Integration of Task-based Language Assessment into Young Learner Classrooms

İskender Hakkı Sarıgöz a * ID , Fatıma Nur Fişne a † ID

a Gazi University, English Language Teaching Department, Ankara 06500, Turkey

Abstract

Task-based language assessment (TBLA) is a way of providing information about language learning in authentic settings. Not only does it present language learners with a meaningful and embedded context of assessment but also promotes communication in the target language. Within the framework of TBLA, this study aims to integrate the 4th grade language testing into language learning process through ILTPY (Integrated Language Testing Program for Young Learners). ILTPY includes ten-unit assessment tasks which have been developed in accordance with the young learners' characteristics and the 4th grade curricular objectives at state primary schools affiliated with the Ministry of National Education (MoNE) in Turkey. In order to test the effectiveness of ILTPY, two-unit assessment tasks were employed in the 4th grade classrooms. At the end of the two-unit implementation, follow-up interviews were conducted with the students and teachers in order to understand the influence of ILTPY on language learning process. The study findings show that the assessment tasks in ILTPY significantly increased the level of language attainment in the 4th grade. The two-unit implementation motivated a majority of language learners by offering them to reinforce what they had learnt. The students' increasing enthusiasm and participation, and the opportunity for professional development motivated most of the language teachers to teach English and assess the students as emphasized in ILTPY. However, task development and implementation were regarded as the demotivating factors in view of practicality. The assessment tasks mostly met the young learners' needs of assessment, but they should be supported with more visuals and technology. Finally, it is considered that ILTPY generally affected language learning positively.

Keywords: Assessing young EFL learners; task-based language assessment; integrated language testing program for young learners

1. Introduction

Foreign language testing has witnessed a number of historical movements on theoretical grounds over the years. There has been a gradual transition from the traditional-structural testing to communicative language assessment (Brown, 2005; Madsen, 1983). Along with the embeddedness of communicative competence into

* Corresponding author. Tel.: +0-312-202-8468
  E-mail address: iskender@gazi.edu.tr

† fatimanurfisne@gazi.edu.tr

language teaching, the communicative aspect of language testing has gained importance in the instructional settings (McNamara, 2000), and a growing interest has been placed on the use of assessment tasks in language classrooms. In other words, the utilization of TBLA has received much attention in the field of language testing. TBLA is terminologically based on the performance assessment of learners by means of the language use tasks which present the learners with real-life experiences. In relation to the use of task-based assessment in language learning, there is a bulk of research which has been carried out in different educational contexts (Brynes, 2002; Elder, Iwashita, & McNamara, 2002; Ke, 2006; Mislevy, Steinberg, & Almond, 2002; Norris, Brown, Hudson, & Bonk, 2002; Robinson & Ross, 1996; Spence-Brown, 2001). Whereas some of these studies focused on how task-based assessment echoes in the foreign or second language instruction, some others investigated the measurement-based specifications, such as task development, task difficulty and task authenticity. According to Shehadeh (2012), TBLA can be administered to the learners who have "different age groups, different proficiency levels, and different educational and cultural backgrounds" (p. 159). The young learner field includes some research studies that are mostly established on the implementation of task-based teaching (Carless, 2002; 2003). There has been little research on the administration of TBLA within the scope of the communicative language testing. In order to promote the use of language tasks in young learner assessment, this study aims to develop an integrated language testing program (ILTPY) including the embedded assessment tasks for the 4th grade young EFL learners at state primary schools in the Turkish MoNE. Throughout the implementation of ILTPY, it is also aimed to increase the language achievement and attainment, and provide the stakeholders with comprehensive information about English language learning in young learner classrooms.

2. Review of Literature

Language testing, as an essential component of English language teaching, performs a high-stake role in shaping language learning procedures and practices. It provides a basis for measuring the success of learning, understanding to what extent language objectives have been achieved, getting familiarized with the learner performance in a classroom setting, receiving feedback on language instruction and teaching materials, and giving information for decision-making (Bachman & Palmer, 2010). Historically, language testing has gone through several stages from traditional and standardized testing to integrative and communicative language assessment (Brown, 2005; Brown & Hudson, 1998; Madsen, 1983; Weir, 2005). Whereas traditional testing emphasizes the classical pen-and-paper based examinations, communicative assessment requires the students "to produce and comprehend the language" (Shohamy, 1997, p. 142). Canale and Swain (1980) discuss the communicative aspect of language assessment and accentuate that learners should present their linguistic gains in an authentic language context. In this regard, language use tasks enable the students to take part in "comprehending, manipulating, producing or interacting in the target language" (Nunan, 2004, p. 4),

and thus these tasks enhance the authenticity and communication in language assessment.

Putting a central emphasis on the use of the target language in assessment, TBLA attempts to measure and interpret to what degree language learning objectives have been attained through the real-life communicative activities, with which the learners may be familiar outside the assessment setting (Brindley, 1994; Elder et al., 2002; Nunan, 2004; Robinson & Ross, 1996; Wigglesworth, 2008). TBLA not only presents learners with a meaningful assessment setting, gives feedback on language performance, provides information for decision makers (Norris, 2009; Norris et al., 2002) but also focuses on the social, cultural and pragmatic aspects of the target language (Mislevy et al., 2002), and it helps bring about a positive washback effect on language teaching (Wigglesworth & Frost, 2017). Norris (2009) discusses the advantages of integrated assessment tasks (*lesson-embedded tasks)* in language classrooms and relates those gains to the formative tenet of language assessment. More precisely, integrated (embedded) tasks demonstrate to what degree language content has been learned, and these tasks raise learners' awareness on language learning, motivate them for setting goals, and give information about how effective instructional activities are (Norris, 2009). This indicates that assessment tasks are authentically used to monitor the language learning processes and outcomes, and therefore they enable YLLs (young language learners) to develop an understanding of their own learning. Assessment tasks may be also used in the young learners' classes to track the pupils' language learning development and language attainment. YLLs have different characteristics in nature, and language assessment should be conducted in accordance with these characteristics (Carr, 2011). To be more specific, assessment practices should be developed in consonance with language learning, and these practices should "support learning and teaching" (Cameron, 2001, pp. 219-221). Assessment tasks should include cognitively appropriate visuals, games, and stories, and young learners should be assessed through the social interaction in the target language (Hughes, 2003). Furthermore, these assessment tasks should motivate young learners (McKay, 2006).

