Adapting the CEFR for the Classroom Assessment of Young Learners' Writing

Angela Hasselgreen

Abstract: This article concerns the contribution that feedback makes to valid classroom assessment of the writing of young learners (YLs), defined here as approximately 9–13 years old. It shows that a scale of descriptors adapted from the *Common European Framework of Reference for Languages* (CEFR; Council of Europe, 2001) can play a central role in this assessment. The article presents a research project, AYLLIT (Assessment of Young Learner Literacy), which developed a CEFR-based instrument and guidelines for teachers, to provide their pupils with feedback that may allow them to see progress and lead to further learning. It sets out the procedures followed in the project and examines the extent to which its outcomes enable teachers to give feedback that contributes to valid classroom assessment.

Keywords: classroom assessment, CEFR, feedback, validity, young learners

Résumé : Cet article souligne l'importance de la rétroaction en relation avec l'apprentissage de l'écriture de jeunes apprenants en salle de classe et le rôle central que le CECR peut jouer dans cette dernière. En effet, le CECR présente un cadre qui aide à juger la validité de la rétroaction dans l'évaluation de l'apprentissage. Cet article présente donc un projet de recherche (évaluation de l'alphabétisation de jeunes apprenants) qui vise à développer un instrument d'évaluation basé sur le CECR et un manuel d'accompagnement pour les professeur(e)s qui puisse les guider dans la rétroaction qu'ils doivent donner à leurs apprenants. Cette rétroaction cherche à développer l'apprentissage des apprenants et à leur permettre d'apprécier leur progrès. L'article présente les démarches qui ont été suivies dans la réalisation du projet et évalue si les rétroactions des professeur(e)s, basées sur le CECR, peuvent réellement contribuer à une évaluation valide en salle de classe.

Mots clés : évaluation en salle de classe, CECR, rétroaction, validité, jeunes apprenants

© 2013 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*,
69, 4, (November / novembre), 415–435 doi:10.3138/cmlr.1705.415

Classroom assessment and valid feedback

Before presenting the Assessment of Young Learner Literacy (AYLLIT) project on which this article reports, it is necessary to establish what is understood here by the term *classroom assessment* and what constitutes valid feedback. For the purposes of the article, *classroom assessment* and *teacher assessment* will be regarded as synonymous. Harlen and Gardner (2010) offer the following definition:

> Teacher assessment comprises a large collection of information gleaned from the daily classroom interactions between pupils and teachers, and between pupils and pupils. The interactions cover many different types of process including the dynamic assessments of questioning and feedback, the reflective assessments of self- and peer-assessment, the sharing of learning goals and the criteria that indicate their achievement, and the long-term progression-related evidence from pupils' work. Such a wealth of evidence is primarily used in an ad hoc support of learning "in the moment" (assessments for formative purposes) but can also be captured in suitable forms for reporting progress and performance (assessments for summative judgements). (p. 7)

According to this definition, there is great diversity in what makes up classroom assessment – its elements, the methods by which information is collected, and the persons involved in the process. While its primary purpose may be formative, particularly in the case of young learners (YLs), it can also contribute to summative assessment. This is relevant to this article insofar as the material produced for teachers in the AYLLIT project was intended to shed light on the general level and progress indicated by pupils' performance as well as lead to further learning. The view taken here is that the ultimate aim of any assessment involving YLs should be to lead to further learning.

In his study of validity in classroom assessment, Stobart (2006) explores conditions that may undermine validity, that is, may prevent the assessment from leading to further learning, and highlights the quality of feedback as crucial. He comments, "If feedback is defined in terms of 'closing the gap' between actual and desired performance then the key consequential validity issue is whether this has occurred" (p. 141). Stobart continues,

> For feedback in the classroom, the following play an important role in the establishment of valid feedback:
>
> - It is clearly linked to the learning intention;
> - The learner understands the success criteria/standard;

- It gives cues at appropriate levels on how to bridge the gap:
  a) self-regulatory/metacognitive
  b) process/deep learning
  c) task/surface learning;
- It focuses on the task rather than the learner (self/ego);
- It challenges, requires action, and is achievable. (p. 141)

These five criteria will be used as a framework for judging the potential validity of the feedback resulting from the use of the AYLLIT project material. Any absolute validation would have to establish that learning had actually occurred, which is beyond the scope of this article. Feedback, in the context of the AYLLIT project, consists of the following elements:

- using a scale of descriptors based on the *Common European Framework of Reference for Languages* (CEFR; Council of Europe, 2001) to make a profile of the written text in terms of several factors, principally linguistic;
- giving written, and preferably also oral, comments following the broad structure of the profile, highlighting strengths and selected weaknesses demonstrated in the text;
- giving the pupil corrective work to do on a limited number of selected linguistic weaknesses that have emerged in the text;
- (periodically) using the profile to establish the overall CEFR-linked level of the text.

