

THE DEVELOPMENT OF A CEFR-BASED SCALE FOR ASSESSING YOUNG FOREIGN LANGUAGE LEARNERS' WRITING SKILLS

1 INTRODUCTION

In the last two decades, many countries across the globe have begun providing foreign language instruction at primary schools, typically with English as the target language. However, English-dominant countries have also started to introduce the statutory provision of foreign language education in primary schools (DfES 2002; Evans/Fisher 2012), with the aim of ensuring that their citizens become efficient lifelong language learners.

As foreign language instruction at primary school has gained popularity worldwide, educational researchers, language specialists and policymakers have expressed concern over the accountability of these programmes, and especially about the inadequate training of their teachers. Unfortunately, there are still many countries that lack appropriately trained teachers. In Vietnam, for instance, Nguyen (2011) reports that most primary school English teachers are not formally trained to teach English at the primary school level. Even where there are enough teachers, such as in Bangladesh or Nepal, many are not adequately trained, nor do they have adequate English language skills (Hamid 2010; Phyak 2011). Hasselgreen, Carlsen and Helness (2004) found that even teachers trained as language specialists expressed a great demand for training in various areas of assessment, such as “defining criteria” and “giving feedback”. Thus it seems that teachers involved in primary school foreign language teaching require assistance and support in both teaching and in assessing young foreign language learners, especially when it comes to giving appropriate feedback.

While learners seem to be well-motivated for communicative, humanistic and learner- and content-based teaching approaches, their language progress needs to be monitored and assessed. Some educational systems (Finland, Sweden etc.) avoid traditional large-scale achievement tests at the primary school level and, instead, strongly promote classroom-based (teacher) forms of assessment. It has been noted that the application of teacher assessment appears to vary tremendously from teacher to teacher (Goto Butler/Lee 2010). On the whole, however, teachers need to assess the performance of individual students in a way that leads to further learning. In this way, teachers are able to improve their own instruction and satisfy the different needs of young language learners. It is the purpose of this article to describe the process of developing an assessment

* *Author's address:* Pedagoška fakulteta, Univerza v Ljubljani, Kardeljeva ploščad 16, 1000 Ljubljana, Slovenia. E-mail: karmen.pizorn@pef.uni-lj.si.

instrument which should support foreign language teachers in assessing writing skills and giving helpful feedback in such a way that learners will be able to develop their language proficiency.

2 ASSESSMENT FOR LEARNING

Assessment covers all of those activities performed by teachers that enable the measurement of the effectiveness of teaching and learning processes. Any kind of assessment should provide a reliable answer to the question *Have the students learnt what they were supposed to learn?* There are three main purposes of assessment: (1) to make schools and teachers accountable for their work, (2) to issue certificates confirming students' attainment, and (3) to advance student learning and help them to improve (usually termed "assessment for learning" or "formative assessment") (Black/Harrison/Lee/Marshall/Wiliam 2004: 10). The present article focuses mainly on the third of these purposes, which reflects the main aim of the Assessment of Young Learner Literacy (AYLLIT) project, on which this article reports.

Assessment for learning has been defined as "any assessment for which the primary aim is to fulfil the purpose of enhancing students' learning" (Black/Harrison/Lee/Marshall/Wiliam 2004: 10). The information derived from the assessment process should be applied by teachers and students alike. In other words:

An assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted and used, by teachers, learners or their peers to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have made in the absence of that evidence (Wiliam 2011: 43).

General principles that underlie assessment for learning, and thus enable students to improve, include:

- the provision of *helpful* and *constructive feedback* to students;
- the active *involvement* of students in their own learning;
- *teacher adjustments* to future instruction, based on the outcome of the results of the assessment;
- making learners aware of the *success criteria* that need to be met, in order to do well in the assessment activity (Faxon-Mills/Hamilton/Rudnick/Stecher 2013).

Assessment for learning is essential for several reasons: (1) a thoughtful and well-informed classroom assessment practice ensures that students are able to achieve their educational potential; (2) formative ways of assessing students take into account variation in students' needs, interests and learning styles, and attempt to integrate assessment and learning activities; (3) a number of research studies have shown that the use of assessment to develop students' future learning makes a substantial difference, not only to students' attainment but also to their attitude towards learning, their engagement with the subject matter, and their motivation to strive for better results at school

(Black/William 1998; Hattie 2012; Murphy 1999); and (4) assessment for learning is viewed as closely related to instruction, and is needed to help teachers make decisions about learning and teaching processes. However, the success of any assessment process depends on the effective selection and use of appropriate tools and procedures, as well as on the proper interpretation of students' performance.

3 ASSESSING WRITING SKILLS OF YOUNG LANGUAGE LEARNERS AND THE IMPORTANCE OF VALID FEEDBACK

Writing seems to be a straightforward and easy skill to assess. It provides the teacher with documentation of what the student can produce at a given time. Corrective feedback on errors may be given, and the writing may be discussed with the student and retained, to allow for subsequent comparisons between earlier or later performances. However, without a thoughtful, planned and systematic way of carrying out this assessment, it may have little formative value and can lead to imprecise summative information. For example, Stobart (2006: 141) explored conditions that may prevent the assessment from leading to further learning, and underlined the quality of feedback as critical. He established five preconditions for valid feedback to occur in the classroom: (1) it is clearly linked to the learning goals; (2) the student is able to understand the success criteria; (3) it gives an indication, at appropriate levels, on how to bridge the gap; (4) it focuses on the task, rather than the student; (5) it challenges and inspires students to do something about their progress, and it is achievable.