In TBLA field, there is a body of research studies which have examined different variables in task-based design or implementation. By and large, these studies can be categorized under four sections: the administration of assessment tasks at tertiary level, task-related aspects in TBLA, TBLA models, and difficulties in TBLA implementation. For instance, Byrnes (2002) investigates task-based assessment of the writing skill in a foreign language program for "collegiate" students (p. 419). This study points out that TBLA in writing makes contribution to "instructed language development" and "interlanguage development" (p. 435). In another study, Ke (2006) draws attention to a model of formative task-based language assessment in the Chinese language program at university level. The main characteristics of this assessment model are "criterion-referenced testing", "skills integration", "componential scoring", and "derivation from class activities" (pp. 212-215). The research findings indicate that this kind of assessment helps language teachers focus

on the current instruction and give feedback on language learning. Considering the aspects of task-based assessment, Elder et al. (2002) examine the concept of task difficulty in oral language assessment in terms of performance conditions and characteristics. The findings of this study illustrate that "the perception of task difficulty is a multidimensional phenomenon" (p. 364), and it may be influenced by some variables including task-related issues and test-takers' characteristics. Similarly, Norris et al. (2002) shed light on the use of language tasks in the performance-based assessment and discuss the results in view of the examinee abilities and task difficulty. The aspect of authenticity is examined by Spence-Brown (2001) in a teaching and assessment task in the Japanese language course at university level. The study reveals that the assessment dimension of a language task affects the degree of task authenticity. Regarding TBLA models, Mislevy et al. (2002) present a conceptual framework for TBLA from the point of "evidence-centered assessment design" (p. 477). They describe the main components of this framework as "student model(s)", "evidence models", "task models", and "assembly model" (p. 481). Another part of the research studies puts accent on the difficulties of task-based implementation. As a basis for TBLA, task-based language teaching (TBLT) itself may have some challenges in task administration. Carless (2002) discusses some issues in the implementation of TBLT, such as "noise and indiscipline", "use of the mother tongue", and "pupil involvement in tasks" (pp. 390-394). In a different study, Carless (2003) elaborates on the teaching-related factors in the use of communicative tasks in primary education, such as teacher understanding, allocated time, and use of textbooks. Likewise, some research underlines the difficulties of TBLA implementation. To illustrate, TBLA is mostly based on the restricted sample of language outcomes, and it is specifically designed in accordance with the assessment needs of the subject group; for these reasons, it may lack generalization (Bachman, 2002). In addition, the construct of task difficulty may need further conceptualization in terms of task characteristics and learners' language performance (Bachman, 2002). In TBLA, there are some concerns about the rater reliability since the raters' judgments may influence the assessment (Brindley, 1994). In addition, it is relatively more expensive and time-consuming than other kinds of assessment with regard to task construction and implementation (Bachman, 2002; Brindley, 1994). As the implementers of TBLA, language teachers may need training about assessing language learners through tasks as there has been a shift from traditional assessment practices to task activities (Brindley, 1994).

Considering the young learners' context, Johnstone (2000) touches upon the embeddedness problem and argues that an assessment task can be valid in the primary learning setting when the pupils are familiar with it in the learning process. In a similar way, Linse (2005) puts forward that language assessment "should be integrated and reflect the type of instruction" (p. 145). Accordingly, this study attempts to develop and suggest an integrated language testing (ILTPY) program for the 4th grade young learner classes in order to overcome the embeddedness problem and integrate language assessment into the instruction. Besides that, this research

aims to implement this suggested program and define its effectiveness in terms of attainment, achievement, motivation, assessment needs and washback effect. The following researh questions guide the data collection and analysis procedures related to the implementation of ILTPY:

1. To what extent does ILTPY promote young learners' language attainment and achievement?
2. How does ILTPY affect the motivation of the stakeholders in YLLs' classrooms?
3. To what extent does ILTPY meet YLLs' assessment needs?
4. Does ILTPY have an impact on learning English in YLLs' classrooms?

3. Method

3.1. Research Design

Mixed-method research methodology has been used in this study. This design provides researchers with "a more complete understanding of research problems" by enabling them to make sense of both qualitative and quantitative data (Fraenkel, Wallen, & Hyun, 2012, p. 557). On the basis of mixed-methods, the quantitative data and qualitative data were collected separately. For quantitative data, unit tasks were employed at the end of the two-unit implementation. As for the qualitative data, interviews were conducted with both EFL teachers and YLLs. Data collection process will be discussed in detail in section 3.3.

3.2. Participants and Setting

The subject group includes a group of the 4th grade YLLs (n=73) and EFL teachers (n=4) at state primary schools in Ankara. Young learners, 38 female and 35 male participants, were aged between 9 and 11 years. EFL teachers were experienced in English language teaching (ELT) from 8 to 30 years. While three teachers were female, one teacher was male. These participants were selected from three primary schools with convenience sampling method. Convenience sampling presents researchers with a practical way to reach the subject group (Cohen, Manion, & Marrison, 2007). Concerning the instructional setting in the 4th grade, YLLs have to attend two-hour English courses per week (MoNE, 2017). ELT curriculum devised by MoNE holistically shapes teaching and assessment practices in this grade.

