

The “Global Scale of English Learning Objectives for Young Learners”: A CEFR-Based Inventory of Descriptors

Veronica Benigno and John de Jong

Abstract This chapter presents an ongoing project to create the “Global Scale of English Learning Objectives for Young Learners” – CEFR-based functional descriptors ranging from below A1 to high B1 which are tailored to the linguistic and communicative needs of young learners aged 6–14. Building on the CEFR principles, a first set of 120 learning objectives was developed by drawing on a number of ELT sources such as ministry curricula and textbooks. The learning objectives were then assigned a level of difficulty in relation to the CEFR and the Global Scale of English and calibrated by a team of psychometricians using the Rasch model. The objectives were created and validated with the help of thousands of teachers, ELT authors, and language experts worldwide – with the aim to provide a framework to guide learning, teaching, and assessment practice at primary and lower-secondary levels.

Keywords Young learners • Descriptors • Assessment • Teaching • Learning objectives • Can Do Statements • Rating • Scaling • CEFR (Common European Framework of Reference for Languages) • GSE (Global Scale of English) Learning Objectives for Young Learners

1 Introduction

The *Common European Framework of References for Languages (CEFR; Council of Europe, 2001)* was compiled with an adult and young adult audience in mind. Consequently, the majority of descriptors refer to communicative acts performed by learners who are likely to use the foreign language in the real world. The CEFR is

V. Benigno (✉)
Pearson English, Pearson Education, London, UK
e-mail: veronica.benigno@pearson.com

J. de Jong
Pearson Assessment Centre, Pearson Education Inc., Iowa City, IA, USA
Amsterdam VU University, Amsterdam, Netherlands
e-mail: john.dejong@pearson.com

therefore less appropriate for describing proficiency of young learners (YL, primary, and lower secondary learners), and particularly of the youngest ones whose life experience is substantially different from that of adults.

In this chapter we discuss an ongoing project at Pearson English which aims to develop a set of functional descriptors for young learners: the “Global Scale of English Learning Objectives for Young Learners” (Pearson, 2015b; here also referred to as “descriptors” or “learning objectives”). These CEFR-based “Can Do” statements cover the levels from below A1 to high B1 and are tailored to motivations and needs of young learners aged 6–14, a period during which they are still developing linguistic and cognitive skills in their own mother tongue. Level B2 and higher are not taken into account because they assume more adult skills. The CEFR was used as a reference guide to identify valid theoretical and methodological principles for the development and the scaling of the new descriptors.

We believe this work represents a contribution to the ongoing debate on what young learners can do and what instruments can be used to assess their performance. Setting standards requires us to define what learners should be able to do with the language at a certain level of proficiency and how to observe proficiency gains in relation to a defined scale. Standard setting does not imply a prescriptive pedagogy but allows for comparability between curricula based on a definition of extraneous, i.e., non-school, functional learning goals. If standards refer to a common framework they will allow the implementation of a transparent link between content development, teaching, and assessment.

Teaching English to Young Learners (TEYL) has recently received much attention. Under the impact of globalization, the last few decades have seen an increasing tendency to introduce English in primary school curricula around the world, particularly in Europe (Nikolov & Mihaljević Djigunović, 2006; Nikolov, 2016). Nowadays, millions of primary age children learn English in response to parents’ expectations and supported by educational policy makers. There has been an increase not only in the number of young learners and their teachers, but also in the volume of documents about and for young learners: language policy documents, teachers’ handbooks, teaching materials, empirical studies, conference reports and proceedings, and academic publications (Nikolov & Mihaljević Djigunović, 2011). Early language learning policies have been promoted by European institutions since the 1990s (Speitz, 2012). According to the European Commission, early language learning yields a positive impact in terms of education and cross-cultural communication:

Starting to learn a second/foreign language early can help shape children’s overall progress while they are in a highly dynamic and receptive developmental stage in their lives. Starting early also means that learning can take place over a longer period, which may support the achievement of more permanent skills. When the young brain learns languages, it tends to develop an enhanced capacity to learn languages throughout life. (European Commission, 2011, p. 7)

Support of intercultural education is claimed to be among the benefits of early language learning: “raising awareness of language diversity supports intercultural

awareness and helps to convey societal values such as openness to diversity and respect” (European Commission, 2011, p. 7).