In order to help students develop their writing skills, teachers need to be able to provide appropriate (corrective) feedback. This should be based on criteria shared between the teacher and the students (Bitchener/Ferris 2012: 124). Moreover, the student should be able to assess his/her own performance using the set criteria, and to assess his/her progress by placing a piece of writing at a level or target point that consists of a description (descriptors) and a sample (benchmark), which illustrate the level in question. For both teacher and students, it is vital that the descriptors are interpreted in the same way. However, it is not just criteria that influence students' progress in FL writing, but also the tasks set by the teacher. Writing tasks have to be designed in a way that allows students to demonstrate their writing abilities. For example, if the task is not close to a young learners' life experiences and interests, it will not stimulate them to show their true communicative competence. The AYLLIT project, therefore, focused on the following vital issues: (1) the development of criteria, (2) the design of guidance for teachers on giving feedback to students, and (3) the design of guidelines for preparing writing tasks (AYLLIT 2007–2011).

4 WHY THE LINK WITH THE CEFR?

The Common European Framework of Reference (CEFR) is a useful and increasingly-known and used tool for the assessment of languages in the classroom. It has two broad aims: to act as a stimulus for reform, innovation and reflection, and to provide

Common Reference Levels, to assist communication across institutional, regional and linguistic boundaries. Before the publication of the CEFR, dialogue about levels of language competence was hindered because each school, institution, testing centre or ministry described, targeted and achieved language levels in their own terms. The CEFR helps to overcome these barriers by providing a common framework for the description of levels, course planning, assessment and certification. It is used for specifying content (what is taught and assessed) and stating criteria (how performance is interpreted).

It is a frame of reference, and must be adapted to fit a particular context. Linking to the CEFR means relating the features of a particular context of learning and teaching to it. Not everything in the CEFR is relevant to any given context, and there are features that may be important for a particular context, but which are not addressed by the CEFR. This is particularly true for young learners, who are not very well-covered in the descriptive scales, as these scales were developed with adults in mind, and do not take into account the cognitive stages prior to adulthood. Adaptation of the CEFR for young learners has been undertaken in many ways: for example, through primary school versions of the European Language Portfolio (ELP), and through national language curricula (Little 2006; Pižorn 2009). However, this is not a straightforward task. Papp and Salamoura (2009) report that assessors attempting to relate Cambridge Young Learners of English examinations to the CEFR found it difficult to map young language learners' performances and tasks against CEFR scales and descriptors.

One of the main, and most influential, parts of the CEFR is its descriptive scheme, which embraces general competence (knowledge or skills, know-how or existential competence, and the ability to learn) and communicative language competence (linguistic, pragmatic, socio-linguistic and sociocultural). It distinguishes four categories of language activity (reception, production, interaction and mediation), four domains of language use (personal, public, occupational and educational), and three types of language use (situational context, text type, and conditions or constraints) (CEFR 2001; Little 2007). For the purposes of classroom assessment, it is necessary to be able to establish not only which tasks the student can perform but also, and importantly, how well s/he can perform them. One of the principal aims of this project was, therefore, to adapt the already-existing ELP scales, with their functional focus, by producing a CEFR-based scale with a linguistic focus.

5 THE PRE-PHASES OF THE AYLLIT PROJECT

The AYLLIT materials were developed in three phases. The first, known as ECML's Bergen "Can Do" project, resulted in a scale that was a forerunner of the AYLLIT scale. The second phase was a preliminary project undertaken immediately prior to the AYLLIT project, and the third phase was the AYLLIT project itself.

In the first phase, two CEFR-based scales of descriptors were developed in Norway for the assessment of writing, as part of the National Testing of English (NTE) in 2004–2005 (Helness 2012). The first focused on the functional aspect of writing, while the second had a linguistic focus and was not task-specific. The latter consisted of four ca-

tegies – textual structure, grammar, words and phrases, and spelling and punctuation – and was primarily based on the CEFR scales of descriptors. The bands of descriptors were only formulated for whole levels (A1, A2, B1, B2), and shaded areas between these levels were given (A1/A2). Next, teachers were asked to rate the scripts using the scale. Hasselgreen (2013) reports that, on the English tests for Grade 10, the inter-rater correlation between experts and teachers was 0.81. For Grade 7, the raters were generally close in their ratings: 34% were in complete agreement, while 40% differed by only half a CEFR level. Hasselgreen (2013) gives further evidence for using the scale, by reporting on teachers' perceptions of the usefulness of the training in the use of the scale, with only 3% answering that it was not useful, while all of the others found it (very) applicable. Teachers also commented that the scale would be very useful for classroom assessment of students' writing. Thus the NTE scale proved to have a high degree of near-agreement in placing students on a CEFR-based scale, and was regarded as useful to teachers in assessing writing. However, the NTE scale was not ready to be used in the AYLLIT project, due to the levels included, and the fact that the descriptors were primarily intended for testing purposes, rather than classroom assessment.

The second phase refers to a preliminary project carried out one year before the AYLLIT project, involving two Grade 5 classes (10–11 years old) in Norway. The project had two purposes. The first was to identify what students of this age could be expected to write, and what kind of assessment tools teachers would find useful. The second was to adapt the NTE scale into a form that both teachers and the project leader would find better-suited to the classroom assessment of students' writing. On the NTE scale, each band of descriptors represented a whole CEFR level, from A1 to B2. According to the research findings (Hasselgreen 2013), it was agreed that the level B2 may be cognitively beyond the reach of students at this age. Furthermore, it was felt that, in order to provide meaningful feedback and allow progress to be shown, descriptors at in-between levels should be provided. As a result, the scale was revised and six bands of descriptors (A1, A1/A2, A2, A2/B1, B1, and above B1) were included. The decision was also made to adjust the categories to include some indication of the functions a student may be expected to perform. These categories were renamed Overall Structure and Range of Information, Sentence Structure and Grammatical Accuracy, Vocabulary and Choice of Phrase, and Misformed Words and Punctuation. This work resulted in a pre-scale leading to the final AYLLIT project scale.