The AYLLIT project: an introduction

The AYLLIT project was part of the 2008–2011 medium-term program of the European Centre for Modern Languages (ECML). Its overall aim was to develop CEFR-linked material and guidelines for primary school teachers to use in their classroom assessment of pupils' reading and writing in languages other than their L1. Four countries were involved – Norway, Lithuania, Slovenia, and Spain. The pupils' reading and writing development in English was studied longitudinally for two years, and materials for assessing them were piloted by teachers and regularly revised. The material and guidelines were finalized following a workshop that was held in 2010 with participants from 30 European countries.

Research in the AYLLIT project was qualitative, mainly involving a small group of teachers and their pupils and evolving continuously over two years. This imposed limitations on the study; the data are not intended to be statistically viable, and the outcomes are based largely on reflective processes. However, the scale of descriptors used in

the project was founded on a forerunner (see below), which had been subjected to quantitative analysis that suggested it was performing in a satisfactory way, at least for the purposes of non-high-stakes assessment. Moreover, the AYLLIT material was refined and retuned on the basis of critical feedback from key groups of users, to the point where it was felt to be working in the context of classroom assessment. Further research on larger groups of users may enhance the material and would be valuable, particularly if the material should ever be used as the basis for high-stakes assessment.

A full account of the AYLLIT project is given in Hasselgreen, Kaledaite, Pizorn, and Martin (2011). The focus of this article is the subproject of AYLLIT involving writing. An account will be given here of the procedures followed in the project and the outcomes. Finally, the validity of these outcomes will be considered with reference to the framework presented above. Before embarking on this, however, it is necessary to discuss the writing of YLs, the assessment needs of their teachers, and the suitability of the CEFR to the assessment of YLs.

The writing of young learners

To undertake a study of L2 writing in primary school, it is necessary to establish the curricular aims in the countries concerned. In the four project countries, although curricula were found to vary in their detail, there emerged a recognizable common overall aim for L2 writing (in this case English), with children being expected to write in a "communicative" and creative way, on familiar and personal topics. They should be able to describe and to narrate in written texts. This view of writing reflects much of the literature on children's L2 written language development. Pinter (2006) maintains that children in this age group "may begin to see clear reason for writing such as . . . to write their own stories" (p. 77). Cameron (2001, p. 156) advocates the use of writing for real communication – writing letters or simple stories for others to read. Drew and Sørheim (2009) state that children "usually have stories to tell about themselves and the world they live in, which they are keen to share with others" (p. 88).

The pupil's text shown in Figure 1, adapted from Hasselgreen et al. (2011), was a first draft of a story intended to be sent to children in another country. It illustrates that despite linguistic shortcomings, children are able and often happy to share episodes from their lives. With the support of pre-writing activities, assessment, feedback, and rewriting, the exercise of writing, besides being an act of communication, creativity, and discovery, can also be an important source of L2 development.

Summy!
My summar holiday.
Aim hvas in Mallorca and am sunbrathling, that was very fun! That was a experienle of the live, and am stay as a camping place, wit my Grandmum and my Grandad, and we fising and have fun that summer. We also play Gitar and Singing and 1 day we go to shopping I don't bay so much

Figure 1: Sample of pupil's writing

The language assessment needs of primary school teachers

Before considering assessment itself, it is worth considering the YL teachers who are likely to be engaged in this activity. Two groups can be broadly identified. The first are those who actually teach the language concerned: the "language teachers." The second group comprises those who teach in a mainstream language, but who have several pupils for whom this language is an L2. This latter group cannot realistically be expected to have specialized in language assessment, but what of the language teachers? A study by Drew, Oostdam, and van Toorenburg (2007) of primary school language teachers in the Netherlands and Norway showed that a minority of these teachers are actually trained as language specialists. Moreover, a survey carried out by the European Association for Language Testing and Assessment (EALTA) among language teachers and other professionals (Hasselgreen, Carlsen, & Helness, 2004) demonstrated that even teachers trained as language specialists have a great need for training in a wide range of areas of assessment, citing "defining criteria" and "giving feedback" as areas that most urgently need to be addressed. Thus it seems safe to conclude that the teachers involved in the assessment of the L2 language skills of primary school children require assistance and support in identifying criteria for describing and assessing these skills and in giving good feedback.

