A case study of a Hungarian EFL teacher's assessment practices with her young learners

Gabriella Hild

University of Pécs, Hungary
gabriella.hild@aok.pte.hu

The case study aims to provide insights into how a Hungarian EFL teacher used tests, assessed her young learners and gave feedback to them in the classroom. This qualitative, exploratory study was a follow-up to a large-scale project. In this single-case study, data were collected from an EFL teacher and five of her seventh graders on what tasks she used to assess them and how. The participants were interviewed. For the purpose of triangulation, the students were also audio- and video-recorded while doing four speaking tasks, and two classes were observed. The results revealed that for the teacher with decades of teaching experience there was room for improvement in her knowledge of age-appropriate teaching methodology and that some of her beliefs and practices reflected a lack of understanding how children develop. She had difficulty diagnosing her students' strengths and weaknesses. The learners were rarely provided with feedback on their performance and language development; therefore, they did not see how much they had progressed. Low achievers had a hard time catching up with their peers; and they lagged further behind. The teacher seemed to be more interested in what her students did not know rather than focusing on what they could do.

Keywords: early language learning; assessment of young language learners; diagnostic assessment; dynamic testing; early foreign language teaching in Hungary

1. Introduction

The increasing popularity of starting to learn foreign languages (henceforth FLs) at an early age is a world-wide tendency (de Bot, 2014; Garton, Copland, & Burns, 2011; Nikolov, 2009a, 2009b, 2016a; Nikolov & Mihaljevic Djigunovic, 2006, 2011). The significance of effective language learning in primary school is crucial since "the foundations for later language learning are laid" in early language programs (Commission of the European Communities, 2003, p. 7). Although common wisdom suggests that children behave like sponges when it comes to learning a new language, success in early language learning is not automatic. Young learners' special characteristics and needs are to be considered and catered for to ensure that the benefits of an early start are realized (Johnstone, 2009; McKay, 2006, p. 5). In an exploratory study examining Hungarian early-start programs, Nikolov (as cited in Medgyes & Nikolov, 2014, p. 516) highlighted the importance of meeting these special requirements by claiming that since the findings on the language teaching methodologies and practices were "devastating," the pupils would have been better off if they had not learned a FL at all.

The process of language development is slower in young learners than in their older peers; therefore, generating and maintaining motivation is crucial for positive outcomes in early language programs (Nikolov, 2011). It has been well established that children's motivation is largely shaped by classroom experience, the teacher, and later by peers (Nikolov, 1999b). Providing young learners with the feeling of success and showing them how much they have progressed are essential means of maintaining motivation at early ages (McKay, 2006, p. 14). In order for young learners to see that they are making progress, teachers should offer detailed feedback and inform them on their strengths and weaknesses on a regular basis. Two recent trends in educational assessment, diagnostic and dynamic testing, can assist language teachers in accomplishing these requirements and, consequently, in boosting their students' learning potential (Nikolov & Szabó, 2012). Unlike in traditional testing procedures in which the aim is to assign grades to performances, in diagnostic assessment the focus is on identifying students' strengths and weaknesses, and then, on the basis of the results, improving their abilities.

Dynamic assessment integrates assessment with teaching (Lantolf & Poehner, 2006, p. 6). Unlike in traditional settings, where teachers take the position of standing back and listening, here, they collaborate with and support students (Poehner, 2008, p. 15; Sternberg & Grigorenko, 2002, p. 29). In order for teachers to be able to apply dynamic assessment, they need to be equipped with "diagnostic competence" (Edelenbos & Kubanek-German, 2004, p. 260). Diagnostic competence is "the ability to interpret students' FL growth, to skillfully

deal with assessment material and to provide students with appropriate help in response to this diagnosis" (Edelenbos & Kubanek-German, 2004, p. 260).

2. Literature review

Research on how teachers assess young language learners (henceforth YLLs) is fairly limited (Nikolov & Mihaljević Djigunović, 2011; Rea-Dickins, 2004). In a case study investigating teachers' performance during assessment in the classroom of English as an additional language in England, Rea-Dickins and Gardner (2000) revealed a lack of systematicity in assessment practices. The reliability of the information the teachers elicited through observation-driven assessment was questionable, which, then, threatened the validity of the inferences they drew for the individual students. The findings also shed light on the teachers' lack of knowledge of developmental theory. This caused them further difficulties with regard to how to interact appropriately with the students and interpret their abilities.

In a study carried out in Germany and the Netherlands, Edelenbos and Kubanek-German (2004) observed the diagnostic behavior of primary school teachers of English. The results revealed that many teachers found the task of capturing their students' level of ability challenging. The amount of time they spent on diagnosing their learners was limited (11.3% of the lesson time, p. 266). Similarly to the study by Rea-Dickins and Gardner (2000), during classroom observations no systematicity regarding the application of diagnostic procedures could be found.

