was determined that the most frequent items were matching, fill-in-the blanks and multiple choice items. That is to say, document analysis enabled us to crosscheck the findings of the questionnaire: there was a perfect match between the findings of these data. It was remarkable that in every single exam paper there were matching and fill-in-the blank items and all of them were prepared to assess grammar and vocabulary.

Similar to our study, in some other studies (Pandian, 2002; Brumen et al., 2009; Han & Kaya, 2014; Basok, 2017) both traditional and alternative assessment types were practiced by EFL teachers. Similar to Turkish teachers of English in our study, Slovenian EFL teachers also mostly preferred fill-in-the blank items; Czech EFL teachers mostly preferred true/false items while Croatian teachers mostly preferred repeat-and-drill practices in the study of Brumen et al. (2009). Additionally, Han and Kaya's (2014) study had also similar findings to our study. For example, in both studies true/false and matching items were the mostly preferred traditional assessment tools in order to assess reading skill. Moreover, in both of the studies, teachers reported that role-plays were mostly preferred alternative assessment tools in order to assess speaking skill.

All in all, teachers did not administer any separate skills examinations in middle schools, but they mostly prepared exams in which grammar, vocabulary and reading had the greatest inclusion. Teachers did not assess speaking and listening skills of young learners at any grade levels with an exception of one exam paper including a listening part. Basok (2017) came up with similar findings to our study findings: EFL teachers declared that they designed structure-based exams including grammar and reading assessment; whereas, they ignored communicative skills of listening and speaking because of the pressure by the central language examinations administered by the government. In the study reported by Pandian (2002), EFL teachers also prepared exams including grammar, vocabulary, reading and writing assessment but ignored listening and speaking skills similar to Basok's (2017) and our studies. Additionally, Yildirim and Orsdemir (2013) had similar findings in terms of the assessment of four skills; for instance, EFL teachers ignored speaking, listening and reading skills totally while preparing performance assessment in young learners' classrooms; and they just included grammar and writing in the performance tasks. However, there was a difference: in our study reading skill was among the mostly assessed language skills while in Yildirim and Orsdemir's (2013) study it was not assessed via performance tasks by the EFL teachers. Han and Kaya (2014) came up with very similar findings to those of our study and also to the aforementioned studies in terms of the assessment of four skills on the grounds that listening and speaking assessment were totally ignored and reading and writing skills were assessed through the exams prepared by the EFL teachers working with young learners.

Brumen et al. (2009) indicated similar findings to our study findings. Slovenian teachers mostly assessed grammar and vocabulary and made use of fill-in-the blanks type of items in their EFL exams. Czech teachers put the emphasis on literacy skills (reading and writing) and overused true/false items in the exams of young learners; Contradictorily Croatian teachers ignored literacy skills and focused on oral skills in company with repeat-and-drill exercises of vocabulary. The study findings obtained by Sarıgöz and Fişne (2018) are also similar to our

study findings as EFL teachers' assessment and evaluation practices of the 4th graders fit the formative purposes and writing and vocabulary components were also among the common assessment types through written exams and assignments.

It can also be argued that the assessment practices of EFL teachers were not consistent with those stated by the Ministry of National Education for the 5th, 6th, 7th and 8th grades English Language Teaching programmes. Together with the results of the quantitative data it could be inferred that EFL teachers tended to implement grammar-based traditional paper and pencil assessment procedures as opposed to policy suggestions by the MoNE (2018, pp. 6-7) including summative and formative, product and process-oriented tests and traditional and alternative assessment tools which cover four skills and all the linguistic components.

In parallel with our study, Basok's (2017) investigation into the consistency between the curriculum and the implementations of EFL teachers working in primary, secondary and high schools showed that assessment implementations did not match the policy. Yildirim and Ordemir's (2013) study also indicated a mismatch between the curriculum proposals and the teachers working with young learners as teachers tended to assess grammar and writing rather than all four skills in their implementations of performance tasks. The contradiction between the in-class practices and assessment procedures may put forth a trouble in the validity and reliability of the exams. Even though teachers may attempt to integrate the language skills into their teaching, assessment practices lacking those skills can cause a mismatch even between their own practices of teaching and assessment before the policy. However, unlike all these study findings and also ours, Kirkgoz, Babanoglu, and Ağçam (2017) came up with contradictory results since EFL teachers of young learners in the Turkish state primary school context preferred performance-based and communication-based assessment more than the traditional assessment. Such a practice is promising and needs to be disseminated in the overall Turkish context as this is one of the major reasons for which we conducted this particular study.