What makes feedback "good" has already been discussed and presented as a five-point framework. This framework illustrates that an awareness of criteria and learning intentions is fundamental to giving useful feedback. To provide good feedback in the area of L2 writing, teachers need to understand how to recognize and judge what makes up writing ability as pupils develop, and how to use this assessment as a basis for feedback that will actively help pupils to improve. Furthermore, they need to be able to assess the overall level of pupils' writing ability, so that the children can perceive their own progress, and see where they are heading.

Attending to these needs became the primary concern of the AYLLIT writing project. The dominating activity in the project was the empirical work that went into producing a scale of descriptors based on the CEFR, while meeting the needs of upper primary school pupils and their teachers (the resulting scale is presented in Appendix 1). However, it was also essential to present to teachers a way to use the material as a basis for feedback to move learners forward, and indeed to get children to write. Through continual contact with the teachers using the material in the project, and systematic discussion within the team, guidelines on these issues were produced for inclusion in the final material for teachers.

The CEFR and young learners

The CEFR was designed with adults in mind. This is clearly reflected in the wording of the descriptors and the contexts of language use referred to. Moreover, the six levels – A1–C2 – span a range of proficiency that represents lifelong learning, culminating in a level that very few adults would expect to reach. Any use of the CEFR levels as a basis for describing the ability of YLs would necessarily involve adapting the content at each level to the communication of children, as well as identifying which levels on the scale are appropriate for children.

This adaptation of the CEFR for young learners has been tackled in many ways, notably through primary school versions of the European Language Portfolio (ELP) and the CEFR's presence, explicit or implicit, in school curricula. In each of the four project countries the ELP had been introduced at primary school level, and in two of the countries the aims of the national curricula were linked to CEFR levels: Norway indirectly and Lithuania directly (see Hasselgreen et al., 2011). In both of these cases the end of primary schooling was associated with a CEFR level around A2–B1.

At this point it is worth considering the nature of CEFR scales, which can be roughly categorized as either functional or linguistic. Perhaps the most familiar CEFR scales are those presented in the self-assessment grid, which are primarily functional in nature (Council of Europe, 2001, pp. 26–27). However, the CEFR also includes several scales that describe aspects of communicative language ability, such as grammatical competence and vocabulary. The descriptor for A2 grammatical accuracy reads as follows: "Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say" (Council of Europe,

2001, p. 114). The ELP, on the other hand, generally reflects the more functional aspects of language ability, giving the learner a clear sense of what s/he can do with the language, and helping the teacher to see roughly where a pupil "is" with respect to the CEFR levels, and what tasks s/he can be reasonably expected to perform. An example from the Norwegian ELP for 6–12-year-olds at level A2 is: "I can write a postcard or an e-mail telling about myself and how I am doing" (Nasjonalt senter for fremmedspråk i opplæringen, 2011).

For the purposes of classroom assessment, perhaps most particularly in the area of writing, it is necessary to be able to establish not only what tasks a pupil can perform, but also, importantly, how s/he can perform them. A principal aim for this project was therefore to supplement the ELP, with its functional focus, by producing a CEFR-based scale with a linguistic focus. It should be emphasized, however, that in giving writing tasks that allow pupils to demonstrate their language ability, close attention has to be paid to the writing functions a pupil can be expected to perform at the rough level s/he is perceived to be at.

Procedure for developing the AYLLIT material

The AYLLIT material, of which the scale of descriptors is a cornerstone, was developed in three phases. The first of these took place some years before the AYLLIT project, resulting in a scale that was a forerunner of the AYLLIT scale. The second phase was a preliminary project immediately before the AYLLIT project, and the third phase was the AYLLIT project itself.

Phase 1: The forerunner scale

In a project that the author was involved in herself, 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 (see Helness, 2012). One scale described the pupils' ability to perform each of a range of tasks, focusing on the functional aspect of writing. The other scale had a linguistic focus and was not task-specific. It described writing in terms of four categories: textual structure, grammar, words and phrases, and spelling and punctuation, being primarily based on the relevant CEFR scales of descriptors, and reflecting much of what has been written on the assessment of writing (e.g. Weigle, 2002). This scale was the forerunner to the AYLLIT scale.

The NTE scales were developed hand in hand with tasks to be used in testing proficiency in written English of all pupils in Norway in Grades 7, 10, and 11. The linguistic scales for the three cohorts were

identical, except that B2 was set as a ceiling for Grade 7, while C1 was the highest level for the other two grades. The scale had descriptors for each of the four categories given above; these bands of descriptors were only formulated for whole levels (A1, A2, etc.), but shaded areas between these levels were given, indicating A1/A2 and so on, and teachers were asked to use these levels as well as the whole levels in rating the scripts. Teachers were required to decide which descriptor in each category best suited the writing and assign an overall level on the basis of these. Teachers themselves carried out the rating after a one-day training course.