Butler (2009a) reviewed studies conducted on teacher-based assessment in three Asian countries: Korea, Taiwan and Japan. He emphasized the importance of central specifications of assessment procedures and noted that such guidelines existed in none of the three countries. In Korea and Taiwan, the government recommended that teachers apply teacher-based assessments or informal observation-based assessments, such as portfolios. However, classroom observations revealed that the teachers found it challenging to use these forms of assessment, which had been traditionally less valued in East Asia (Butler, 2009a). The difficulties mainly stemmed from teachers' lack of time and knowledge to apply teacher-based assessment, large class sizes, and the lack of assistance given to teachers in how to utilize the information gained through these assessment procedures. In Japan, teachers were not required to carry out assessment at primary school level; however, they did apply some form of self-assessment. Butler (2009a), however, pointed out that how accurately these assessments demonstrated students' performance and how the information collected during these procedures could be utilized to improve teaching was often unclear.

In Hungary, similarly to the international arena (e.g., Brumen, Cagran, & Rixon, 2009; Butler, 2009a, 2009b, 2015; Edelenbos & Kubanek-German, 2004;

Peng & Zheng, 2016; Wilden & Porsch, 2016) research on teachers' assessment practices is fairly limited. A few studies, however, explored classroom assessment. In a retrospective study (Nikolov, 2001) which examined 185 Hungarian adults' language learning experiences, all the participants claimed that in primary school the assessment of their performances had always induced anxiety in their classrooms. Other studies (Bors, Lugossy, & Nikolov, 2001; Nikolov, 2003; 2008) reported similar trends: Students dreaded and disliked most tests and assessment. The most common forms of assessment were translation and tests of vocabulary and grammar. Oral competence was tested primarily by having students recite a text they had crammed. In other words, the dominance of the grammar translation and audio-lingual methods and activities inappropriate for young learners, which also prevail in English as a foreign language (EFL) classrooms in Hungary (Lugossy, 2009; Medgyes & Nikolov, 2014; Nikolov, 1999a, 2001, 2009c), were found to be typical ways of assessment.

In terms of classroom activities appropriate for this age group, in their review of research on early language learning programs in Europe, Edelenbos and Kubanek (2009, pp. 52-53) stated that one of the eight aims of these programs was "to improve the level of communicative competence reached by students through their educational system." A study by Lundberg (2007), conducted in Sweden to improve teaching EFL in young learners classrooms, found that using communicative and authentic materials and involving pupils in the planning and evaluation of teaching and learning stimulated the motivation of the participating students aged 10 to 11 the most. In contrast, "a slow pace and a lack of physical activities, too much revising and too few challenges, lack of variety, uninteresting material, too much silent work on your own, inconsistency of target language use and isolated lesson units without cohesion" proved to be demotivating (Lundberg, 2007, p. 30).

The objective of the present case study is to offer more data on classroom assessment practices and their impact on YLLs. It provides a thick description of how a Hungarian English-as-a-foreign-language (EFL) teacher assessed and gave feedback to her students, and of what assessment tasks she used and how. The study is a follow-up to a large-scale project (Nikolov & Szabó, 2012), which consisted of three phases. The first phase aimed to explore teachers' classroom assessment practice and to establish a baseline to build on good practice. Next, building on the experience of the first phase, a list of 18 tests was drawn up. A total of 18 EFL primary school teachers volunteered to choose eight that they considered suitable for their students' language proficiency. They were asked to pilot and evaluate them according to a set of given criteria. Based on the results of the previous phases, diagnostic tests were developed and piloted in the third phase. The ultimate aim of the project was to design, pilot, and calibrate diagnostic tests Hungarian primary school EFL teachers could use for diagnosing their students' progress.

In the first phase of the large-scale project, 12 primary school EFL teachers volunteered to choose and characterize 10 tasks they had used successfully with their students for testing their EFL knowledge (Hild, 2014b). They were given a form and asked to complete it for each task they chose. The findings of this qualitative exploratory study coincided with the results of studies on FL teaching and learning in Hungary (Medgyes & Nikolov, 2014; Nikolov, 2009c). They indicated that the teachers did not have a clear view of what task, task difficulty, skills and sub-skills meant and, thus, had difficulties when applying these categories to the tasks they chose (see Hild, 2014a; Hild & Nikolov, 2011 for more details). While describing learners' performances they applied loose and fuzzy terms and no clear criteria. They put emphasis on errors and accuracy rather than fluency and vocabulary, and on what students could not do, as opposed to what they could achieve. Feedback was often provided in the form of rewards (red points) for top performance only. Lower achievers, however, did not receive any reward, which may easily result in a motivational decrease in less able learners. The description of the assessment procedures of the tasks showed that the students were rarely, if at all, provided with detailed feedback regarding their strengths and weaknesses. The results also suggested that the teachers tended to separate teaching and assessment. They did not scaffold their students' performance during assessment, which would have assisted learning and the successful completion of the tasks and also revealed prospective development. Many of the activities the teachers had chosen were not in line with the needs and cognitive abilities of their pupils.

This qualitative exploratory study (Creswell, 2003; Mackey & Gass, 2005) is a follow-up to the large-scale project. My aim was twofold. Firstly, since only reading and listening tasks developed and piloted in the large-scale project had been analyzed (Nikolov & Szabó, 2012), I intended to find out how the oral tasks had worked. Secondly, since the literature on classroom assessment is fairly limited, I intended to explore early EFL teachers' assessment practices from an emic perspective involving one teacher. The present article aims to answer three research questions:

1. What kind of oral assessment tasks did the teacher use in her class?
2. How did the teacher assess and give feedback to her students?
3. What did the teacher and her students think about the diagnostic speaking tests?