6 THE AYLLIT PROJECT

6.1 Introduction

The third phase refers to the AYLLIT project itself, which was part of the 2008–2011 medium-term programme of the European Centre for Modern Languages (ECML), and was aimed at designing CEFR-linked guidelines and materials for primary school foreign language teachers to use in their classroom assessment of their students' reading and writing skills. The guidelines and materials for teachers were finalised following a workshop with participants from 30 European countries. Although research in the AYLLIT project

was qualitative, the AYLLIT material was thoroughly discussed and revised in the project group, and with teachers and students from all of the participating countries, until it was perceived to be appropriate for the context of classroom assessment.

The AYLLIT project team consisted of four experts representing Lithuania, Norway (coordinator), Slovenia and Spain. Two classes of students (aged 9 at the beginning) and their teachers in each country took part in the project, over a two-year period. The common foreign language for the main part of the project was English. In each of the four countries, it was assumed that children, at this stage, are able to read and write English. There was close cooperation and regular contact between the team members and the teachers in their respective countries. The role of the teachers was to be closely-involved in the whole process: administering, assessing and commenting on writing tasks, and collecting the reactions of students. The role of the team members was to draft and assign writing tasks and procedures, to revise the scale of descriptors using samples of students, to assess students' scripts already assessed by teachers, to send scripts to schools abroad, and to collect comments from teachers using the materials. The data consisted of tasks designed and revised by team members and teachers, as well as students' writing scripts, teachers'/experts' comments, and ratings of students' texts by teachers and experts. Finally, before finalising the materials, a workshop with 30 participants (most of whom were not part of the project) was organised.

6.2 AYLLIT writing process

Curricula for literacy in English in the four countries proved to be quite diverse. However, concerning foreign language writing skills, students were expected to be able to write communicatively, and at some length, on personal topics, in a descriptive and narrative way. Learners at this age should do tasks that are intrinsically motivating and challenging (McKay 2006: 250–251; Wilford 2000: 1). Cameron (2001: 156) argues in favour of writing for real communication. The idea that children are motivated when they are encouraged to talk about themselves, and to share such information with their peers from other countries through writing, was crucial to the way writing was conducted in the AYLLIT project. The writing tasks that the team designed for students reflected “can do” statements for the appropriate levels in the countries' ELPs. The initial tasks were descriptive in nature, such as introducing oneself, and sending letters and postcards from the students' towns, with attached drawings. They did not require language ability higher than around A2 on the CEFR scale, which was a fairly typical upper level for the students involved in the project. Later, the tasks became more narrative in nature, such as describing one's summer holidays. Thus, students were able to demonstrate their ability as far as B1, or slightly beyond.

The students wrote three or four tasks per year. Guidelines, with rough procedural steps, were prepared for the teachers. The students were first involved in the pre-writing phase, in the form of classroom discussion and/or Power Point presentations, which helped to activate students' schematic knowledge. The pre-writing stage requires more activities on the activation of the schematic knowledge than the other two stages: the writing stage and the post-writing stage (revising and editing). In the pre-writing pro-

cess, the teacher should consciously activate the students' content and formal schemata (Zheng/Dai 2012: 86). After the first stage, the students received feedback and guidance from their teachers, and revised their texts to make them suitable to be sent to students from another country—for example, the Norwegian students sent their texts to the Slovene students, the Slovenes sent theirs to the Spanish, the Spanish to the Lithuanian, etc. Thus, as well as being a potential source of pleasure and discovery, writing can be a major source of language development. The actual assessment of the scripts was undertaken by the students' own teacher and a corresponding expert.

6.3 Revision of the assessment scale and feedback profile

The revision of the scale of descriptors was the other major task of the AYLLIT project (see Appendix 1). The most significant revisions occurred as a result of analysing individual students' writings. Sets of three or four scripts were collected longitudinally, from a large number of students over a two-year period. A selection was then made of several of these sets, representing different students, countries and relative levels. The texts were then closely analysed, with the team members constantly referring to the drafted descriptors, and trying to answer the question, *What has Student A demonstrated in his/her most recent text that s/he did not demonstrate in the previous text?* In this way, valuable insight was gained into the development of the individual student's writing ability and his/her language progress.

In revising the scale, other materials were used, including school curricula, comments collected from teachers, and the team members' own experiences in using the descriptors. It was also essential to ensure that the essence of the CEFR levels was preserved. Similar to the findings by Papp and Salamoura (2009: 17), it was identified that a number of students were only able to copy words or write phonetically (see Figure 1) and did not satisfy the criteria for the A1 level. It was therefore necessary to introduce a new level labelled "Approaching A1", which in some other educational contexts is referred to as the pre-A1 level (Negishi/Takada/Tono 2012).

| |
|--|
| MAJNEJMIZ XXXX (a boy) AJLIV IN XXXX AJM 10 JERZ OLD AJM IN 4 KLAS AJHEB 1BRADR END 1 SISTER AJHEB PEC: 1DOG, 2 KEC, 6 BRC IND 4 FIŠ. |
|--|

Figure 1: Example of a feedback profile

In order to give appropriate feedback, teachers need to be aware of the assessment criteria and learning goals. They also have to understand how to recognise and judge what constitutes writing ability, how students develop in writing, and how to use this feedback in such a way that it will actively help students to improve. Moreover,

teachers need to be able to assess the overall level of students' writing ability, so that students can see how they are progressing. In the AYLLIT project, teachers were asked to decide on a rough level, and only refer to the part of the scale that extended slightly above and below the selected level. It was recommended that the teacher shade all of the descriptors that seemed to apply to the student's script, in order to construct a writing profile that demonstrated the student's writing abilities.