A study of the Norwegian national testing in 2005 was carried out on behalf of the Norwegian State Board of Education (Lie, Hopfenbeck, Ibsen & Turmo, 2005). Samples of the test scripts were graded by external experts (who were also teachers at the grades concerned), and teachers' own ratings of their pupils were compared with the external ratings. In the study's report on the English tests for Grade 10 (N = 201), the inter-rater correlation between experts and teachers was given as 0.81. The percentage of scripts placed on the same level was 55%, while a further 33% differed by only one level (i.e., one half of a CEFR level) (Lie et al., 2005, p. 94). In most of these cases, the teachers gave higher ratings to their own pupils' scripts than those of the external raters (perhaps telling more about teachers than the scale). Thus in 88% of cases there was agreement to within half a CEFR level. For Grade 7 (N = 383), the inter-rater correlation between experts and teachers was considerably lower, at 0.26. However, here again the raters were generally close in their ratings: 34% were in complete agreement, while 40% differed by only half a CEFR level (Lie et al., 2005, p. 97).

A further study (Kavli & Berntsen, 2005) investigated the tests from the perspective of external evaluators. In the case of the NTE writing tests, these evaluators consisted of 77 teachers representing the three school grades involved. In response to the question on the usefulness of the training to use the scale, 22% opted for *very high degree*, 44% for *quite high degree*, 32% for *to some degree*, with only 3% answering *not useful* (Kavli & Berntsen, 2005, p. 13). During their training, teachers often commented that the scale would be very useful for everyday assessment of writing.

Thus the NTE scale used as a forerunner of the AYLLIT scale, while not proven to be suitable for high-stakes testing, has shown itself to have a high degree of near-agreement in placing pupils on a CEFR-based scale, and to be regarded as useful to teachers in assessing writing. However, several issues had to be resolved before this scale could be regarded as a suitable prototype for the AYLLIT project. These issues concerned the levels included and the fact that the descriptors

were primarily for testing (rather than classroom assessment) in both upper primary and secondary schools.

Phase 2: The preliminary project

In 2007–2008, one year before the AYLLIT project, a preliminary project was carried out in Norway. Two Grade 5 classes (10–11 years old) were drafted into this project; they would also take part in the AYLLIT project. One purpose of the preliminary project was to gain experience in what pupils of this age could be expected (and motivated) to write, and what kind of assessment tools teachers would find useful. Another purpose was to change the NTE (forerunner) scale into a form that both teachers and the project leader considered better suited to the classroom assessment of the writing of upper primary school pupils.

On the NTE scale, each band of descriptors represented a whole CEFR level, from A1 to B2. Lie et al. (2005) had reported that only 1% of a sample of 621 pupils were placed by the external raters at B2 on the language scale, and 6% placed at B1/B2. This suggested that although there appear to be some Norwegian pupils at the end of primary school (13 years) for whom B1 is not a ceiling, the level B2 may be cognitively out of the reach of pupils at this age. Furthermore, it was felt that, to provide meaningful feedback and allow progress to be shown, descriptors at in-between levels should be provided.

Thus the scale was reworked with six bands of descriptors: A1, A1/A2, A2, A2/B1, B1, and above B1. The in-between levels were formulated partly by studying the CEFR scales for communicative language ability (Council of Europe, 2001), which in some cases included upper and lower divisions of a level, and partly by intuition. It was also decided to adjust the categories to include some indication of the functions a pupil can 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*. As the levels below A2 represent a "pre-grammatical" stage, with a limited number of words and phrases largely learned as "chunks," it was decided to conflate the categories into one at A1 and two at A1/A2. This work resulted in a prototype for the AYLLIT project scale, which was to be refined in the course of the project.

Phase 3: The AYLLIT project

The AYLLIT project proper started in the autumn of 2008, with a team of four "experts" (using ECML terminology) representing Lithuania, Norway, Slovenia, and Spain, with the author as coordinator. Two

classes and their teachers in each country took part in the project for two years, starting at Grade 5 or 6.

The project consisted of two one-year cycles, with each cycle containing the following elements:

- Team meetings (preceding each year), to include:
  - planning of writing tasks and procedures for the coming year
  - revision of scale of descriptors using longitudinal samples of pupils' writing and comments from users
- Assignment of writing tasks to pupils at regular intervals throughout the year, to be:
  - assessed by teacher and/or team expert
  - revised (usually) and sent to pupils in another country
- Team monitoring activities, such as:
  - observing and collecting comments from teachers using the material
  - assessing samples of writing already assessed by teachers

In addition, a Central Workshop with 30 mainly non-project participants was held immediately after the completion of Year 2, before finalizing the material.