3. Method

3.1. Participants

An EFL teacher, Anikó (all participants' names are pseudonyms), and five of her seventh graders from her class of 17 students, Robi, Béla, Balázs, Anett and Lili, agreed to participate in the study. After Lili fell ill, her classmate, Anett, replaced her in the study. The participating students were aged 12 to 13. They all attended the same class in a prestigious primary school in a large town in South-West Hungary. It is one of the schools affiliated to the university in town, where in-service teachers do their teaching practice. Therefore, its teaching staff is considered to be well qualified.

3.2. Data collection

Data were collected with semi-structured interviews with the teacher and the students. For the purpose of triangulation (Mackey & Gass, 2005, p. 181), the students were also audio- and video-recorded while doing four speaking tasks, and I also observed two classes and took notes. The datasets elicited during the interviews and video recordings were analyzed for themes and issues (Creswell, 2003, pp. 190-195; Mackey & Gass, 2005, pp. 178-179). The participants' answers were translated into English.

The students were asked about their opinion of the oral tasks immediately after completing them so that the drawback of selective recall or memory loss could be minimized (Mackey & Gass, 2005, p. 174). They were also interviewed at a later stage during which the following topics were addressed:

- typical English classes: tasks, work mode, activities;
- assessment in class: self-, peer-assessment and teacher assessment;
- the diagnostic tests they had piloted with their teacher a month earlier in the framework of the final stage of the large-scale project.

The questions of the interview with the teacher tapped into the following topics:

- the students: general abilities, language proficiency, motivation, out-of-school language use;
- typical English classes: tasks, teaching materials, work mode, assessment, practice, tasks popular and unpopular among students;
- assessment procedures in class;
- piloting the diagnostic speaking tests: the teacher's opinion about the tasks and the assessment procedures.

The language of the interviews was Hungarian since the aim was not to test the participants' language proficiency but to elicit information regarding their views

and experiences. While piloting the oral tasks with the students English was used, except for the introductory discussions and discussion of their opinions about the tasks. However, the students were also told that they were free to switch to their mother tongue whenever they felt the need.

3.3. Procedure

In January 2011, teachers who had participated in the third phase of the large-scale project were sent an email inquiring if they were willing to take part in a follow-up study. One teacher volunteered to participate with her students. The students' parents signed a consent form in which they agreed to their children's participation in the research. The interview with the teacher was conducted in two sessions during her breaks between classes because otherwise it would have taken up too much of her time.

Before I started to work with the students the teacher allowed me to observe two of her English classes. The week after the interviews with the teacher, the students carried out the four oral tasks in pairs. I audio- and video-taped one pair at a time during their regular English classes in a free classroom; thus, I did not use their free time. They were highly motivated and intrigued knowing they were participating in a research study. In order to put them at ease I informed them that this was not a test and their performance would remain confidential and would not count towards their assessment at school. I explained to them in detail the steps of the procedure. The interviews with the students were carried out during class time the week after all of them completed the oral tasks. I interviewed them in their classroom; therefore, the environment was familiar.

4. Results and discussion

4.1. Oral assessment tasks and classroom activities

As has been mentioned in the literature review, in young learners' FL classroom, short, communicative and interesting activities should constitute the core of their teaching. As to the oral tests Anikó used in class to assess her students' language learning development, the findings revealed that she mostly required the students to cram texts in the coursebook and recite them in class; this practice was in line with the results of other Hungarian studies on early language learning (Bors et al., 2001; Nikolov, 2003; 2008). Contrary to the communicative approach, Anikó believed that "*having students cram a text*" was a good "*method*" because it could be assessed "*objectively*," and it "*is even good for the less able learners: They do not have to think about what to say.*" When some

of the students tried to talk "*right off the top of their head*," that is, use the FL freely to express an idea because they had not learned it by heart, Anikó considered it "*unacceptable.*" One of the boys, Robi, reported similar tendencies: "*It has also happened that one of my classmates did not learn it* [the story from the textbook] *but told us the way he remembered it. She* [Anikó] *said it was ok, but next time he should learn it by heart.*" Anikó also added: "*Those who are more able and creative may change it* [the text] *in whatever ways they can.*" Such a distinction between low and high achievers can demotivate the less able children and, thus, make them lag further behind.

Children are more likely to participate in the activities willingly if they find the tasks interesting and enjoyable, that is, if they are driven by intrinsic motivation (Nikolov, 1995, 1999b). In terms of classroom activities, during the interviews both Anikó and her students reported that most of the time they covered the exercises and texts of the coursebook, *Project 3* (Hutchinson, 2003). Although Anikó claimed that she had regularly supplemented the book, as one of the students, Anett, put it, these activities tended to be "*a little bit similar*" and were mainly form-focused. One of the exceptions was a reading diary, which the students enjoyed and which seemed to have a positive impact. Three out of the five participants read books in English out of school. Variety, which is a must in an early language learning classroom, did not characterize Anikó's classes. As two of the participating children replied when asked about the most and the least frequently used tasks, there were no such activities, because "*what we do we do regularly.*" In the interview the students mainly mentioned classroom activities such as gap-filling, "*true or false*," and, the activity described by Béla in this way: "*There is a picture, and we have to write down what the people were doing yesterday at 12.*" Robi mentioned another activity: "*There was this exercise when there was a word,* have*; and we had to replace it for* will have."