By being presented with only the relevant part of the scale, the student was able to observe the degree to which s/he had developed his/her writing skills, compare his/her own writings, and identify where s/he was heading, without being pressured by the group's achievement. This profile was intended to be used as a basis for giving feedback to students, and making learners aware of the *success criteria* (Faxon-Mills/Hamilton/Rudnick/Stecher 2013: 419). The feedback was intended to reflect the four scale criteria (Overall Structure and Range of Information, Sentence Structure and Grammatical Accuracy, Vocabulary and Choice of Phrase, and Misformed Words and Punctuation) and to draw the students' attention to what they could already do, and to what further work remained to be done. Teachers were also strongly encouraged to provide feedback, in spoken interaction with the student, in the most encouraging and positive way. This is in line with a study carried out by Bitchener and Ferris (2012), which found that the combination of written and conference feedback had a significant effect on the accuracy levels of specific grammatical structures. Furthermore, Fluckiger/Vigil/Pasco/Danielson (2010) claim that such feedback is typically formative and, as such, is intended to help students to develop, not merely to grade their performance in a task. The absence of a summative grade can reduce student anxiety and encourage risk-taking, as students perceive their errors merely as part of a work in progress. In addition, teachers were advised to give the student corrective communicative tasks related to the key weaknesses disclosed. A sample of writing, accompanied by its profile and written feedback, is given in Figure 2.

Summer Holiday (a girl)

This is about my summer holiday. First i travelled to xxx (a city) in xxx (a country), for one week. I travelled with my mom, dad and my hamster. But then we found out that we couldn't take the hamster with us to Denmark. But fortunately we found a nice girl who worked in the animal hospital. She offered to take care of my hamster for one week, while we were in Denmark. We travelled with car and boat to Denmark. We rented a holiday house in Denmark. It was a nice house. After one or two days we drove to a beautiful beach. It was very windy. It is not mountains in Denmark so the wind just blew everywhere. Then we went to Legoland. It was so incredible! Many LEGO houses So cool! And a big, cool Rollercoaster. It rained that day so I didn't do so much. Then we went to Odense zoo. It was fun but the animals had to little space to walk and play! And after a while we travelled to Germany. Just for a short visit. Then at the last day in Denmark we went to see the famous Moonfish. Then we travelled back to Bergen.

From xxx

| Levels | Overall Structure and Range of Information | Sentence Structure and Grammatical Accuracy | Vocabulary and Choice of Phrase | Misformed Words and Punctuation |
|----------|--|--|---|--|
| Above B1 | Is able to create quite complicated texts, using effects such as switching tense and interspersing dialogue with ease. The more common linking words are used quite skilfully. | Sentences can contain a wide variety of clause types, with frequent complex clauses. Errors in basic grammar only occur from time to time. | Vocabulary may be very wide, although the range is not generally sufficient to allow stylistic choices to be made. | Misformed words only occur from time to time. |
| B1 | Is able to write texts on themes which do not necessarily draw only on personal experience and where the message has some complication. Common linking words are used. | Is able to create quite long and varied sentences with complex phrases, e.g., adverbials. Basic grammar is more often correct than not. | Vocabulary is generally made up of frequent words and phrases, but this does not seem to restrict the message. Some idiomatic phrases used appropriately. | Most clauses do not contain misformed words, even when the text contains a wide variety and quantity of words. |
| A2/B1 | Is able to make a reasonable attempt at texts on familiar themes that are not completely straightforward, including very simple narratives. Clauses are normally linked using connectors, such as <i>and, then, because, but</i> . | Sentences contain some longer clauses, and signs are shown of awareness of basic grammar, including a range of tenses. | Vocabulary is made up of very common words, but is able to combine words and phrases to add colour and interest to the message (e.g., using adjectives). | Clear evidence of awareness of some spelling and punctuation rules, but misformed words may occur in most sentences in more independent texts. |

This is quite a long narrative text, which has complicating factors, such as the episode with the hamster and how it was resolved. There is good linking, e.g., *after a while*, including the use of adverbs such as *fortunately*. We get a clear sense of what happened and her reactions, including her reservations: *It was fun but the animals had to little space*. She provides reasons for things: *It is not mountains in Denmark so the wind just blew everywhere*. Her grammar is generally correct, apart from *it/there* error, and *travelled with car*. The text lacks a certain fluency, with many very short sentences which are not well linked to the adjacent ones. The vocabulary seems sufficient to allow her to fully tell her story, and there are a few quite idiomatic phrases, such as *she offered to take care of* Her spelling is good, the only errors being 'i' and *to* (too).

Figure 2: Sample script for B1 level, example of profile and feedback form

The actual assessment process of students' scripts was performed by teachers and experts. All scripts were assessed by a teacher, team member and coordinator, independently. It should be noted that the difference between levels assigned to a student's script rarely exceeded half a CEFR level, or one level in the scale. As Hasselgreen (2013: 426) notes, "Any bigger differences tended to be sporadic rather than systematic, and the three raters were all given access to each other's ratings, which acted as a form of training for all involved".