The focus in Year 1 was on familiarizing both pupils and the teachers/experts with their respective roles and responsibilities and on trialling and developing tasks and assessment material. In Year 2 the material was in a near-ready state, so the focus was on using it fully and compiling users' comments to carry out final adjustments and to produce guidelines for using it. The research was qualitative in nature, and the data mainly consisted of pupils' writing, teachers'/experts' comments, and actual assessments of pupils' texts by different raters.

Team meetings – planning and revising material

The team meetings were intended both for planning activities and for working on the material. Designing writing tasks for pupils during the coming year occupied a central role in the meetings. These tasks reflected "can do" statements for the appropriate levels in the countries' ELPs. The tasks designed for the first year were descriptive in nature, such as writing personal introductions and postcards from the pupils' towns. These tasks do not require language ability higher than around A2 on the CEFR scale, which was a fairly typical upper level for the pupils in the first year of the project. In the second year, the tasks became more narrative in nature, for example, recounting summer holidays. These tasks gave pupils rather more scope to demonstrate their ability, as far as B1 or slightly beyond. Pupils at lower levels tended to tackle these tasks in a simple, fairly descriptive way, as is illustrated by the sample shown in Figure 1 (which was assessed at A2).

The revision of the scale of descriptors was the other major task for the team meetings. The final version of this scale is given in Appendix 1. The most significant revisions took place as a result of analyzing pupils' texts. Sets of three or four texts had been collected longitudinally, over one or two years, from a large number of pupils. A selection was made of several of these sets, representing different countries and relative levels. These texts were worked through, pupil by pupil, asking what this pupil demonstrated at level Z (most recent text) that s/he had not demonstrated at level Y (the previous text), and what s/he demonstrated at level Y that s/he had not demonstrated at level X (the earliest text). In this way, valuable insight was gained into what it was that individual pupils had added to their ability as they progressed from level to level. By going through this procedure with pupils at a range of starting and current levels, it was possible to modify the descriptors so they reflected this progression.

In addition, other material was drawn on in revising the scale, including school curricula, comments collected from teachers, and our own experiences in using the descriptors. It was also essential at all times to ensure that the essence of the CEFR levels was preserved. A significant result of the Year 2 meeting was the introduction of a level "Approaching A1," as it was found that there were several pupils who did not, in their earliest texts, satisfy the criteria for A1.

Pupils' writing – procedure and assessment

The writing tasks were carried out by the pupils three or four times a year, using rough guidelines that had been agreed upon for the procedure at the team meeting. The texts were collected after writing and sent to the recipient class. The tasks were always sent to a school in another country, which was very motivating for the children. The actual assessment was undertaken only by the pupils' own teacher or corresponding expert.

The assessment itself, using the descriptor scale, was also intended to follow a procedure that was gradually refined over the two years. It was always carried out on first drafts, and for each text it was recommended that the teacher decide on a rough level and only use that part of the scale which extended a bit below and above this level. The teacher could use the computer to isolate the appropriate part of the scale, then shade all the boxes that seemed to apply to the text, producing a profile of the writing that showed the pupil's strengths and weaknesses.

The profile was to be used as a basis for feedback. To help pupils become familiar with the criteria, feedback was to reflect the four components (columns) of the scale, drawing attention to what the pupil

had done well and a few things that were within the pupil's reach but needed further work. Ideally, feedback was to be given to the pupil through dialogue with the teacher. In addition, teachers were encouraged to give the pupil corrective tasks relating to key weaknesses revealed, for example, grammar, spelling, or word choice. A sample of writing accompanied by its profile and written feedback is given in Appendix 2 and an example of corrective feedback tasks in Appendix 3.

By being presented with only the relevant part of the scale, as shown in Appendix 2, the pupil was able to see how s/he had progressed (e.g., by comparison with earlier profiles) and also what s/he should be aiming at, without getting any impression of how low or high s/he was relative to the group. An overall level might be given by the teacher, based on the level which the profile most closely approximates.

Monitoring activities: analyzing teachers' assessments and collecting comments

The team carried out a variety of monitoring activities. In some countries the expert was involved in the actual assessment process, while in others this was done exclusively by the teachers, yielding two different perspectives on the assessment process. Where the assessment was mainly carried out by the teachers, it was essential that the local expert was available for questions and receptive to comments from the teachers. This feedback on tasks, procedures, and material was a vital source of data to draw on, especially during the team meetings. Another activity carried out by the team involved a comparison of the team's and the teachers' overall assessment of pupils' tasks. This was primarily done to ensure that teachers were generally interpreting the levels in the scale of descriptors in the way the team intended. After each writing round, tasks were independently assessed by teacher, team member, and coordinator. It was notable that the difference between levels assigned to a text rarely exceeded one level on the scale (i.e., half a CEFR level). 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.