3.3. Data Collection

In this research, qualitative and quantitative data were collected from different sources. First, assessment tasks in two sequential units of ILTPY were used to measure to what extent young learners attained the language objectives in two units. After the implementation of the assessment tasks, follow-up interviews were conducted to understand the effectiveness of suggested testing program in terms of

the motivation, assessment needs and washback. ILTPY and main components are illustrated as follows:

3.3.1. ILTPY

Integrative Language Testing Program for the 4th Grade Young EFL Learners (ILTPY) was originally developed for this study with reference to the research problems and aims. It is theoretically constructed within the framework of TBLA. This program proposes a language assessment model for young learners in order to reflect on what they have learnt in an authentic and communicative setting. ILTPY is embedded in the flow of language learning in the 4th grade classrooms. This enables assessment to be regarded as the part and parcel of language teaching. In this respect, this program has four key components: language skills and objectives, communicative functions and language use, construct definitions, and assessment tasks (see Appendix B). These components are explained below:

3.3.1.1. Language Skills and Objectives

Instructional objectives are defined as "specific statements that describe the particular knowledge, behaviors, and/or skills that the learner will be expected to know or perform at the end of a course or program" (Brown, 1995, p. 73). In other words, they describe what kind of performance is expected at the end of learning processes. In keeping with this definition, ILTPY is based on the achievement of language skills and objectives, which are originally presented in ELT curriculum developed by MoNE (2013). The level of the target group in this study (the 4th grade) represents A1 level (breakthrough) in the Common European Framework for References (CEFR) (Council of Europe [CoE], 2001). In consideration of A1 level in the CEFR, ELT curriculum proposes speaking and listening along with very limited writing and reading as the main language skills for the 4th grade classes (MoNE, 2013), and language objectives are formulated with respect to these skills. In ILTPY, the curricular objectives perform an active role in the operationalization of the constructs, development of assessment tasks, and conceptualization of task characteristics.

3.3.1.2. Communicative Functions and Language Use

ILTPY attaches importance to the communicative aspects of language assessment by offering young learners to interact in the target language through language use tasks. At this point, this component includes the basic communicative functions and language use which ELT curriculum (MoNE, 2013) covers. Since the communicative functions and language use are influential in selecting content and organizing the learning process, ILTPY puts emphasis on the utilization of these functions in assessment to get feedback on language learning and sharpen the understanding of what communicative language objectives have been attained.

3.3.1.3. Construct Definitions

In language assessment context, an assessment task response should be built upon "construct definition" which refers to "making an inference to an observable variable"

(Fulcher & Davidson, 2007, p. 78). According to Bachman and Palmer (1996), a construct definition "provides a basis for the development, in the operationalization stage, of test tasks" (p. 89). From this point of view, 36 task constructs were operationalized for the development of the assessment tasks in ILTPY through the adaptation of language learning objectives, curricular outcomes, and communicative functions in ELT curriculum published by MoNE in 2013. These constructs are in attempt to describe the observable attributes in mostly speaking and listening skills according to the cognitive and linguistic level of the subject group. All task components in ILTPY were designed in line with those 36 constructs with the aim of ensuring construct validity.

3.3.1.4. Assessment Tasks

Young learners "try to find a meaning and purpose for activities that are presented to them" in language classrooms (Cameron, 2001, p. 21). It means that they have a dynamic role in making sense of their own learning environment. For the construction of meaningful learning, language tasks are the effective tools since they "call for primarily meaning-focused language use" (Ellis, 2003, p. 3). According to Cameron (2001) and McKay (2006), language learning and teaching tasks may be used as the assessment practices in YLLs' classrooms. To put it differently, young learners may be assessed through the language use tasks with which they are familiar in the learning process. In this sense, the assessment tasks in ILTPY are similar to learning and teaching activities in this grade. Broadly speaking, ILTPY includes 36 speaking and listening tasks (see Appendix C and D for sample tasks) which were developed in accordance with the language objectives, communicative functions, and language use highlighted in ELT curriculum published by MoNE in 2013. These assessment tasks have an overall purpose to integrate language assessment into communicative learning setting in the 4th grade. They also aim at an anxiety-free assessment environment for YLLs (Hasselgreen, 2005). During the design stage of the 4th grade integrative assessment tasks, the framework of task characteristics (Bachman & Palmer, 1996; McKay, 2006) was taken into consideration to develop task specifications. Figure 1 shows how the assessment tasks have been developed in ILTPY. The basic components in Figure 1 were adapted from the task characteristics (*i.e. characteristics of setting, assessment activity procedure, input, expected response and relationship between input and response*) presented in the framework (Bachman & Palmer, 1996; McKay, 2006).
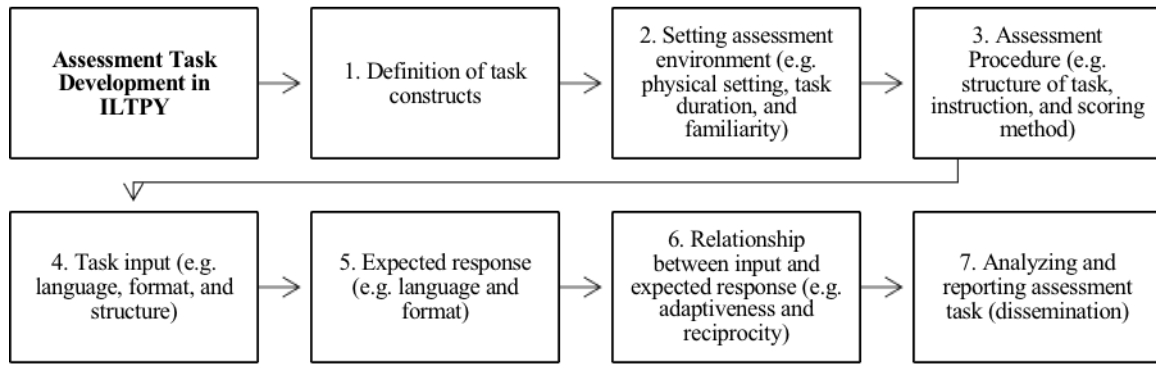
Figure 1: Development process of assessment tasks in ILTPY