Throughout the interview with Anikó it became evident that she attributed particular importance to grammar. She believed that "*it is necessary to give a grammatical basis to their* [students'] *knowledge so that they can use the language accurately. Naturally, the emphasis is not on this, but I strive to keep them in balance, grammar and vocabulary, so that they feel that both are important.*" Anikó feared that her students did not like practicing grammar, but believed they "[should do] *it properly, anyway.*" When it comes to teaching grammar to YLLs, teachers should ensure that in the classroom the focus is on the meaning, and, while paying attention to the meaning, young learners can implicitly use their inductive logic to infer the underlying language rules and gain knowledge of the grammar (McKay, 2006, p. 42). Providing learners with forms and rules and expecting them to apply those to specific cases, which was often the case in Anikó's classroom, are cognitively too demanding for this age group and, hence, highly demotivating.

The two classes I observed also provided valuable information about Anikó's classroom practices. Before the first class I observed Anikó told me: "*I cannot show you anything about these students today; whether they will participate or not; you will not see anything about them.*" She was proved right. The whole class was devoted to discussing the homework, which was to make a "spidergram," a list of words organized in a cobweb fashion. The students were to collect words in connection with traffic and group them under headings such as rules, people, jobs and vehicles. In class the students mostly did nothing but shouted or read aloud the words they had previously collected at home in connection with the topic of traffic and completed their lists with new words whenever it was necessary. Those who came up with unfamiliar words did have the opportunity to use English freely and explain the meanings of these words, which most of them did really well. But the rest took the easy way out and rather gave the Hungarian equivalent without the teacher encouraging or helping them to use the English language. On those few occasions when the students had an opportunity to use English, Anikó did not provide them with feedback on their performance. She mainly asked for more words that could go under the various headings of the spidergram. However, as mentioned earlier, providing detailed feedback is especially important in the case of children since it helps them see how much they have progressed, motivating them to succeed (Edelenbos, Johnstone, & Kubanek, 2006, p. 80; Edelenbos & Kubanek, 2009; McKay, 2006, p. 23). Although the task itself might allow for more use of English in context, the aim seemed to be to gather as many words as possible. During the class, however, it did not turn out what the students were supposed to do with the complete spidergram, which covered a whole A4-size piece of paper. Since children easily get diverted (McKay, 2006, p. 6), it is preferable to apply shorter activities in the classroom. Interesting, engaging and challenging tasks can motivate them to participate and pay attention (McKay, 2006, p. 41). Spending the whole 45-minute class on collecting words proved neither short nor interesting/challenging. This was reflected in the students' behavior; they lost interest and started to talk and fidget towards the end of the class.

In the first 30 minutes of the second class I observed, the students carried out a motivating oral task. They were to give a presentation in pairs about an object they invented. However, as it later turned out, these presentations had been written previously by the students and then corrected by Anikó. Therefore, the students' opportunity to use the language freely and improve their communicative skills was again compromised. Except for one case, neither during nor after the presentations did Anikó take the opportunity to encourage the students to use the target language in context by asking further questions or inviting the classmates to inquire about these objects after the presentations, but

instead she said: "What else have we got for today?" Initiating discussion would have also allowed for all the other students to participate in classwork rather than just listening to their peers. Giving immediate feedback about students' performance is very important so that they know in what area they should improve (Alderson 2005, p. 208). However, besides thanking the students for their performance, Anikó failed to comment on their presentations.

The remaining 15 minutes of this class were devoted to revision. They discussed a story from the coursebook, which they had covered in the previous lessons. Afterwards, Anikó quizzed the students on the vocabulary of the story. She gave the English definitions of certain words and expressions in the story on the basis of which the students were to give the English terms. They often did not wait for the teacher to ask them to reply but shouted the answers, which Anikó had no problem with. She was content if she heard the correct answer and did not attempt to find out who said it, which was often impossible. It seemed that she assessed the vocabulary knowledge of the whole class, rather than the individual students' vocabulary. As for the feedback, she sometimes corrected the students or thanked them for the correct answer, which had no diagnostic value. The class was finished off with a competition. She asked for two volunteers who had to stand next to one of the desks. She said words, expressions or sentences in Hungarian, and the student who was quicker to translate them into English could take a step forward. The winner, who received a red point as a reward for good performance, was the student who reached the teacher's desk first. The other student's performance was neither rewarded nor commented on. As McKay (2006, p. 192) pointed out, "vocabulary is best assessed in an integrated way through language use in language use tasks." However, these two latter tasks required the students to memorize words, expressions and sentences out-of-context, and had little communicative value.