The Central Workshop

The Central Workshop was attended by 30 participants who were all directly involved in primary school language education. The focus was on the scale of descriptors, how this related to the CEFR, and the usefulness of the scale in assessing texts and as a basis for feedback. The participants brought texts written by their pupils, in English or French. After a brief familiarization with the CEFR levels, participants worked in groups of five to assign isolated AYLLIT descriptors to

levels/in-between levels of the CEFR. Each group assigned every descriptor to the "correct" level. This was a way of validating the descriptors, as they all proved to be recognizable as belonging to the intended CEFR levels.

Next, the groups were asked to assign texts to levels on the AYLLIT scale of descriptors, starting with the AYLLIT sample texts, selected as benchmarks for each of the AYLLIT levels and moving on to other sets of texts, including the participants' own pupils' texts. This activity showed that the scale could be used with texts at a wide range of levels in two languages. In the case of the AYLLIT sample texts, there was a high degree of consensus among groups: overall levels never differed by more than one level above or below the level assigned by the AYLLIT team. The focus then moved to feedback based on the AYLLIT profiles. Examples of AYLLIT feedback were presented and discussed, followed by group work to decide on appropriate feedback on individual texts. The discussion within the groups proved valuable in compiling the final versions of the material and guidelines.

It should be mentioned here that, before the central workshop, an online workshop (with different participants) took place, in which participants rated tasks with no training other than reading the material provided. This resulted in little agreement in rating scripts, which underlines the importance of training teachers in the use of the scale.

AYLLIT outcomes

The outcomes of the writing part of the AYLLIT project principally consist of assessment material and guidelines for its use (Hasselgreen et al., 2011, and the ECML/AYLLIT project website [AYLLIT, 2007-2011]). Central to the assessment material is the scale of descriptors (Appendix 1), which is accompanied on the website by eight sample texts representing all four countries and a wide range of levels. The texts chosen were all assigned the same levels by the team. Each of the eight texts has a completed profile form, with feedback comments.

The guidelines for using the assessment material address issues such as how to make a profile of the pupil's writing based on the four criteria for writing ability shown in the scale, how to use this profile to give feedback on the strengths and weaknesses revealed in the writing (as in Appendix 2), how to give corrective feedback tasks for pupils to work on (as in Appendix 3), and how to use the criteria in self-assessment. Finally, because the optimal way of initiating teachers in the use of the material appears to be by hands-on training, there is a step-by-step guide for those wishing to stage workshops to introduce the material, either in initial or in-service training for teachers.

Validation of the feedback resulting from the use of the AYLLIT material

Stobart's (2006, p. 141) identification of five factors that play an important role in the establishment of valid feedback has given rise in this article to a framework for assessing the validity of the feedback resulting from the use of the AYLLIT material. These five factors will be considered in turn.

Feedback is clearly linked to the learning intention

The feedback in the AYLLIT assessment is based on the profile (as in Appendix 2), which is made up of four elements (or fewer in the case of the lowest levels). These are *Overall structure and range of information*, *Sentence structure and grammatical accuracy*, *Vocabulary and choice of phrase*, and *Misformed words and punctuation*. These reflect the dimensions commonly recognized in the literature on writing (e.g., Weigle, 2002). Furthermore, the progression within each of these elements in the scale is rooted in the descriptors in the CEFR (Council of Europe, 2001), adapted to the aims for writing described in the curricula of the four countries concerned. Thus the scale can be regarded as representing a recognized way of depicting general writing ability and the way it develops. Feedback on these aspects of writing not only shows what the pupil appears to have mastered, but also indicates what s/he is aiming at through the presence of the next level in the feedback the pupil receives. The feedback can be regarded as being clearly linked to the intention of learning to write as it is perceived in the literature on writing, including the CEFR, and in the aims expressed in the school curricula of the four project countries.

The learner understands the success criteria/standard

This point has to some extent been covered by the discussion of learning intentions, which are inextricably linked to criteria. By using the AYLLIT scale systematically as a basis for feedback, teachers will gradually accustom pupils to regarding their writing in terms of these criteria. It is also recommended in the guidelines that pupils assess their own writing, using broadly similar criteria, as is exemplified in Appendix 4.