At the end of the development process, sample assessment tasks were piloted to ensure reliability and validity. In light of the piloting results, some modifications were carried out in relation to the task administration. Assessment tasks were revised in order to minimize the guessing factors, take control of the first language use, and manage the classroom interaction. After the revision, the two-unit implementation of ILTPY was conducted in the subject group. In this implementation, Unit 2 (Children's Day) and Unit 3 (Free Time) assessment tasks* were administered in connection with the 4th grade language teaching program and then scored with task-specific rubrics group (see Table 1 for the task characteristics and construct definitions). Task-specific rubric was developed for each task in consideration of task specifications. Following the scoring process, task results were explained to YLLs through the colorful cards showing the level of their performance with symbolic stars. Symbolic stars ranging from 1 to 5 indicate the levels of task performance respectively as "Needs improvement, Developing, Average, Satisfactory, and Very Satisfactory". These levels were generated in line with the description levels of sample rubrics (Brookhart, 2013; O'Malley & Pierce, 1996) and official pre-school, primary and secondary school regulations (MoNE, 2014). In addition to this report, their task sheets were distributed to young learners, and they had an opportunity to examine their learning strengths and weaknesses.

Table 1: The Characteristics and Construct Definitions of the Assessment Tasks

UNIT 2: Children's Day	
Task 5: MATCHING TASK "Countries and Nations"	*Construct: Recognizing personal information about other people such as where they are from and where they live
Task 6: TPR Activity "(Chart-filling)" Find Someone Who….	*Construct: Asking and answering the questions about people, their nationalities and home countries if articulated slowly and clearly

* In revised ELT Curriculum (MoNE, 2017), the 4th grade units were rearranged. Accordingly, Unit 2 was titled "Nationality", and "Free Time" was included in the curriculum as Unit 4.

Task 7: INFORMATION GAP "Where is s/he from?"	*Construct: Recognizing personal information about other people such as where they are from and where they live, Identifying different people's personal possessions
UNIT 3: Free Time	
Task 8: MULTIPLE CHOICE "Keloğlan likes…" & "Keloğlan doesn't like…"	*Construct: Understanding the gist and key information in short, recorded passages of people talking about their likes and dislikes
Task 9: ORAL INTERVIEW "Survey: Likes and Dislikes in the Classroom"	*Construct: Asking people questions about what they like and answering such questions addressed to them provided they are articulated slowly and clearly
Task 10: INFORMATION GAP "Who is he/she?"	*Construct: Saying what they like doing

*These constructs were adapted from the curricular objectives devised by MoNE (2013).

3.3.2. Follow-up Interviews

In order to collect qualitative data on the effectiveness of the two-unit implementation of ILTPY, follow-up interviews were conducted with young learners in written and EFL teachers through note-taking. Both interviews were performed in the participants' native language (Turkish) in order to minimize the impact of foreign language "concerns" on data collection (Mackey & Gass, 2005, p. 174). The interview questions were developed in accordance with the study aims, and five experts examined the questions in terms of the content validity. According to the expert opinions, interview questions were revised, and prepared for the follow-up administration (see Appendix A).

3.4. Data Analysis

In this study, qualitative and quantitative data were separately analyzed by means of different data analysis techniques. Qualitative data were analyzed through content analysis which refers to "reading the data over a variety of times, looking for key ideas or topics and labelling these ideas by marginal notes or post-its" (McKay, 2006, p. 57). Dörnyei (2007) emphasizes the four steps in the analytic process of the content analysis as "(a) transcribing the data, (b) pre-coding and coding, (c) growing ideas-memos, vignettes, profiles, and other forms of data display, and (d) interpreting the data and drawing conclusions." (p. 246). To follow these steps, the written qualitative data were reviewed and translated at first, and then coded and categorized according to the themes emerged in the research questions. As for the quantitative data analysis, descriptive statistics like mean and standard deviation were calculated in SPSS 21. In order to investigate whether there was a statistically significant difference between the two-unit assessment tasks, non-parametric test, *Wilcoxon Signed Rank Test*, was utilized since data sets were not normally distributed.

4. Results and Discussion

In this section, the study findings are presented and discussed under the titles of four research questions. The research questions focus attention on the language attainment and achievement, stakeholders' motivation, assessment needs and washback.

4.1. To what extent does ILTPY promote young learners' language attainment and achievement?

The descriptive statistics of the task scores given in Table 2 show to what degree YLLs have attained the unit objectives. As given in Table 2, the mean score in Unit 3 (*M*=87) is higher than the mean score of Unit 2 (*M*=80). Considering the unit scores as the indicator of the attainment and achievement level, it may be stated that YLLs have a high level of language attainment in Unit 3.

Table 2. Descriptive Statistics of Assessment Tasks

	Mean	N	Std. Deviation	Std. Error Mean	Minimum	Maximum
Unit 2	79.93	73	17.46	2.04	27.50	100
Unit 3	86.64	73	12.46	1.46	50.00	100

According to the mean scores, YLLs have higher scores in Unit 3. It shows that there is an increase in the assessment task scores in Unit 3. That is to say, there is a difference between the assessment scores in two separate units. Table 3 indicates whether this difference is statistically significant or not. As explained in Table 3, the difference between the assessment scores in Unit 2 and Unit 3 is statistically significant (Z=-3.634, p<0.05). The statistical findings point out that the students have higher achievement in Unit 3. Put it another way, their level of achievement increased at the end of two-unit implementation.