4.2. Classroom assessment

The results regarding the teacher's assessment practices were in line with the findings of the first phase of the large-scale project and other international studies carried out in this field (Butler, 2009a; Edelenbos & Kubanek-German, 2004; Rea-Dickins & Gardner, 2000). While describing the students' language knowledge, the unclear terms Anikó used revealed that she had difficulties with capturing her students' level of ability. She failed to define their proficiency levels in relation to any criteria: "*Well, they are at a good level . . . they can communicate and talk.*" She also added: "*The majority has very good abilities. Two of them are at a little lower level, but even they have abilities above average.*"

Feedback is a crucial element of assessment. It should always come right after students' performance, and be individualized and stimulating for further learning (Alderson, 2005, p. 208; Nikolov, 2011). Shohamy (1992, p. 515) emphasizes that "for assessment information to be used effectively it needs to be detailed, innovative, relevant and diagnostic, and to address a variety of dimensions rather than being collapsed into one general score." The development of strategy use, which is one of the main aims of early language learning, can also be assisted by providing students with regular feedback on their performance (Nikolov, 2011). Feedback should be provided regularly to enable young learners to see they are progressing and achieving what they are expected to (Nikolov, 2011). The interviews and the classroom observations demonstrated that the feedback Anikó gave to her students was often too general or fuzzy and contained hardly any specific information on the strengths and weaknesses of their performance, which suggests she did not bear in mind any criteria for assessment. After a test, Anikó usually distributed the papers and only provided a general evaluation of the class's performance. According to one of the students, Robi, on such occasions she told the whole class that "*there were very good ideas she could give red points for. But there were quite a lot of mistakes, too.*" According to Lili, in tests Anikó "*underlines the mistakes or puts a question mark next to them.*" After getting back their test papers, as Béla voiced it to me, they sometimes "*correct the mistakes, and . . . write down the correct version in our exercise book.*" Lili reported that "*it happens that she* [Anikó] *goes up to a student and discusses with him what he needs to practice.*" When Anikó sometimes took the students' exercise books home, as Lili put it, "*to check if we have done everything, and if we have done it correctly,*" she did not talk to them individually, but "correct[ed] *our mistakes, . . . and . . .* [gave] *it back to us to see it.*" During the interview Anikó mentioned that when she took home the students' exercise books, the comments she usually wrote were like this: "*This was good or wrong. Nice job. Well done. . . . This was unsatisfactory, and you can do better than that,*" which gave hardly any clue to the students as to what they should practice or improve. One of her best students, Balázs, also complained that when Anikó was not satisfied with his performance she only said: "*I expect more from a student at your level,*" which Balázs did not consider very useful since "*when I do what she asked, I don't understand why she is saying that.*"

In terms of detailed, individual feedback, Anikó stated that the students usually received such "*feedback after we finish a topic, or, at least, at the end of each semester, when they get a mark for their end of term test.*" On such occasions, she informed the students who had "*performed at the level expected of them,*" whom Anikó would have "*expected more of, who should add what to his performance and who should practice what.*"

In class, as a reward, Anikó gave red points rather than marks, unless some students put in an "*outstanding performance*," which deserved grade 5. Those who did not perform well did not receive anything because, as she formulated it, "*they can see anyway that it did not go well and they need to do more, because out of 10 only seven were correct.*"

When asked about classroom assessment and feedback, the students also described techniques contradicting one another. Two stated that during an oral test Anikó took notes, but after it she only said, "*It was good*," without further detail. Three students claimed that the teacher read out her notes and told them what to pay attention to next time. Robi, for example, stated: "*She tells us what exactly we need to pay attention to. For example, I tend to forget about the future tense.*" This implies that Anikó did not always provide the students with information about the strong and weak points of their performance, and they had no idea what to improve.

The two "good" students, Balázs and Béla, claimed that they did not really receive any feedback on their performance because they did not make mistakes: "*I don't make mistakes, so she doesn't tell me anything.*" According to them, however, Anikó commented on the lower achievers' performance more often. As Béla formulated it, "*she doesn't usually come up to me, . . . but it happens more often if somebody is not so good at English.*" This practice suggests that in Anikó's class feedback mainly focused on what the students could not do and rarely on what they did well, which is not conducive to early language learning (McKay, 2006, p. 9, Nikolov, 1999a). The students' reports also suggested that Anikó concentrated on accuracy while giving information on their performance.

Contrary to principles of dynamic assessment, Anikó believed that if the aim was progress testing, as she formulated it, "*to find out what they* [the students] *could do*," no assistance was acceptable. This belief is in line with the assessment traditions in Hungarian education in general, not only in FL classrooms.

In the classroom of YLLs, self- and peer-assessment play an essential role in the establishment and development of learning strategies, one of the main aims of early language learning (Nikolov, 2011). However, Anikó was not content with the idea of applying self-assessment because then the assessment "*will not be objective or realistic.*" She believed that "*then the assessment will certainly not be appropriate and relevant . . . because students cannot really decide if that sentence was grammatically correct or not. The best ones may be able to judge it, but it isn't for sure.*" Her concern, which then proved right while I was piloting the oral tasks with the students, was that students disregarded grammatical mistakes and only concentrated on whether they understood what their classmates had said, or, in other words, on the meaning they meant to convey. Hence, she thought that "*it is done properly only when I do the assessment.*" As Ioannou-

Georgiou and Pavlou (2003, p. 99) also stated, although communication is the priority in the early language learning classroom, accuracy does count and also has to be dealt with. However, the focus should be on meaning, especially if the aim of the task is to assess communicative ability (Ioannou-Georgiou & Pavlou, 2003, p. 99), which was the case in these oral tests. Therefore, the students did what they were supposed to do at this age.