In conjunction with this, a workshop was organised, attended by 30 participants from 30 countries, all of whom were directly involved in primary school language education. The focus of the workshop was to validate the scale of descriptors, discover its potential usefulness in assessing texts, and try out its appropriateness as a basis for providing feedback. The participants were asked to deliver texts written by their students. Working in groups of five, the participants were first familiarised with the CEFR. They were asked to assign isolated AYLLIT descriptors to the levels set by the AYLLIT writing assessment scale. The participants agreed with the levels assigned to the descriptors by the AYLLIT team and, thus, this activity served as a validating procedure of the descriptors/levels, as they all proved to be recognisable as belonging to the intended CEFR levels.

Next, the participants were asked to assign seven texts (selected as benchmarks) to each of the AYLLIT levels, thus relating the descriptors to real texts. It was clear that the participants mostly agreed with the levels assigned by the AYLLIT team, as the overall levels never differed by more than one level above or below the level assigned by the AYLLIT team. This activity was followed by the participants working in groups with their own texts, and assigning them to the AYLLIT levels. They found this activity very useful and were able to identify appropriate descriptors in the AYLLIT scale that mirrored their students' achievement in writing. Prior to the central workshop, an online workshop took place, in which participants, with no training other than reading the material provided, rated scripts according to the AYLLIT levels. It was not surprising that there was little agreement in rating the scripts, which underlines the importance of training teachers in the use of assessment scales (Becker/Pomplun 2006: 720).

The second part of the workshop aimed at providing feedback using the AYLLIT profiles. The AYLLIT team first designed samples of AYLLIT feedback (eight scripts with feedback), after which the samples were discussed in smaller groups of participants. The discussion within the groups proved beneficial in composing the final version of the AYLLIT scale, feedback profiles and guidelines.

7 OUTCOMES OF THE AYLITT PROJECT

The outcomes of the AYLLIT project consist of assessment material and guidelines for its use (Hasselgreen/Kaledaite/Pižorn/Maldonado 2011 and the ECML/AYLLIT project website [AYLLIT, 2007–2011]). The key achievement is the scale of descriptors (Appendix 1), accompanied on the website by eight sample texts ranging from pre-A1 to above B1 levels. Each text is linked to its feedback profile. The guidelines

for assessing writing are found in Chapter 2 of the handbook (Hasselgreen/Kaledaite/Pizorn/Maldonado 2011), where teachers can find information on the assessment of young language learners' literacy, writing processes in primary school, their own needs regarding the assessment of writing, and the use of the materials and methods in the classroom. Teachers can learn how to construct a profile of the student's writing based on the AYLLIT scale, how to use this profile to stimulate learners to improve their writing abilities, how to give corrective feedback (see Figure 3), and how to use the criteria in self-assessment. As experienced teacher trainers themselves, AYLLIT team members believe that many teachers prefer face-to-face training. Thus, a step-by-step guide for teacher trainers, who wish to give workshops to novice and inexperienced teachers, is available as part of the online downloadable handbook.

| | |
|---|---|
| <p>Summy!</p> <p>My summer holiday.</p> <p>Aim hvas in Mallorca and am sunbrathling, that was very fun! That was a experienle of the live, and <u>am stay</u> as a camping place, <u>wit</u> my Grandmum and my Grandad, and we fising and <u>have</u> fun that summer. We also <u>play Gitar</u> and Singing and 1 day we <u>go</u> to shopping I <u>don't buy</u> so much.</p> | |
| <p>1. Spelling: copy these words carefully</p> <p>Summer Was Fishing Guitar Buy With</p> <p>Now correct the spelling of the words shaded in your text.</p> | <p>2. Grammar:</p> <p>When we tell about things that happened at a time in the past, we use the past tense of verbs. The underlined verbs in the text should be in the past tense. Find the past tense of these verbs and write them in the phrases below. The first one is done for you.</p> <p>We have "We had" I am We stay I play We go I don't</p> <p>Now correct all the verbs underlined in your text.</p> |

Figure 3: Example of corrective feedback (Hasselgreen/Kaledaite/Pizorn/Maldonado 2011: 30)

8 DISCUSSION

The decisive question is whether the AYLLIT outcomes enable teachers to assess their students' writing skills in a valid way – to stimulate further learning. Stobart (2006: 141) identifies five factors that must be established, in order to provide valid feedback. The first factor refers to the clear linking of feedback to the learning goals/intentions. Here the feedback is based on the AYLLIT scale, which consists of four criteria: (1) overall Structure and range of information; (2) sentence structure and grammatical accuracy; (3) vocabulary and choice of phrase; and (4) misformed words and punctuation. These criteria are generally recognised in the literature on (assessing) writing (Weigle 2002; Lee 2007). Moreover, the progression within the AYLLIT scale and its levels is derived from the descriptors in the CEFR, as well as being closely linked to the goals for writing described in the curricula of the four countries concerned. Therefore, the feedback can be regarded as clearly linked to the goals/intention of learning to write, due to the perception of writing in the research literature, its roots in the CEFR and its close linkage to curricular goals.

The second factor refers to the requirement that a learner should understand the success criteria. Stobart (2012: 236) adds another dimension to the comprehension of the criteria: in order for formative assessment to lead to learning, the classroom climate has to be supportive. Some of the AYLLIT project teachers reported that learners needed some time to become accustomed to feedback profiles using the AYLLIT scale. However, after a few months' experience, and with oral support and encouragement from the teacher, in the form of so-called oral conferences (Bitchener/Ferris 2012), students gained a deeper insight into the individual descriptors, and were able to identify their own strengths and weaknesses, as well as setting their own individual writing goals, to bridge the gaps between the levels. In addition, they were highly motivated by writing texts for "real" people who would read and respond to their texts.