Feedback gives cues at appropriate levels on how to bridge the gap

Three levels of cue are cited by Stobart (2006, p. 141): *self-regulatory/metacognitive*, *process/deep learning*, and *task/surface learning*. All three levels are represented in the feedback provided by the AYLLIT assessment. Self-regulatory/metacognitive cues are given through the corrective feedback tasks learners are required to do after getting their

work back (Appendix 3). Process/deep learning and task/surface learning cues are given both in the corrective feedback and to some extent in the written feedback based on the profile. This may, for instance, include help in understanding how and why the past tense is used, as illustrated in the corrective feedback in Appendix 3. Task/surface learning cues are sometimes necessary, for example, in drawing attention to certain misspellings, as shown in the feedback in Appendix 2.

Feedback focuses on the task rather than the learner (self/ego)

The AYLLIT feedback is clearly focused on the task itself. It is worth noting that teachers are urged to use the overall levels with caution. While it is useful and rewarding to see, over time, how a pupil gradually moves up the levels, the level of any individual piece of writing will only be based on evidence on the page, and should not necessarily be interpreted as the level a pupil is at. Moreover, a pupil will probably not be able to give evidence of being, say, at B1, on the basis of a task only requiring A2 ability, such as writing a postcard. Similarly a pupil faced with a task which is too demanding may perform at a level below his/her norm. Thus, to obtain a true assessment of a pupil's capacity, a task should be chosen to elicit what s/he can do around his/her level of ability.

Feedback challenges, requires action, and is achievable

Teachers using the AYLLIT material are urged to give the kind of corrective feedback to tasks as shown in Appendix 3. This requires children to work on their weaknesses. Teachers are strongly advised not to draw attention to all errors, but rather to a limited number of those that are within reach if the pupils are given support. The focus should preferably be on one or two types of error revealed in the writing; an ambitious child should not be demotivated by "red ink." The practice advocated here can, however, run counter to the culture in a country or school, and parents as well as children should be informed of the reason for this practice and prepared for the fact that that some errors will be ignored.

Conclusions

The validation process presented in the previous section has shown that the feedback resulting from the materials and guidelines of the AYLLIT project do, potentially, satisfy the five conditions laid down by Stobart (2006) for establishing validity. Thus, using the material as it is intended is a significant step toward carrying out valid classroom assessment, where the ultimate aim is to move learning forward.

Whether learning does move forward is in the hands of many people, most particularly the learner. What has also been demonstrated is that the CEFR, implemented in primary schools through the ELP, is able to play a direct role in the classroom assessment of young learners' L2 writing.

Correspondence should be addressed to **Angela Hasselgreen**, Faculty of Teacher Education, Bergen University College, 5096 Landås, Bergen, Norway. E-mail: amh@hib.no.

References

AYLLIT. (2007-2011). *Assessment of young learner literacy, linked to the CEFR*. Graz, Austria: European Centre for Modern Languages. http://AYLLIT.ecml.at

Cameron, L. (2001). *Teaching languages to young learners*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511733109

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Drew, I., Oostdam, R., & van Toorenburg, H. (2007). Teachers' experiences and perceptions of primary EFL in Norway and the Netherlands: A comparative study. *European Journal of Teacher Education*, *30*(3), 319–341. http://dx.doi.org/10.1080/02619760701486159

Drew, I., & Sørheim, B. (2009). *English teaching strategies* (2nd ed.). Oslo, Norway: Det Norske Samlaget.

Harlen, W., & Gardner, J. (2010). Assessment to support learning. In J. Gardner, W. Harlen, L. Hayward, & G. Stobart (Eds.), *Developing teacher assessment* (pp. 15–28). Maidenhead, UK: McGraw-Hill Education.

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs*. http://www.ealta.eu.org/resources.htm

Hasselgreen, A., Kaledaite, V., Pizorn, K., & Martin, N.M. (2011). *AYLLIT – Assessment of young learner literacy linked to the Common European Framework of Reference for Languages*. Graz, Austria: ECML Publications.

Helness, H.L. (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* (pp 145-158). Bergen, Norway: Fagbokforlaget.

Kavli, H., & Berntsen, W. (2005). *Kartlegging blant eksternvurderere av nasjonale prøver [A survey of the external evaluation of the national tests]*. Oslo, Norway: Utdanningsdirektoratet.

Lie, S., Hopfenbeck, T.N., Ibsen, E., & Turmo, A. (2005). *Nasjonal prøver på ny prøver [National tests re-tested]*. *Acta Didactica*, 1/2005; special issue. Oslo, Norway: University of Oslo.

Pinter, A. (2006). *Teaching young learners*. Oxford, UK: Oxford University Press.

Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 133–146). London, UK: Sage.