Table 3. Wilcoxon Signed Rank Test Result

		N	Mean Rank	Sum of Ranks	z	p
Unit 3 – Unit 2	Negative Ranks	20	22.95	459.00	-3.634	.000
	Positive Ranks	42	35.57	1494.00		
	Ties	11				
	Total	73				

According to Norris, Brown, Hudson, and Yoshioka (1998), "task-based testing involves the assessment of task outcomes in terms of learner/examinee success or failure to accomplish the task" (p. 53). With reference to this perspective, it can be stated that YLLs had higher scores in Unit 3, and it shows that ILTPY unit implementation may promote the attainment and achievement of language learning in the 4th grade. Also, Norris (2009) expresses that embedded assessment tasks explain to what extent learners have learnt the content. From this formative aspect, ILTPY may give feedback on the progress of language learning in young learner classrooms.

4.2. How does ILTPY affect the motivation of the stakeholders in YLLs' classrooms?

In the follow-up interviews, a majority of YLLs expressed that assessment tasks in ILTPY helped them feel motivated to learn English. There are some examples of their responses below:

S2: *Yes, because my enthusiasm increases when I get 5 stars.*

S32: *They increased because I have learnt what I do not know.*

S48: *Yes, when I learnt what I did not know, I wanted that activity to be implemented more, and I liked that unit very much.*

S71: *They increased my motivation [because] I answered the questions more quickly.*

As it is seen from the learner responses, assessment tasks provide learners with opportunities to strengthen language learning. Also, they motivate YLLs by providing reinforcement. When YLLs feel motivated to learn English, they have more chances to get involved in their own language learning. According to McKay (2006), language assessment in YLLs' classes should motivate the pupils. Regarding the student responses, it can be argued that assessment tasks in ILTPY mostly motivate young learners to learn English.

As for the teacher responses, it may be concluded that most of the language teachers are motivated for language teaching at the end of the two-unit implementation. There are various triggers for their motivation, such as the professional development, students' in-class participation, willingness and motivation. More specifically, Teacher 1 and Teacher 2 explained that student motivation had an impact on their teaching motivation, and therefore integrated assessment tasks affected their motivation in a positive way. In a similar vein, Teacher 4 shared that those tasks could motivate teachers professionally and make teaching process more useful.

T1: *Namely, it is expected that they affect a teacher's motivation in a good way. Interaction between teacher and student makes a considerable contribution to their exams in terms of helping them have good listening. The more the pupil learns, the more teacher motivation increases.*

T2: *I think that they will affect positively. They affected me positively since they increased the students' willingness to participate [in the activities].*

T4: *Sure, they affect us, too. They increase teacher's motivation in terms of teacher's professional development. Yes, they did. I wish I could teach the students in that way. I believe that I will be more helpful in this way.*

However, Teacher 3 stated that there were some difficulties in the administration and evaluation of the assessment tasks even though those implemented in the classroom were the activities to be used in order to conduct effective assessment and evaluation. In addition, he/she stressed that the implementer of those tasks was not himself/herself; so, his/her motivation was not positively affected as follows:

T3: *I want to say that teachers may not want to implement those tasks because it is difficult to implement and evaluate them, but they are activities to be employed for more effective assessment and evaluation. As you implemented them, they did not affect me but if I had implemented, my motivation would have been affected negatively.*

Carless (2002; 2003) elaborates on the issues in task-based language teaching ranging from the classroom setting to instructional factors. Since TBLA is theoretically shaped by TBLT (Shehadeh, 2012), there may be similar problems with which language teachers and students may confront in the assessment task administration. For example, Bachman (2002) touches on the issues of the generalizability and task difficulty in TBLA. What's more, Brindley (1994) calls attention to the inter-rater reliability and teacher training in TBLA. Also, the practicality of task-based assessment is regarded as one of the difficulties (Bachman, 2002; Brindley, 1994). All these problems in task-based practices may have a negative impact on teacher motivation and performance as stated by Teacher 3.

4.3. To what extent does ILTPY meet YLLs' assessment needs?

After the implementation stage of ILTPY, teachers evaluated whether the integrated tasks met the learners' needs of assessment or not. It can be deduced from the EFL teacher responses that the two-unit implementation largely met the students' needs; nevertheless, there were some points to be taken into consideration for more effective assessment and evaluation practices. In particular, Teacher 3 and Teacher 4 explained that those tasks were interesting and appropriate to their age level; that is why the two-unit tasks mostly catered for the students' needs of language assessment.

T3: *They meet the students' expectations largely because they do not like classical exams. They like the activities which they take part in actively. They met my expectations but helped me see the difficulties in implementation.*

T4: *They met [the assessment needs] largely. Materials [and] questions are appropriate to the level of young learners. Young learners have difficulties in the exams carried out with the classical methods. The activities which you implemented attract their attention more. Making effort in the classroom [and] asking questions to each other were appreciated by the teacher.*

In other words, two-unit implementation promoted the classroom interaction, and young learners enjoyed the activities in which they actively participated. Carr (2011) underlines the importance of the assessment process, which has been developed in accordance with the young learners' age and characteristics. Similarly, Cameron (2001) puts emphasis on the age factor which makes young learner assessment different from other levels. Also, Hughes (2003) discusses that language assessment in young learner classrooms should employ cognitively appropriate materials and offer social interaction. Taken together, it can be elucidated that ILTPY largely meets YLLs' assessment needs by presenting the students with familiar assessment setting and age-appropriate assessment materials, and facilitating the student involvement

and interaction. On the other hand, Teacher 1 and Teacher 2 mentioned the need for more visual aids and technology in order to assess the skill of listening.