Valid feedback should also give signals at appropriate levels, and indicate how to bridge the gap between one level and another. Stobart's (2006) requirements include the following cues: metacognitive, deep learning and task learning. All three cues are part of the corrective feedback tasks prepared by the teacher, and completed by students on their own (see Figure 3). Students need to be cognitively-involved in the task, in order to accomplish it successfully. A number of deep learning processes are necessary: for example, understanding and applying the grammatical rule about forming the past tense, identifying and correcting spelling mistakes, paying attention to certain ways of spelling words, etc.

The next factor underlying the provision of valid feedback concerns whether feedback predominantly focuses on the task or on the learner. The AYLLIT feedback profile is distinctly focused on the task itself (see Figure 2). The feedback is based on the descriptors used in the AYLLIT scale (evidence- and criterion-based assessment). Such feedback encourages teachers to compare students' writing products with the criteria, and not with other students. It also supports teachers and students in providing evidence and arguments for their decisions: for example, in assigning a certain level to the student's overall writing ability. It is recommended that teachers apply overall

levels with great caution (Hasselgreen/Kaledaite/Pižorn/Maldonado 2011: 27–30), as writing is a complex process, and a one-off product cannot reliably indicate which (AYLLIT/CEFR) level the student has reached. There are further factors influencing the writing process that teachers should be aware of, such as students' motivation, task purpose (authenticity), task cognitive demands, the background knowledge required to complete the task, vocabulary and grammar knowledge, classroom climate, the status of language, L1 literacy development, etc.

The final (fifth) factor in establishing valid feedback refers to the following: *Feedback needs to be challenging, it should require action, and it has to be achievable*. As can be seen in Figure 3, when students are given corrective feedback, rather than receiving it passively, they should act upon it. They need to understand, apply the rules, correct, identify and check relevant linguistic forms, etc. This requires students to work on their errors. Developmental psychologist, Reuven Feuerstein, and his colleagues (1980, 2006) have indicated that dealing with error should be seen as a mark of respect for the learner. Errors cannot be viewed solely as failures: their origin and reason must be pursued.

In doing so the teachers demonstrate their respect for the student as a thinking being who has arrived at a response through reasons that may not correspond to the task, but which, nonetheless, exist and must be explored. (Feuerstein/Feuerstein/Falik/Rand 2006: 353).

Furthermore, while reading the shaded descriptors that the teacher has highlighted in the AYLLIT scale, students need to be able to relate them to their own writing, and find evidence for the descriptors selected.

Thus, the feedback resulting from the materials and guidelines of the AYLLIT project can be regarded as achieving the five conditions proposed by Stobart (2006) for establishing validity. They also follow the general principles of assessment for learning, established by Faxon-Mills, Hamilton, Rudnick and Stecher (2013), which presuppose the provision of helpful and constructive feedback (the AYLLIT scale descriptors, samples of corrective feedback), active involvement of students in their own learning (the revision of corrective feedback, self-assessment of writing abilities), teacher adjustments to future instruction based on the outcomes of the results of the assessment (the guidelines for assessing writing produced in the AYLLIT project), and making students aware of the success criteria needed to do well in the assessment activity (students' comprehension of the AYLLIT scale descriptors). Moreover, the materials and guidelines of the AYLLIT project, if applied appropriately do, in fact, promote assessment for learning, as they also support the development of students' metacognitive and linguistic skills.

References

- AYLLIT (2007–2011) *Assessment of young learner literacy, linked to the CEFR*. Graz, Austria: European Centre for Modern Languages. 30 May 2014. <http://AYLLIT.ecml.at>.
- BECKER, Douglas/Mark POMPLUN (2006) “Technical reporting and documentation.” In: S. Downing/T. Haladyna (eds), *Handbook of test development*. Mahwah/ New Jersey/London: Lawrence Erlbaum, 711–723.
- BITCHENER, John/Dana R. FERRIS (2012) *Written Corrective Feedback in Second Language Acquisition and Writing*. New York/London: Taylor and Francis.
- BLACK, Paul/Christine HARRISON/Clare LEE/Bethan MARSHALL/Dylan WILIAM (2004) “Working inside the Black Box: Assessment for Learning in the Classroom.” *Phi Delta Kappan* 86/1, 8–21.
- BLACK, Paul/Dylan WILIAM (1998) “Assessment and Classroom Learning.” *Assessment in Education: Principles, Policy and Practice* 5/1, 7–71.
- CAMERON, Lynne (2001) *Teaching languages to young learners*. Cambridge, UK: Cambridge University Press.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (2001). Cambridge: Cambridge University Press.
- DEPARTMENT FOR EDUCATION AND SKILLS (2002) *Languages for All: Languages for Life. A Strategy for England*. London: DfES. 9 June 2014. <http://www.teachernet.gov.uk/docbank/index.cfm?id%411879>.
- EVANS, Michael/Linda FISHER (2012) “Emergent communities of practice: secondary schools' interaction with primary school foreign language teaching and learning.” *The Language Learning Journal* 40/2, 157–173.
- FAXON-MILLS, Susannah/Laura S. HAMILTON/Mollie RUDNICK/Brian M. STECHER (2013) *New Assessments, Better Instruction? Designing Assessment Systems to Promote Instructional Improvement*. Santa Monica et al.: Rand Corporation. 10 June 2014. http://www.rand.org/content/dam/rand/pubs/research_reports/RR300/RR354/RAND_RR354.pdf.
- FEUERSTEIN, Reuven (1980) *Instrumental Enrichment: an Intervention Program for Cognitive Modifiability*. Baltimore: University Park Press.
- FEUERSTEIN, Reuven/Raphael S. FEUERSTEIN/Louis FALIK/Yaacov RAND (2006) *Creating and Enhancing Cognitive Modifiability: the Feuerstein Instrumental Enrichment Program*. Jerusalem: ICELP Publications.
- FLUCKIGER, Jarene/Yvonne Tixier y VIGIL/Rebecca PASCO/Kathy DANIELSON (2010) “Formative Feedback: Involving Students as Partners in Assessment to Enhance Learning.” *College Teaching* 58/4, 136–140.
- GOTO BUTLER, Yuko/Jiyoon LEE (2010) “The effects of self-assessment among young learners of English.” *Language Testing* 27/5, 5–31.
- HAMID, M. Obaidul (2010) “Globalisation, English for everyone and English teacher capacity: Language policy discourses and realities in Bangladesh.” *Current Issues in Language Planning* 11/4, 289–310.
- HASSELGREEN, Angela/Cecilie CARLSEN/Hildegunn HELNESS (2004) *European survey of language testing and assessment needs*. 1 June 2014. <http://www.ealta.eu.org/resources.htm>.