Utdanningsdirektoratet (2009). *Europeisk språkperm [European Language Portfolio] (6-12)*. Oslo, Norway: Author. Retrieved from http://elp.ecml.at/Portals/1/documents/Norway-100-2009-Model-for-young-learners-aged-6-12.pdf_

Weigle, S.C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511732997

Appendix 1: AYLLIT scale of descriptors

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 reasonable attempt at texts on familiar themes that are not completely straightforward, including very simple narratives. Clauses are normally linked using connectors, such as *and, then, because, but.*	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.
A2	Can write short, straightforward,	Is able to make simple	Vocabulary is made up of very	Some evidence of knowledge of

(*continued on next page*)

Appendix 1 (*continued*)

Levels	Overall structure and range of information	Sentence structure and grammatical accuracy	Vocabulary and choice of phrase	Misformed words and punctuation
	coherent texts on very familiar themes. A variety of ideas are presented with some logical linking.	independent sentences with a limited number of underlying structures.	frequent words but has sufficient words and phrases to get across the essentials of the message aspired to.	simple punctuation rules, and the independent spelling of very common words.
A1/A2	Can adapt and build on to a few learnt patterns to make a series of short and simple sentences. This may be a short description or 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 recognizable 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 recognizable 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.			

Appendix 2: Example of profile and feedback form

Levels	Overall structure and range of information	Sentence structure and grammatical accuracy	Vocabulary and choice of phrase	Misformed words and punctuation
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 reasonable attempt at texts on familiar themes that are not completely straightforward,	Sentences contain some longer clauses, and signs are shown of awareness of basic grammar,	Vocabulary is made up of very common words, but is able to combine words and phrases to add colour and interest	Clear evidence of awareness of some spelling and punctuation rules, but misformed words may occur

(*continued on next page*)

Appendix 2 (*continued*)

Levels	Overall structure and range of information	Sentence structure and grammatical accuracy	Vocabulary and choice of phrase	Misformed words and punctuation
	including very simple narratives. Clauses are normally linked using connectors, such as *and, then, because, but.*	including a range of tenses.	to the message (e. g., using adjectives).	in most sentences in more independent texts.
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. Text 6. A2/B1+

This is a colourful introductory text, which covers a range of aspects and which is not completely straightforward. It gives a reason why you have many friends: *because I am a good friend,* and tells something about your future hopes: *My bigest wishes to have a trip to Egipt and to see the piramides.* You also mention what you don't like and haven't got. You use present and past tenses appropriately and correctly. Your sentences are occasionally complex, although mainly short and simple, and the text lacks a certain flow, rather jumping from topic to topic. You use adjectives – *smart mother*, *strong father*, *blue Mediterranean sea*. Your spelling is generally good, with a few misspelt words, which are not usually of the simplest words. However, notice how you have spelt "biggest", "Egypt", and "Pyramids".

Dear Friend, (a girl)

Hello! My name is xxx. I am 11 years old. I have a smart mother and a strong father. I also have a brother. He is 12 years old. I like to play basketball, but I don't like to do my homework. I have many friends, because I am a good friend. I haven't got any pets. I am waiting my summer holiday. Last year I was in Turkey. I lived in a Club hotel Mirabell near blue Mediterranean sea. It was hot, interesting summer. My bigest wishes to have a trip to Egipt and to see the piramides. With best wishes! Xxx

Appendix 3: Example of corrective feedback

***Summy*!**

My *summar* holiday.

Aim hvas in Mallorca and am sunbrathling, that was very fun! That was a experienle of the live, and am stay as a camping place, *wit* my Grandmum and my Grandad, and we *fising* and have fun that summer. We also play *Gitar* and Singing and 1 day we go to shopping I don't *bay* so much

Example of corrective feedback

1. **Spelling**: Copy these words carefully:

Summer
Was
Fishing
Guitar
Buy
With

Then correct the **spelling** of the words in your text marked with italics.

2. **Grammar**:

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.

We have **We had**
I am
We stay
I play
We go
I don't

Now correct all the underlined verbs in your text.

Appendix 4: Extract from self-assessment form for writing

How true are these? Ring round the best number.

(4 = *true*, 3 = *more or less true*, 2 = *partly true*, 1 = *not true*)

I managed to write what I wanted	4	3	2	1
I made a text that suited the task	4	3	2	1
I made a text with a "thread" running through	4	3	2	1
My paragraphs each covered a main point	4	3	2	1
I used words & phrases typical of texts like this	4	3	2	1
I knew enough grammar	4	3	2	1
I checked for spelling, punctuation, "endings," etc.	4	3	2	1
I liked doing this	4	3	2	1
I got help from: (dictionary, internet, etc.)	4	3	2	1