It is generally believed that early foreign language (FL) introduction provides substantial benefit to both individuals (in terms of linguistic development, social status, and opportunities) and governments (as a symbol of prestige and economic drive). However, some concerns have been raised about the dangers of inadequate preparation and limited knowledge about who young learners are, how they develop, and what they need. This has led some researchers to argue against the validity of “the earlier the better” hypothesis. Among the most common arguments against this principle are: (a) learning is not exclusively determined by age but also by many other factors, e.g., the effectiveness of teaching; and (b) younger learners have an imprecise mastery of their L1 and poorer cognitive skills in comparison to older learners. Studies on the age factor (e.g., Lightbown & Spada, 2008) have shown that, at least in the early stages of second language development, older learners progress faster than younger ones, questioning the benefit of the early introduction of an FL in the curriculum. Other studies (e.g., Singleton, 1989), however, have argued that early language learning involves implicit learning and leads to higher proficiency in the long run. There is indeed some evidence to support the hypothesis that those who begin learning a second language in childhood in the long run generally achieve higher levels of proficiency than those who begin in later life (Singleton, 1989, p. 137), whereas there is no actual counter evidence to disprove the hypothesis.

It is worth highlighting that “the earlier the better” principle is mainly questioned in FL contexts, whereas several studies on bilingual acquisition show great benefits for children who learn two linguistic systems simultaneously (Cummins, 2001).

Another major concern among TEYL educators and stakeholders is the lack of globally (or widely) accepted guidelines to serve as a reference for standard setting. Although there is some consensus on who young learners are and how their proficiency develops at different cognitive stages, there seems to be a lack of consistency in practices around the world. According to Inbar-Lourie and Shohamy (2009, pp. 93–94, cited in Nikolov & Szabó, 2012, p. 348), early programmes range from awareness raising to language focus programmes and from content-based curricula to immersion. It appears to be particularly problematic to develop a global assessment which fits the richness of content aimed at young learners of different ages and with different learning needs worldwide. While the CEFR has become accepted as the reference for teaching and assessment of adults in Europe, different language institutions have produced different, and sometimes conflicting, interpretations of what the different levels mean. Moreover, there is no single document establishing a common standard for younger learners, but rather several stand-alone projects that try to align content to the CEFR or to national guidelines (e.g., Hasselgren, Kaledaité, Maldonado-Martin, & Pizorn, 2011). Pearson’s decision to develop a CEFR-based inventory of age-appropriate functional descriptors was motivated by the awareness of (1) the lack of consensus on standards for young learners and (2) the consequent need for a more transparent link between instructional and assessment materials, on the one hand, and teaching practices, on the other.

Although it is not the purpose of the present study to provide a detailed picture of all aspects of TEYL, we will briefly touch upon some of the main issues related to its implementation (see Nikolov & Curtain, 2000 for further details). In the first section of this chapter we present the heterogeneous and multifaceted reality of TEYL and discuss the need for standardisation. We outline the linguistic, affective and cognitive needs which characterize young learners. This brief overview is intended to provide the reader with some background on the current situation of TEYL and to support our arguments for the need of a set of descriptors for young learners. In the second section we discuss the limitations of the CEFR as a tool to assess young learners. We also describe the reporting scale used at Pearson, the Global Scale of English -henceforth GSE- (Pearson, 2015a), which is aligned to the CEFR. Then, we move to the main focus of our paper and explain how we developed the learning objectives by extending the CEFR functional descriptors and how we adapted them to the specific needs of a younger audience. Our descriptor set is intended to guide content development at primary and lower-secondary levels and to serve as a framework for assessment for EFL learners aged 6–14 and on the CEFR levels below A1 to high B1. The last section discusses the contribution of our paper to the research on young learners and briefly mentions some issues related to assessment.

2 The Heterogeneous Reality of TEYL and the Characteristics of Young Learners

2.1 The Need for Standardisation in TEYL

One of the major concerns related to TEYL is the absence of globally agreed and applied standards for measuring and comparing the quality of teaching and assessment programmes. Nikolov and Szabó (2012) mention a few initiatives aimed at adapting the CEFR to young learners' needs and examinations, along with their many challenges. According to Hasselgren (2005), the wide diffusion of the European Language Portfolio checklists developed by the Council of Europe (2014) for young learners has shown the impact of the CEFR on primary education. However, a glimpse into the different local realities around the world reveals a chaotic picture. Consider the obvious variety of foreign language programmes across Europe in terms of starting age, hours of instruction, teachers' proficiency in the foreign language, teachers' knowledge of TEYL, and support available to them (McKay, 2006; Nikolov & Curtain, 2000). Although there may be arguments for using different methods, approaches, and practices, a problem arises when no or little effort is made to work toward a common goal. Because of the absence of agreed standards, even within national education systems, existing learning, teaching and assessment resources are extremely diverse, leading to a lack of connectedness and resulting inefficacy. The implementation of a standard is therefore needed