4.3. The teacher's and the students' opinion about the diagnostic tests

In the interview Anikó was also asked to give her opinion about the diagnostic speaking tasks she had tried out with her students a month earlier. Although Anikó was an experienced EFL teacher at a school known for its English language program, she provided little concrete information as to what exactly she liked or disliked in the diagnostic tests she had tried out with her students. Besides emphasizing how easy most of these tasks were and that her class had already covered most of the topics and grammatical structures the tests focused on, she did not go into any detail regarding how well the tasks had worked with her students.

While describing her experiences with the diagnostic tests Anikó was thinking in terms of normative and holistic testing following the traditions typical of the Hungarian educational system. Anikó was concerned that the scoring guide on how many words and mistakes an utterance could contain to be worth a particular number of scores did not take into account the differences between students' language knowledge: "*For some students a 3-word sentence would be an excellent performance, whereas others can say 5 or 6-word-long sentences.*" She also added that though this assessment procedure was "*more nuanced, it was more difficult to follow. So, now is it five, four or three points. And then how shall we convert it into grades? Our kids and educational system think in terms of grades.*" Although she realized that this method was, as she put it, "*more nuanced*" and provided more detailed information on students' knowledge, she still preferred "*grades,*" that is, holistic testing.

Anikó mainly had problems with those tasks that had multiple solutions, such as the picture description activities or the 99 questions. For example, in one of the picture description tasks, she could not decide if "*the man is walking or going somewhere, or the girl is putting the picture on the wall or taking it off.*" In these cases Anikó told the students to write about what they thought there was in the picture. However, as she formulated it, this "*made checking more difficult, and . . . lengthy . . . since it took us a whole class* [to complete it]. *It takes time for all of them to ask if it is ok in this or that way, because there were so many variations.*" In class, she usually solved this kind of problem by specifying the grammatical structures students should use to formulate sentences. In

order to make assessment, as she put it, "*straightforward,*" she also applied this technique while piloting the diagnostic task of picture description: "*I had to make sure and convince the students not to overcomplicate the description of the pictures. So they should use this structure: This is a man who does this and that. And this is where they should stop.*" In other words, Anikó prevented the students from demonstrating how they could use English creatively and freely, which was one of the main objectives of these tasks.

During the interview, the students stated that the oral diagnostic tests they had tried out with Anikó were mostly unfamiliar to them. They remembered doing activities similar to the 99 questions, when they had to write and then ask their partner questions. They also mentioned that they sometimes described pictures. However, the students' report suggested that during these tasks the focus was on practicing grammatical rules rather than on using the language for communicative purposes: "*When we discussed present continuous there was this exercise where there were pictures, and we described what the people in the picture were doing at that moment.*"

Similarly to Anikó, the only difficulty the students had with the written tasks was that some of the pictures were, as they put it, "*blurred.*" This really meant that in some cases more than one description of the picture was appropriate. In the classroom of communicative language teaching, this should not cause any problems. Multiple solutions can increase the opportunities for learners to use the FL creatively and come up with as many ideas as they can, which is one of the aims of early language teaching.

Though the students hardly ever carried out self-assessment in class, they claimed they did not find it difficult to follow the instruction and assess their performance while doing the diagnostic tests. However, according to Robi's report, when they were asked to do it, some of them did not take it seriously and wrote the maximum score everywhere, suggesting that they attributed little importance to their own opinion: "*Some of us thought that ouch, it is not Anikó who will check it, we have to do it. Then I will write a 2* [the maximum score] *everywhere to fool them.*"

5. Conclusions

The aims of the paper were to determine: (a) what kind of oral assessment tasks a primary school EFL teacher used in class, (b) how she assessed and gave feedback to her students, and (c) what the teacher and her students thought of the diagnostic tests they had previously tried out in the last phase of the large-scale project. Since little is known about teachers' classroom practices, a qualitative, exploratory approach was adopted (Dörnyei, 2007, p. 39). As a follow-up to the

large-scale project, the case study intended to provide a more in-depth look at assessment practices in the early language learning classroom. The findings were in line with the results of the first phase of the large-scale project (Hild, 2014a, 2014b; Hild & Nikolov, 2011) and other Hungarian (Bors et al., 2001; Nikolov, 2003, 2008) and international (Butler, 2009a; Edelenbos & Kubanek-German, 2004; Rea-Dickins & Gardner, 2000) studies carried out in this field.

The study shed light on deficiencies in the teacher's knowledge of age-appropriate methodology and assessment practices. Instead of short, intrinsically motivating tasks where the focus is on fluency rather than accuracy, teaching and practicing grammatical rules, memorizing out-of-context words and the monotony of the use of the coursebook activities or similar tasks characterized Anikó's classes. She considered having students recite a text the best way to assess oral skills. Since she attributed great importance to accuracy, even at the expense of creative language use, she accepted the "*more able*" students' deviations from the original text, but discouraged the less able learners from using English freely. The students' creative language use was limited even when they presented their own inventions because Anikó wanted them to stick to the written version of their presentations she had previously edited for them.