- HASSELGREEN, Angela/Violeta KALEDAITE/Karmen PIŽORN/Martin Natalia MALDONADO (2011) *AYLLIT – Assessment of young learner literacy linked to the Common European Framework of Reference for Languages*. Graz, Austria: ECML Publications. 11 June 2014. <http://ayllit.ecml.at/Handbook/tabid/2507/language/en-GB/Default.aspx>.
- HASSELGREEN, Angela (2013) “Adapting the CEFR for the Classroom Assessment of Young Learners’ Writing.” *The Canadian Modern Language Review/La revue canadienne des langues vivantes* 69/4, 415–435.
- HATTIE, John (2012) “Know Thy IMPACT.” *Educational Leadership* 70/1, 18–23.
- HELNESS, Hildegunn Lahlum (2012) “A comparative study of the vocabulary of 7th and 10th graders in scripts from the National Tests of Writing in English.” In: A. Hasselgreen/I. Drew/B. Sørheim (eds), *The Young Language Learner*. Bergen: Fagbokforlaget, 145–158.
- LEE, Icy (2007) “Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom.” *Canadian Modern Language Review-revue/Canadienne Des Langues Vivantes* 64/1, 199–213.
- LITTLE, David (2006) “The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact.” *Language Teaching* 39, 167–190.
- LITTLE, David (2007) “The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy.” *Modern Language Journal* 91, 645–655.
- MCKAY, Penny (2006) *Assessing Young Language Learners*. Cambridge: Cambridge University Press.
- MURPHY, Patricia (ed.) (1999) *Learners, Learning and Assessment*. London: Paul Chapman Publishing in association with The Open University.
- NEGISHI, Masashi/Tomoko TAKADA/Yukio TONO (2012) “A progress report on the development of the CEFR-J.” *Studies in Language Testing* 36, 137–165.
- NGUYEN, Thi Minh Hoa (2011) “Primary English language education policy in Vietnam: Insights from implementation.” *Current Issues in Language Planning* 12/2, 225–249.
- PAPP, Szilvia/Angeliki SALAMOURA (2009) “An exploratory study linking young learners examinations to the CEFR.” *Research Notes* 37, 15–22.
- PHYAK, Prem Bahadur (2011) “Beyond the façade of language planning for Nepalese primary education: Monolingual hangover, elitism and displacement of local languages.” *Current Issues in Language Planning* 12/2, 265–287.
- PIŽORN, Karmen (2009) “Designing proficiency levels for English for primary and secondary school students and the impact of the CEFR.” In: N. Figueras/J. Noijons (eds), *Linking to the CEFR levels: research perspectives*. Arnhem: Cito, Institute for Educational Measurement; Council of Europe; European Association for Language Testing and Assessment, 87–101.
- STOBART, Gordon (2006) “The validity of formative assessment.” In: J. Gardner (ed.), *Assessment and Learning*. London: Sage, 133–146.

- STOBART, Gordon (2012) "Validity in formative assessment." In: J. Gardner (ed.), *Assessment and Learning*. London: Sage, 233–242.
- WEIGLE, Sacha Cushing (2002) *Assessing writing*. Cambridge, UK: Cambridge University Press.
- WILFORD, Sara (2000) *From play to literacy: Implications for the classroom* (Report No. PS028746). Bronxville, NY: Child Development Institute, Sarah Lawrence College. 8 June 2014. http://www.slc.edu/cdi/media/pdf/Occasional%20Papers/CDI_Occasional_Paper_Wilford_2000.pdf.
- WILIAM, Dylan (2011) *Embedded Formative Assessment*. Bloomington: Solution Tree Press.
- ZHENG Shigao/Weiping DAI (2012) "Studies and Suggestions on Pre-writing Activities." *Higher Education Studies* 2/1, 79–87.

Abstract

THE DEVELOPMENT OF A CEFR-BASED SCALE FOR ASSESSING YOUNG FOREIGN LANGUAGE LEARNERS' WRITING SKILLS

The Common European Framework of Reference (CEFR) was designed with adults in mind, which is clearly reflected in the six levels encompassing a range of proficiency that represents lifelong learning. Therefore, any use of the CEFR levels as a basis for describing the ability of young learners requires adapting the content of each level, as well as identifying which levels on the scale are appropriate for children.