to describe what learners are expected to know at different levels of schooling. At the national level, common learning goals should be clearly defined and students’ gains at each transition should be accounted for in order to guarantee continuity between different school grades. At the international level, standardisation should be promoted so as to increase the efficacy of teaching programmes in order to meet the requirements from increasing international mobility of learners and to allow for the comparison of educational systems.

2.2 Who Are Young Language Learners?

According to McKay (2006), young language learners are those who are learning a foreign or second language and who are doing so during the first 6 or 7 years of formal schooling. In our work we extend the definition to cover the age range from 6 to 14, the age at which learners are expected to have attained cognitive maturity. In our current definition, the pre-primary segment is excluded and age ranges are not differentiated. In the future, however, we may find it appropriate to split learners into three groups:

1. Entry years age, usually 5- or 6-year-olds: teaching often emphasizes oral skills and sometimes also focuses on literacy skills in the children’s first and foreign language
2. Lower primary age, 7–9: approach to teaching tends to be communicative with little focus on form
3. Upper primary/lower secondary age, 10–14: teaching becomes more formal and analytical.

In order to develop a set of learning objectives for young learners, a number of considerations have been taken into account.

- Young learners are expected to learn a new linguistic and conceptual system before they have a firm grasp of their own mother tongue. McKay (2006) points out that, in contrast to their native peers who learn literacy with well-developed oral skills, non-native speaker children may bring their L1 literacy background but with little or no oral knowledge of the foreign language. Knowledge of L1 literacy can facilitate or hinder learning the foreign language: whilst it helps learners handle writing and reading in the new language, a different script may indeed represent a disadvantage. In order to favour the activation of the same mechanisms that occur when learning one’s mother tongue, EFL programmes generally focus on the development of listening and speaking first and then on reading and writing. The initial focus is on helping children familiarize themselves with the L2’s alphabet and speech sounds, which will require more or less effort depending on the learners’ L1 skills and on the similarity between the target language and their mother tongue. The approach is communicative and tends to minimize attention to form. Children’s ability to use English will be

affected by factors such as the consistency and quality of the teaching approach, the number of hours of instruction, the amount of exposure to L2, and the opportunity to use the new language. EFL young learners mainly use the target language in the school context and have a minimal amount of exposure to the foreign language. Their linguistic needs are usually biased towards one specific experiential domain, i.e. interaction in the classroom. In contrast, adolescents and adult learners are likely to encounter language use in domains outside the classroom.

- The essentials for children’s daily communication are not the same as for adults. Young children often use the FL in a playful and exploratory way (Cazden, 1974 cited in Philp, Oliver & Mackey, 2008, p. 8). What constitutes general English for adults might be irrelevant for children (particularly the youngest learners) who talk more about topics related to the here and now, to games, to imagination (as in fairy tales) or to their particular daily activities. The CEFR (2001, p. 55) states that children use language not only to get things done but also to play and cites examples of playful language use in social games and word puzzles.
- The extent to which personal and extra-linguistic features influence the way children approach the new language and the impact of these factors are often underestimated (to this regard, see Mihaljević Djigunović, 2016 in this volume): learning and teaching materials rarely make an explicit link between linguistic and cognitive, emotional, social and physical skills.

Children experience continuous growth and have different skills and needs at different developmental stages. The affective and cognitive dimensions, in particular, play a more important role for young learners than for adults, implying a greater degree of responsibility on the part of parents, educators, schools, and ministries of education. One should keep in mind that because of their limited life experience each young learner is more unique in their interests and preferences than older learners are. Familiar and enjoyable contexts and topics associated with children’s daily experience foster confidence in the new language and help prevent them from feeling bored or tired; activities which are not contextualised and not motivating inhibit young learners’ attention and interest. From a cognitive point of view, teachers should not expect young learners to be able to do a task beyond their level. Tasks requiring metalanguage or manipulation of abstract ideas should not come until a child reaches a more mature cognitive stage. Young learners may easily understand words related to concrete objects but have difficulties when dealing with abstract ideas (Cameron, 2001, p. 81). Scaffolding can support children during their growth to improve their cognition-in-language and to function independently. In fact children are dependent upon the support of a teacher or other adult, not only to reformulate the language used, but also to guide them through a task in the most effective way. Vygotsky’s (1978) notion of the teacher or “more knowledgeable other” as a guide to help children go beyond their current understanding to a new level of understanding has become a foundational principle of child education: “what a child can do with some assistance today she will be able to do by herself tomorrow” (p. 87). The implication of this for assessing what young learners can do in a new language has been well expressed by Cameron (2001, p. 119):