6. Conclusion

CEFR has been accepted and implemented as a pathfinder in Turkey since 2006 and accordingly ELT programmes have been revised several times in terms of language teaching, language learning and assessment of language (MoNE, 2013; MoNE, 2018). Considering the policy innovations in the ELT programmes of the MoNE, this study aimed to find out the assessment practices of EFL teachers in middle schools (5th, 6th, 7th, and 8th grades) by examining the types of assessment tools EFL teachers used in young learners' classrooms and also to find out the extent of consistency between the proposed course outcomes of the ELT Programme suggested by the MoNE and the EFL teachers' assessment practices in middle schools.

Findings of the descriptive statistics revealed that EFL teachers preferred to assess accuracy more frequently than communicative competence as they used traditional assessment more frequently than alternative assessment while they mostly preferred matching, true/false, fill-in-the blank and multiple choice items rather than translation, journals, oral exams and creative writing. The findings of the document analysis substantially supported such findings; namely, the skills assessed in the exam papers were grammar and vocabulary with a hundred percent. In nearly half of the exam papers, there were parts for reading assessment and in some exam papers writing skill was also assessed. However, in any of the exam papers there were no questions for the assessment of listening skill. As for the speaking skill, since it cannot be assessed through written materials, and none of the teachers shared any documents or declared they assessed speaking skill, it was interpreted that teachers did not assess speaking skill at all. Above and all, the findings of this study revealed that the assessment practices of the EFL

teachers working in state middle schools did not match the CEFR-oriented ELT policy of the MoNE since the EFL teachers tended to design traditional structure-based tests instead of a harmony of all kind of assessment tools and techniques based on communicative competence. Why EFL teachers prefer traditional assessment types may be related to their language assessment knowledge (LAK) as Ölmezer-Öztürk and Belgin (2019) found out in their study that EFL teachers received low scores in the LAK scale they developed. Tavassoli and Farhady (2018) also support such results as EFL teachers in their study reported that they needed to improve their language assessment knowledge. Such studies might indicate the urgency of in-service training of EFL teachers in alternative assessment types.

This study being one of the few studies examining the assessment practices of EFL teachers of young learners and their consistency with the policy in Turkey might contribute to the comprehension of the policy and its implementation in a more compatible way in the Turkish context. The study may provide feedback to the teachers, teacher trainers, and policy makers in order to find a common ground in language assessment. Above all, it is significant to determine the underlying reasons of the inconsistency between policy and practice to be able to produce applicable solutions. Therefore, this study paves the way for further research on investigating the background problems of this inconsistency. Researchers may conduct more research focusing on the underlying reasons of such problems in the implementation of the policies at schools.

References

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics*, 14(2), 115-129.
- Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, 1998(74), 5–16. <https://doi.org/10.1002/tl.7401>
- Brumen, M., Cagran, B., & Rixon, S. (2009). Comparative assessment of young learners' foreign language competence in three Eastern European countries. *Educational Studies*, 35(3), 269–295. <https://doi.org/10.1080/03055690802648531>
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Basok, E. (2017). *Language teaching policies and practices in the Turkish EFL context and the effects on English teachers' motivation*. (Unpublished M.A. Thesis). The University of Texas at San Antonio.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRI0902027>
- Brown, H.D. (2004). *Language assessment. Principles and classroom practices*. NY, USA: Pearson Education.
- Brown, H.D. (2007). *Teaching by principles. An interactive approach to language pedagogy*. NY, USA: Pearson Education.
- Cameron, L. (2001). *Teaching languages to young learners*. UK: Cambridge University Press.
- Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, 31(4), 485–500. <https://doi.org/10.1016/j.system.2003.03.002>
- Cheng, L., & Fox, J. (2017). *Assessment in the language classroom. Teachers supporting student learning*. UK: Macmillan Education.
- Clark, J.L.D. (1972). *Foreign language testing: theory and practice*. Philadelphia: Center for Curriculum Development.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education (8th ed.)*. London and New York: Routledge.
- Council of Europe (CoE). (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.
- Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.
- Çalışkan, H., & Kaşıkçı, Y. (2010). The application of traditional and alternative assessment and evaluation tools by teachers in social studies. *Procedia - Social and Behavioral Sciences*, 2(2), 4152–4156. <https://doi.org/10.1016/j.sbspro.2010.03.656>