The present article examines the contribution that feedback, in the form of an assessment scale, can make to valid classroom assessment of the writing of young learners, in the age group of 9–13 years. It shows that a scale of descriptors adapted from the CEFR can play a central role in this assessment. The article presents the AYLLIT (Assessment of Young Learner Literacy) research project, which developed a CEFR-based writing scale and guidelines for teachers, enabling them to provide their students with feedback, and to gain a clearer insight into their students' progress. After describing the procedures followed in the project, the article examines the extent to which its outcomes may enable teachers to give feedback that could contribute to valid classroom assessment.

Keywords: language assessment, young learners, writing skills, CEFR.

Povzetek
RAZVOJ OCENJEVALNE LESTVICE, PRILAGOJENE NA OKVIR SEJO
ZA OCENJEVANJE PISNIH SPRETNOSTI MLAJŠIH UČENCEV

Ko so oblikovali Skupni evropski jezikovni okvir (SEJO), so imeli v mislih predvsem odrasle uporabnike (tujega) jezika, kar se odraža tudi v šestih ravneh opisnikov, ki se nanašajo na znanja, ki jih posameznik pridobi v času vseživljenjskega učenja. Zato je prilagajanje vsebine in števila SEJO ravni (opisnikov) za potrebe mlajših učencev nujno.

Članek preučuje možen prispevek povratne informacije v obliki ocenjevalne lestvice pri jezikovnem preverjanju/ocenjevanju v razredu mlajših učencev v starosti 9 do 13 let.

Avtorica predoči, da lahko opisniki kot del ocenjevalne lestvice, prilagojene na SEJO, igrajo pomembno vlogo pri preverjanju/ocenjevanju znanja v razredu. V članku je predstavljen raziskovalni projekt AYLLIT (and. Assessment of Young Learner Literacy; slo. Preverjanje/ocenjevanje pismenosti mlajših učencev jezika), v okviru katerega so strokovnjaki razvili ocenjevalno lestvico za preverjanje pisne spretnosti in jo prilagodili na ravni in opisnike SEJO ter oblikovali smernice za učitelje. Ti so s tem pridobili orodje za podajanje povratne informacije, ki naj bi jih opolnomočila pri prepoznavanju učenčevega napredka. Avtorica opiše postopek raziskave in evalvira možnosti uporabe lestvice kot pomoč učiteljem pri dajanju povratnih informacij učencem pri pisnih izdelkih, ki lahko izboljšajo veljavnost preverjanja in ocenjevanja pisnih spretnosti v razredu.

Ključne besede: jezikovno preverjanje/ocenjevanje, mlajši učenci, pisne spretnosti, SEJO.

APPENDIX 1: AYLLIT scale of descriptors (AYLLIT, 2007–2011)

| Levels | Overall Structure and Range of Information | Sentence Structure and Grammatical Accuracy | Vocabulary and Choice of Phrase | Misformed Words and Punctuation |
|---------------|--|--|---|--|
| Above B1 | Is able to create quite complicated texts, using effects such as switching tense and interspersing dialogue with ease. The more common linking words are used quite skilfully. | Sentences can contain a wide variety of clause types, with frequent complex clauses. Errors in basic grammar only occur from time to time. | Vocabulary may be very wide, although the range is not generally sufficient to allow stylistic choices to be made. | Misformed words only occur from time to time. |
| B1 | Is able to write texts on themes which do not necessarily draw only on personal experience and where the message has some complication. Common linking words are used. | Is able to create quite long and varied sentences with complex phrases, e.g., adverbials. Basic grammar is more often correct than not. | Vocabulary is generally made up of frequent words and phrases, but this does not seem to restrict the message. Some idiomatic phrases used appropriately. | Most sentences do not contain misformed words, even when the text contains a wide variety and quantity of words. |
| A2/B1 | Is able to make a reasonable attempt at texts on familiar themes that are not completely straightforward, including very simple narratives. Clauses are normally linked using connectors, such as <i>and</i> , <i>then</i> , <i>because</i> , <i>but</i> . | Sentences contain some longer clauses, and signs are shown of awareness of basic grammar, including a range of tenses. | Vocabulary is made up of very common words, but is able to combine words and phrases to add colour and interest to the message (e.g., using adjectives). | Clear evidence of awareness of some spelling and punctuation rules, but misformed words may occur in most sentences in more independent texts. |

| Levels | Overall Structure and Range of Information | Sentence Structure and Grammatical Accuracy | Vocabulary and Choice of Phrase | Misformed Words and Punctuation |
|----------------|--|--|--|--|
| A2 | Can write short straightforward coherent texts on very familiar themes. A variety of ideas are presented with some logical linking. | Is able to make simple independent sentences with a limited number of underlying structures. | Vocabulary is made up of very frequent words but has sufficient words and phrases to get across the essentials of the message aspired to. | Some evidence of knowledge of simple punctuation rules, and the independent spelling of very common words. |
| A1/A2 | Can adapt and build on a few learnt patterns to make a series of short and simple sentences. This may be a short description or a set of related facts on a very familiar personal theme. | | Can use some words which may resemble L1, but on the whole the message is recognisable to a reader who does not know the L1. Spelling may be influenced by the sound of the word and mother tongue spelling conventions. | |
| A1 | Can write a small number of very familiar or copied words and phrases and very simple (pre-learnt) sentence patterns, usually in an easily recognisable way. The spelling often reflects the sound of the word and mother tongue spelling conventions. | | | |
| Approaching A1 | Makes an attempt to write some words and phrases, but needs support or model to do this correctly. | | | |