> T1: *Listening is good for the students. Besides that, there should be visual activities. If technological devices are utilized, what is taught becomes more permanent in students' memory. They met [the assessment needs] but they would have met my expectations more if there had been visuals, too.*

> T2: *They considerably meet [the assessment needs] because I think that they meet the students' need of learning with fun. Even so, there are some constraints. There are technological limitations in listening.*

Simply put, even if ILTPY meets the assessment needs of YLLs to a large extent, it should be supported with more visual and technological tools especially in assessing listening more effectively.

4.4. Does ILTPY have an impact on learning English in YLLs' classrooms?

Washback effect is defined as the impact of testing on language learning, and this effect can be either positive or negative (Hughes, 2003). In this study, teachers elucidated the impact of integrated language assessment on young learners' language learning. Teacher responses point out that ILTPY provides young learners with a comfortable learning setting, helps them take an active role in learning, promotes their motivation and attention, and reinforces their language learning. So, it generally influences language learning in a positive way. More specific examples are presented as follows:

> T2: *I think that they generally have a positive impact on the students' learning. I think they paid more attention [to the tasks] because they were the activities which the students took actively part in.*

> T3: *Yes, they did. First of all, their attention and motivation in the lesson increased in line with the activities implemented. The reason for the increase in their motivation is that they actively take part in the activities, [they] don't have to [attend] the lesson in a boring way, and they participate in a comfortable learning setting by acting freely and listening to the songs in the classroom…*

> T1: *They affected positively. I mean, the results which they had, their [language] learning, and the fact that this process makes them pleased may get them to like English.*

It can be understood that young learners may feel more pleased to learn English when they consider their learning gains. Wigglesworth and Frost (2017) explain that task-based assessment provides positive washback effect on language learning. According to the teacher responses, young learners actively participated in ILTPY tasks, and it can be interpreted from three teachers' responses that integrated language assessment tasks might affect the students' learning process positively. However, Teacher 4 stated that the effect of the two-unit implementation on language learning could be only exerted in a longer term, and allocated time for two-unit tasks was not enough to observe the long-term effect of the integrated assessment program. It means that more instructional time is required to implement the assessment tasks representatively and understand the washback effect.

T4: *In order to observe the impact of the activities on learning, more time is necessary, and they should be implemented in more units.*

In summary, two-unit implementation of ILTPY brought about an effect on YLLs' language learning by mainly promoting in-class participation and offering the students an authentic assessment and learning setting. Nonetheless, all the assessment tasks developed for ten curricular units in the 4th grade should be implemented in order to see the long-term effect of ILTPY on YLLs' language learning.

5. Conclusion and Implications

Task-based assessment has gained importance in the field of language testing with the increasing emphasis on the communicative aspect of language testing and performance-based assessment practices in the real-life contexts. The use of language tasks as the assessment tools in young learner classrooms generates the evaluative information on the pupils' authentic language performance (McKay, 2006). Hence, this study aimed to measure YLLs' language performance through the assessment tasks in ILTPY, which is a congruent program specifically developed for the 4th grade language learners with respect to the curricular objectives in ELT curriculum devised by MoNE (2013). The study findings showed that the two-unit implementation of ILTPY significantly promoted the level of language achievement and attainment in the 4th grade. Since ILTPY presents the students with an opportunity to strengthen what they have learnt, it can be generally concluded that the suggested two-unit implementation has a motivating effect on language learning. Besides, YLLs' increasing motivation, participation and interest in language activities have also promoted EFL teachers' motivation. However, the demanding process in the development and administration of the assessment tasks may be regarded as the demotivating dimension of ILTPY. Regarding the characteristics of YLLs, it can be deduced from the interviews that ILTPY largely meets the assessment needs of the pupils by taking the curricular specifications of assessment into consideration, presenting age-appropriate materials, and encouraging the active involvement in the classroom. Nevertheless, it should be supported with more visuals and advanced technological devices.

Concerning the washback effect, it can be stated that the two-unit implementation of ILTPY generally influences language learning in a positive way. However, it is required to administer all assessment tasks in other units to inspect the overall impact of this integrated program on the pupils' language performance. This can be considered as one of the limitations of this research. Since ILTPY includes ten-unit assessment schedule, and the implementation of the whole program will take two academic semesters (i.e. almost nine months), the whole program could not be implemented in this research. Only two-unit tasks were conducted in the 4th grade because of the time limitation. Another limitation of the study is the use of technology

during the task administration. All materials used in the implementation of the assessment tasks were additionally brought into the classroom. Due to the practicality issues, advanced technological devices could not be employed in the assessment setting. For that reason, it is stated that ILTPY needs improvement in terms of the technology use in listening assessment. Final limitation of the present study is the research sample. In order to generalize the study findings, it is required to implement the suggested program in a larger sample. Since such an implementation is not practically applicable due to the limitations described above, the results can be discussed within the range of convenience sampling employed.

For further research, task-based assessment of language skills may be embedded in the language learning processes at other levels (e.g. secondary and tertiary education). More particularly, the assessment tasks may be developed in accordance with the specific needs of the language learners, and these tasks may be integrated into the language instruction with formative purposes. Moreover, the usefulness of the assessment tasks can be examined from the perspectives of language test qualities proposed by Bachman and Palmer (1996). As another study, the washback effect of TBLA can be investigated in different contexts of teaching English as a foreign language.

References

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, *19*(4), 453-476.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice.* Oxford: Oxford University Press.

Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. USA: Heinle & Heinle Publishers.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment.* New York: Mc Graw Hill.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, *32*(4), 653-675.