As for classroom assessment, Anikó rarely provided her students with detailed feedback on their performance. When she, however, did so, she found it hard to diagnose their strengths and weaknesses. She did not apply any criteria but provided her students with a general, often fuzzy description of their language knowledge and skills. Instead of focusing on fluency and emphasizing what the students did know, she was more concerned with how many mistakes they had made and whether what they had said or written was accurate. Similarly to the outcomes of the first phase of the large-scale project, rewarding the best performances and ignoring the other students' results were also typical of Anikó' assessment practice. According to her, scaffolding students while doing a task was not consistent with testing their knowledge. She refrained from the use of self- and peer-assessment because she believed students tended to focus on meaning rather than accuracy. Hence, she was of the opinion that only she could do a proper job when it came to assessing students.

With regard to the diagnostic tests Anikó had previously tried out in the last phase of the large-scale project, she formulated only a few general thoughts. She believed that her students had been able to tackle them easily. She had problems with the tasks that had several possible answers. In the picture description tasks, where the students came up with several ideas, she had difficulty scoring the answers. She wanted tests where there was only one correct solution, the number of which could then add up to a grade. Even though she realized that the diagnostic tests allowed her to obtain a more comprehensive

analysis of students' language competence, she still preferred tasks where the assessment was more "*straightforward.*" These findings suggested that she insisted on the normative and holistic testing methods, which are typically applied in the Hungarian educational context, and found it hard to deliver diagnostic assessment. However, since children learn FLs slowly, it is essential to test their progress regularly to maintain their motivation by showing them how much they have developed and that hard work was worth the effort.

5. Limitations and implications for further research

Despite its valuable findings, the research porject has limitations. One of them concerns the number of the participants. This qualitative, exploratory study set out to provide a more in-depth perspective and, therefore, was designed to work with only a few participants; however, only one teacher could be recruited. Further research is necessary to find out more about how teachers diagnose their students' strengths and weaknesses, how they give feedback and how they use the information they gained from tests. It would also be interesting to see how other EFL teachers would use these diagnostic tests, how they could follow the assessment instructions, and whether and how they would use them in the long-run.

A further limitation of the paper is that data were mainly collected through self-reports. More classroom observations would be necessary to support and supplement the information elicited by the questionnaire and the interviews. However, the teacher seemed to be threatened by the thought of me observing her classes. Therefore, besides the two classes I had observed, I was not provided with other opportunities.

On the basis of the results of the large-scale project the case study was a follow-up to, a list of can do statements, as well as topics and task types have been developed, which teachers can adopt in their own contexts (for more details, see Nikolov, 2016b).

Acknowledgments

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London: Continuum.

Bors, L., Lugossy, R., & Nikolov, M. (2001). Az angol nyelv oktatásának átfogó értékelése pécsi általános iskolákban [A comprehensive evaluation study of the teaching of English in Pécs primary schools]. *Iskolakultúra, 11*(4), 73-88.

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. Retrieved from http://wrap.warwick.ac.uk/2275/1/WRAP_Rixon_Brumen_Assessment_ok.pdf

Butler, Y. G. (2009a). How do teachers observe and evaluate elementary school students' foreign language performance? A case study from South Korea. *TESOL Quarterly, 43*, 417-444.

Butler, Y. G. (2009b). Issues in the assessment and evaluation of English language education at the elementary school level: Implications for policies in South Korea, Taiwan, and Japan. *The Journal of Asia TEFL, 6*(2), 1-31.

Butler, Y. G. (2015). Task-based assessment for young learners: The role of teachers in changing cultures. In M. Thomas & H. Reinders (Eds.), *Contemporary task-based language teaching in Asia* (pp. 348-365). London: Bloomsbury.

Commission of the European Communities. (2003). *Promoting language learning and linguistic diversity: An action plan 2004-2006.* Retrieved from http://www.eu.int/comm/education/doc/official/keydoc/actlang/act_lang_en.pdf

Creswell, J. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: SAGE.

de Bot, K. (2014). The effectiveness of early foreign language learning in the Netherlands. *Studies in Second Language Learning and Teaching*, *4*, 409-418.

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

Edelenbos, P., Johnstone, R., & Kubanek, A. (2006). *The main pedagogical principles underlying the teaching of languages to very young learners: Languages for the children in Europe: Published research, good practice and main principles.* Strasbourg: Council of Europe. Retrieved from http://ec.europa.eu/education/policies/lang/doc/youngsum_en.pdf.

Edelenbos, P., & Kubanek, A. (2009). Early foreign language learning: Published research, good practice and main principles. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 39-58). Berlin: Mouton de Gruyter.

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of diagnostic competence. *Language Testing*, *21*(3), 259-283.

Garton, S., Copland, F., & Burns, A. (2011). *Investigating global practices in teaching English to young learners.* London: British Council. Retrieved from http://englishagenda.britishcouncil.org/sites/ec/files/B094FINALAs tonUniversityA4report_2column_V3.pdf

Hild, G. (2014a). Assessing young language learners' development: Hungarian teachers' views on assessment tasks and practices. In J. Horváth & P. Medgyes (Eds.), *Studies in honour of Marianne Nikolov* (pp. 72-86). Pécs: Lingua Franca Csoport.