- Edmonds, W, A., & D. Kennedy, T. (2017). *An applied guide to research designs quantitative, qualitative, and mixed methods (Second Edition)*: ISAGE Los Angeles.
- Han, T., & Kaya, H. İ. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education*, 4(1), 77. <https://doi.org/10.5296/jse.v4i1.4873>
- Harmer, J. (2007). *The practice of English language teaching*. Essex: Pearson Education.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Kırkgöz, Y. (2007). English language teaching in Turkey: policy changes and their implementations. *RELC Journal*, 38(2), 216-228.
- Kirkgoz, Y., Babanoglu, M. P., & Ağçam, R. (2017). Turkish EFL Teachers' perceptions and practices of foreign language assessment in primary education. *Journal of Education and e-Learning Research*, 4(4): 163-170.
- McKay, P. (2006). *Assessing young language learners*. UK: Cambridge University Press
- McKay, P. (2009). Assessing Young Language Learners. *ELT Journal*, 63(1), 91-94. <https://doi.org/10.1093/elt/ccn063>
- Ministry of National Education (MoNE). (2013). *İlköğretim İngilizce dersi (2, 3, 4, 5, 6, 7 ve 8. sınıflar) öğretim program [Primary education English language curriculum (the 2nd-8th grades)]*. Retrieved from <http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=327>.
- Ministry of National Education (MoNE). (2018). *İlköğretim İngilizce dersi (2, 3, 4, 5, 6, 7 ve 8. sınıflar) öğretim program [Primary education English language curriculum (the 2nd-8th grades)]*. Retrieved from <http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=327>.
- Nikolov, M. (2016). *Assessing young learners of English: Global and local perspectives*. Switzerland: Springer International Publishing. DOI10.1007/978-3-319-22422-0
- Oller, J.W. (1976). Language testing. In R. Wardhaugh and H.D. Brown (Eds.). *A survey of applied linguistics* (275-300). Ann Arbor: University of Michigan Press.
- Ölmezer-Öztürk, E. & Belgin, A. (2019). Investigating language assessment knowledge of EFL teachers-İngilizceyi yabancı dil olarak öğreten öğretmenlerin dilde ölçme değerlendirme bilgilerinin araştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education)*. 34(3): 602-620. doi: 10.16986/HUJE.2018043465
- Pandian, A. (2002). English Language Teaching in Malaysia today. *Asia Pacific Journal of Education*, 22:2, 35-52, DOI: 10.1080/0218879020220205

- Paker, T. (2013). The Backwash Effect of the Test Items in the Achievement Exams in Preparatory Classes. *Procedia-Social and Behavioral Sciences*, 70, 1463–1471. <https://doi.org/10.1016/j.sbspro.2013.01.212>
- Sarıgöz, İ. H., & Fişne, F. N. (2018). English language assessment and evaluation practices in the 4th grade classes at mainstream schools. *Journal of Language and Linguistic Studies*, 14(3), 380-395.
- Shaaban, K. (2005). Assessment of young learners. *English Teaching Forum*, 43(1), 34–40. Retrieved from <http://americanenglish.state.gov/resources/english-teaching-forum-2005-volume-43-number-1#child-567>
- Simonson, M., Smaldino, S, Albright, M. & Zvacek, S. (2000). *Assessment for distance education (ch 11)*. *Teaching and Learning at a Distance: Foundations of Distance Education*. Upper Saddle River, NJ: Prentice-Hall.
- Tavassoli, K. & Farhady, F. (2018). Assessment knowledge needs of EFL teachers. *Teaching English Language*. 12(2), 45-65.
- Yildirim, R., & Orsdemir, E. (2013). Performance tasks as alternative assessment for young EFL learners: Does practice match the curriculum proposal? *International Online Journal of Educational Sciences*, 5(3), 562–574. Retrieved from <http://ezproxy.scu.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ejh&AN=93436242&site=ehost-live>