Brindley, G. (1994). Task-centred assessment in language learning: The promise and the challenge. In B. Norman (Ed.), *The annual international language in education conference* (pp. 73-94). Hong Kong.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. USA: ASCD.

Byrnes, H. (2002). The role of task and task-based assessment in a content-oriented collegiate foreign language curriculum. *Language Testing*, *19*(4), 419-437.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1-47.

Cameron, L. (2001). *Teaching languages to young learners.* UK: Cambridge University Press.

Carless, D. (2002). Implementing task-based learning with young learners. *ELT Journal*, *56*(4), 389-396.

Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, *31*(4), 485-500.

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford: Oxford University Press.

Cohen, L., & Manion, L., Marrison, K. (2007). *Research methods in education (6th Ed.)*. New York: Routledge.

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

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. New York: Oxford University Press.

Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer?. *Language Testing*, *19*(4), 347-368.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education (8th Ed.)*. New York: Mc Graw Hill.

Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York: Routledge.

Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing*, *22*(3), 337-354.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Linse, C. (2005). *Practical English language teaching (PELT): Young learners*. New York: McGraw Hill.

Johnstone, R. (2000). Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience. *Language Testing*, *17*(2), 123-143.

Ke, C. (2006). A model of formative task-based language assessment for Chinese as a foreign language. *Language Assessment Quarterly*, *3*(2), 207-227.

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. New Jersey: Lawrence Erlbaum Associates Publisher.

Madsen, H. S. (1983). *Techniques in language testing*. Oxford: Oxford University Press.

McKay, S. L. (2006). *Researching second language classrooms*. London: Lawrence Erlbaum Associates Publisher.

McKay, P. (2006). *Assessing young language learners*. UK: Cambridge University Press.

McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, *19*(4), 477-496.

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 578-594).

Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing*, *19*(4), 395-418.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. USA: National Foreign Language Resource Center.

Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.

O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Boston: Addison-Wesley Publishing Company.

Robinson, P., & Ross, S. (1996). The development of task-based assessment in English for academic purposes programs. *Applied Linguistics*, *17*(4), 455-476.

Shohamy, E. (1997). Second language assessment. In G. R. Tucker & D. Corson (Eds.), *Encyclopedia of language and education (Volume 4)* (pp. 141-149). USA: Springer.

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, *18*(4), 463-481.

Shehadeh, A. (2012). Task-based language assessment: Components, development, and implementation. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.) *The Cambridge guide to second language assessment* (pp. 156-163). Cambridge University Press.

T.C. Milli Eğitim Bakanlığı [MoNE] (2013). *English language teaching program (for the 2nd, 3rd, 4th, 5th, 6th, 7th and 8th grade)*. Ankara.

T.C. Milli Eğitim Bakanlığı [MoNE] (2014). *Pre-school, primary and secondary school regulations. T.C. Resmi Gazete,* 29072, 26 July 2014.

T.C. Milli Eğitim Bakanlığı [MoNE] (2017). *English language teaching program (for the 2nd, 3rd, 4th, 5th, 6th, 7th and 8th grade)* [revised]. Ankara.

Wigglesworth, G. (2008). Task and performance-based assessment. In N. H. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 111-122). Springer Science+Business Media, LLC.

Wigglesworth, G., & Frost, K. (2017). Task and performance-based assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment*, (pp. 121-133). Springer International Publishing.

Weir, C. J. (2005). *Language testing and validation.* UK: Macmillan.

Appendix A. Follow-up Interview Questions

A.1. Follow-Up Interview Questions

A.1.1. Student Follow-up Interview Questions

1. Do you think that these activities, which were used in your classrooms to assess and evaluate your language performance, are useful? Please explain the reason(s).

2. Did you get information about your strengths and weaknesses from your report cards given after these implementations? Please explain.

3. Did these activities increase your motivation for learning English? If yes, please explain the reason.

4. Do you think that these activities have an impact (positive or negative) on your learning English? Please explain the reason(s).

5. If you have additional opinions related to these activities, please share them.

A.1.2. Teacher Follow-up Interview Questions

1. Do you think that the assessment and evaluation activities, which were administered in your classrooms, are useful in terms of teaching English? Please explain the reason(s).

2. In your opinion, how do these activities affect teachers' motivation? Do you think that they have an impact on your motivation? Please explain.

3. In your opinion, did these activities facilitate the 4th grade students' attainment of language objectives? Please explain the reasons(s).

4. Do you think that these additional activities have met your and your students' assessment and evaluation needs and expectations? Please explain your answer with the reason(s).

5. Do you think that these activities have an impact (positive or negative) on learning English in general? Please, explain the reason(s).

6. If you have additional opinions related to these activities, please share them.

Appendix B. Integrated Language Testing Program (Outline) for the 4th Grade Young EFL Learners

A.2. Outline of Testing Program

UNIT: .... (This part defines the title of the unit in the curriculum)			
Language Skills & Objectives*	**Communicative Functions & Language Use***	**Assessment Tasks**	**Construct Definition ****
This column gives information about English language learning objectives. These statements of learning objectives were directly taken from the curriculum developed by MoNE (2013). Assessment tasks, methods and scoring rubrics were designed in line with these objectives.	This column gives information about the communicative functions and skills. It is aimed to assess these functions and skills in the integrated language testing program. These functions, skills, and related examples were directly taken from English language teaching curriculum developed by MoNE (2013). Tasks mainly aim to assess these communicative elements in language learning setting.	This column describes the assessment tasks and methods integrated with language learning. Language assessment tasks were developed in accordance with the curriculum objectives, learning activities, and pedagogical needs. The main structural schemes for designing assessment tasks and methods were adapted from Bachman and Palmer (1996), O'Malley and Valdez Pierce (1996), Cameron (2001) and McKay (2006).	This column gives the operational definitions of constructs which give shape to the construction of the assessment tasks. Each task is developed with regard to the constructs based on the objectives given in English language teaching curriculum developed by MoNE (2013).