Hild, G. (2014b). *Assessment of young EFL learners in the Hungarian educational context* (Unpublished doctoral dissertation). University of Pécs, Hungary.

Hild, G., & Nikolov, M. (2011). Teachers' views on tasks that work with primary school EFL learners. In M. Lehmann, R. Lugossy, & J. Horváth (Eds.), *UPRT 2010: Empirical studies in English applied linguistics* (pp. 47-62). Pécs: Lingua Franca Csoport.

Hutchinson, T. (2003). *Project English 3.* Oxford: Oxford University Press.

Ioannou-Georgio, S., & Pavlou, P. (2003). *Assessing young learners.* Oxford: Oxford University Press.

Johnstone, R. (2009) An early start: What are the key conditions for generalized success? In J. Enever, J. Moon, & U. Raman (Eds.), *Young learner English language policy and implementation: International perspectives* (pp. 23-30). Reading: Garnet.

Lantolf, J. P., & Poehner, M. E. (2006). *Dynamic assessment in the foreign language classroom: A teacher's guide.* Pennsylvania: CALPER University Park. Retrieved from http://www.scribd.com/doc/57598521/DA-Guide

Lugossy, R. (2009). "I will think about this": A case study with a lower-primary school teacher of English. In R. Lugossy, J. Horváth, & M. Nikolov (Eds.), *UPRT 2008: Empirical studies in English applied linguistics* (pp. 59-69). Pécs: Lingua Franca Csoport.

Lundberg, G. (2007). Developing teachers of young learners: In-service for educational change and improvement. In M. Nikolov, J. M. Djigunović, M. Mattheoudakis, G. Lundberg, & T. Flanagan (Eds.), *Teaching modern languages to young learners: Teachers, curricula and materials* (pp. 21-34). Graz: European Centre for Modern Languages, Council of Europe Publishing.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* Mahwah, NJ: Lawrence Erlbaum.

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

Medgyes, P., & Nikolov, M. (2014). Research in foreign language education in Hungary (2006-2012). *Language Teaching, 47*, 504-537

Nikolov, M. (1995). Általános iskolás gyerekek motivációja az angol mint idegen nyelv tanulására [Primary school students' motivation to learn English as a foreign language]. *Modern Nyelvoktatás, 1*(1), 7-20.

Nikolov, M. (1999a). Classroom observation project. In H. Fekete, É. Major, & M. Nikolov (Eds.), *English language education in Hungary: A baseline study* (pp. 221-246). Budapest: British Council.

Nikolov, M. (1999b). "Why do you learn English?" "Because the teacher is short": A study of Hungarian children's foreign language learning motivation. *Language Teaching Research, 3*(1), 33-65.

Nikolov, M. (2001). A study of unsuccessful language learners. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 149-170). Honolulu, HI: The University of Hawaii, Second Language Teaching and Curriculum Center.

Nikolov, M. (2003). Angolul és németül tanuló diákok nyelvtanulási attitűdje és motivációja [Attitudes and motivation of English and German learners]. *Iskolakultúra, 13*(8), 61-73.

Nikolov, M. (2008). "Az általános iskola, az módszertan!" Alsó tagozatos angolórák empirikus vizsgálata ["Primary school means methodology!" An empirical study of lower-primary EFL classes]. *Modern Nyelvoktatás, 10*(1-2), 3-19.

Nikolov, M. (Ed.). (2009a). *Early learning of modern foreign languages: Processes and outcomes*. Clevedon: Multilingual Matters.

Nikolov, M. (Ed.). (2009b). *The age factor and early language learning*. Berlin: Mouton de Gruyter.

Nikolov, M. (2009c). The dream and the reality of early programmes in Hungary. In J. Enever, J. Moon, & U. Raman (Eds.), *Young learner English policy and implementation: International perspectives* (pp. 121-129). Reading: Garnet Education.

Nikolov, M. (2011). Az angol nyelvtudás fejlesztésének és értékelésének keretei az általános iskola első hat évfolyamán [A framework for developing and assessing proficiency in English as a foreign language in the first six years of primary school]. *Modern Nyelvoktatás, 17*(1), 9-31.

Nikolov, M. (2016a). Trends, issues, and challenges in assessing young language learners. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 1-18). New York: Springer.

Nikolov, M. (2016b). A framework for young EFL learners' diagnostic assessment: 'Can do statements' and task types. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 65-92). New York: Springer.

Nikolov, M., & Mihaljevic-Djigunovic, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics, 26*, 234-260.

Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics, 31*, 95-119.

Nikolov, M., & Szabó, G. (2012). Developing diagnostic tests for young learners of EFL in grades 1 to 6. In E. D. Galaczi & C. J. Weir (Eds.), *Exploring language frameworks* (pp. 345-361). Cambridge: Cambridge University Press.

Peng, J., & Zheng, S. C. (2016). A longitudinal study of a school's assessment project in Chongqing, China. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 213-242). New York: Springer.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development.* Berlin: Springer.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*, 249-258.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17*, 215-243.

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal, 76*, 513-521.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential.* Cambridge: Cambridge University Press.

Wilden, E., & Porsch, R. (2016). Learning EFL from year 1 or year 3? A Comparative study on children's EFL listening and reading comprehension at the end of primary education. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 191-212). New York: Springer.