*These headings were directly taken from ELT curriculum devised by MoNE (2013).

**This part was adapted from the objectives given in ELT curriculum devised by MoNE (2013).

Appendix C. Assessment Task in Unit 2

ASSESSMENT TASK – 6*

Title of Assessment Task:	"FIND SOMEONE WHO…" TPR Activity (Chart-filling)
**Objective of Assessment Task:	Students will be able to ask and answer the questions about people, their nationalities and home countries if articulated slowly and clearly.
***Construct Definition:	Asking and answering the questions about people, their nationalities and home countries if articulated slowly and clearly
Characteristics of Learners:	Grade 4, beginning EFL learners (A1), aged 9-10
Learning Context:	The assessment task is integrated into teaching activities in which they ask and answer the questions about the countries and nationalities.
Resources:	Mini Flashcards (Countries and Nationalities) Chart-filling Cards Task Specific Rubric
Characteristics of Setting: - **Physical Environment** - **Time of Task** - **Participants**	The task will be performed in the classroom setting. Physical environment is familiar. It is a whole-class activity in which all students and EFL teacher will get involved. Time for the task will be 10 minutes.
Characteristics of The Assessment Task Procedures: - **Instructions** - **Structure** - **Time Allotment** - **Scoring Method**	EFL teacher will give oral instructions in the target language and simplify the task by providing examples. There is one part in the assessment task. First of all, teacher will distribute the nationality cards to the students. Then they will ask and answer the questions according to these cards. Their task performance will be observed and then scored with task specific rubric.
Characteristics of The Input: - **Format** - **The Language of the Input**	The input of the assessment task will be given in both non-language form (including visual cards) and language form (written phrases). The length of the task is limited to the simple phrases for each person. Since it is similar to puzzle activity, the task will raise the interest of the learners. The language of input includes organizational knowledge in the target language. It is made up of single words and phrases in addition to cultural references.
Characteristics of The Expected Response: - **Format** - **The Language of the Expected Response**	Format of the expected response will be both spoken and written. First, the learners are expected to draw meanings from the context (pictures) and then answer the questions related to them. They are to write the answers down correctly. The expected response includes simple sentences and phrases. Also, it is appropriate to the learners' cognitive and linguistic level.
Relationship between Input and Response	Assessment task is non-reciprocal. Teacher will not give immediate feedback. The task response requires drawing meanings from the visual context with respect to their prior topical knowledge on the basis of the input processing.
Analyzing & Reporting Assessment Process	Chart-filling cards will be examined after the lesson, and learners' performance will be scored in view of task-specific rubric. The rubric results (reports) will be given to EFL learners and discussed at the end-of-the-unit conference.

*Task characteristics were adapted from Bachman and Palmer (1996) and McKay (2006).
**The objectives were directly taken from the curriculum devised by MoNE (2013).
***The construct definitions were adapted from the objectives given in the curriculum developed by MoNE (2013).

Appendix D. Assessment Task in Unit 3

ASSESSMENT TASK – 8*	
Title of Assessment Task:	"KELOĞLAN LIKES …" & "KELOĞLAN DOESN'T LIKE …" (Multiple Choice)
**Objective of Assessment Task:	Students will be able to understand the gist and key information in short, recorded passages of people talking about their likes and dislikes.
***Construct Definition:	Understanding the gist and key information in short, recorded passages of people talking about their likes and dislikes
Characteristics of Learners:	Grade 4, beginning EFL learners (A1), aged 9-10
Learning Context:	The assessment task is integrated into teaching activities in which they listen to the recorded passages about likes and dislikes.
Resources:	Multiple Choice Sheets Audio-player Task Specific Rubric
Characteristics of Setting: - **Physical Environment** - **Time of Task** - **Participants**	The task will be performed in the classroom. Physical environment is familiar to the students. It is a whole-class activity in which all the students and EFL teacher will get involved. Time for the task will be 3+3 (6) minutes.
Characteristics of The Assessment Task Procedures: - **Instructions** - **Structure** - **Time Allotment** - **Scoring Method**	EFL teacher will give oral instructions in the target language. There are two parts (Comprehension & Performance) in the assessment task. Firstly, learners will listen to the audio-record for general comprehension and then choose the correct pictures describing Keloğlan's likes and dislikes. They have 3 + 3 minutes sequentially for these parts. Learners' task performance will be graded with task specific rubric.
Characteristics of The Input: - **Format** - **The Language of the Input**	The input of the assessment task will be given in the target language. It will be appropriate to the cognitive maturity. Also, the task will be supported with EFL teacher's explanations. The length of the task is limited to single simple sentences. The language of input includes organizational and topical knowledge.
Characteristics of The Expected Response: - **Format** - **The Language of the Expected Response**	Format of the expected response will be visual (in non-language form). The students are expected to choose the most appropriate picture describing Keloğlan's likes or dislikes. It is limited to the pictures describing free time activities. It requires topical knowledge.
Relationship between Input and Response	Assessment task is non-reciprocal. Teacher will not give immediate feedback. The expected response is primarily based on input processing and drawing meaning from the context.
Analyzing & Reporting Assessment Process	Multiple choice sheets will be examined after the lesson, and learners' performance will be graded in view of task specific rubric. The rubric results (reports) will be given to EFL learners and discussed at the end-of-the-unit conference.

*Task characteristics were adapted from Bachman and Palmer (1996) and McKay (2006).
**The objectives were directly taken from the curriculum devised by MoNE (2013).
***The construct definitions were adapted from the objectives given in the curriculum developed by MoNE (2013).