Vygotsky turned ideas of assessment around by insisting that we do not get a true assessment of a child’s ability by measuring what he can do alone and without help; instead he suggested that what a child can do with helpful others both predicts the next stage in learning and gives a better assessment of learning.

3 Project Background: The CEFR and the Global Scale of English

The above brief overview of the main characteristics of young learners shows the need for learning objectives that are specifically appropriate for young learners. Following the principles laid out in the CEFR, we created such a new, age-appropriate set of functional descriptors. Although adult and young learners share a common learning core, only a few of the original CEFR descriptors are suitable for young learners.

Below we first discuss the limitations of the CEFR as a tool to describe young learners’ proficiency and present our arguments for the need to complement it with more descriptors across the different skills and levels. Then, we present the Global Scale of English, a scale of English proficiency developed at Pearson (Pearson, 2015a). This scale, which is linearly aligned to the CEFR scale, is the descriptive reporting scale for all Pearson English learning, teaching, and assessment products.

3.1 The CEFR: A Starting Point

The CEFR (Council of Europe, 2001) has acquired the status of the standard reference document for learning, teaching, and assessment practices in Europe (Little, 2006) and many other parts of the world. It is based on a model of communicative language use and offers reference levels of language proficiency on a six-level scale distinguishing two “Basic” levels (A1 and A2), two “Independent” levels (B1 and B2), and two “Proficient” levels (C1 and C2). The original Swiss project (North, 2000) produced a scale of nine levels, adding the “plus” levels: A2+, B1+ and B2+. The reference levels should be viewed as a non-prescriptive portrayal of a learner’s language proficiency development. A section of the original document published in 2001 explains how to implement the framework in different educational contexts and introduces the European Language Portfolio, the personal document of a learner, used as a self-assessment instrument, the content of which changes according to the target groups’ language and age (Council of Europe, 2001).

The CEFR has been widely adopted in language education (Little, 2007) acting as a driving force for rigorous validation of learning, teaching, and assessment practices in Europe and beyond (e.g., CEFR-J, Negishi, Takada & Tono, 2012). It has been successful in stimulating a fruitful debate about how to define what learners

can do. However, since the framework was developed to provide a common basis to describe language proficiency in general, it exhibits a number of limitations when implemented to develop syllabuses for learning in specific contexts. The CEFR provides guidelines only. We have used it as a starting point to create learning objectives for young learners, in line with the recommendations made in the original CEFR publication:

In accordance with the basic principles of pluralist democracy, the Framework aims to be not only comprehensive, transparent and coherent, but also open, dynamic and non-dogmatic. For that reason it cannot take up a position on one side or another of current theoretical disputes on the nature of language acquisition and its relation to language learning, nor should it embody any one particular approach to language teaching to the exclusion of all others. Its proper role is to encourage all those involved as partners to the language learning/teaching process to state as explicitly and transparently as possible their own theoretical basis and their practical procedures. (Council of Europe, 2001, p. 18)

The CEFR, however, has some limitations. Its levels are intended as a general, language-independent system to describe proficiency in terms of communicative language tasks. As such, the CEFR is not a prescriptive document but a framework for developing specifications, for example the *Profile Deutsch* (Glabionat, Müller, Rusch, Schmitz & Wertenschlag, 2005). The CEFR has received some criticism for its generic character (Fulcher, 2004) and some have warned that a non-unanimous interpretation has led to its misuse and to the proliferation of too many different practical applications of its intentions (De Jong, 2009). According to Weir (2005, p. 297), for example, “the CEFR is not sufficiently comprehensive, coherent or transparent for uncritical use in language testing”. In this respect, we acknowledge the invaluable contribution of the CEFR as a reference document to develop specific syllabuses and make use of the CEFR guidelines as the basis on which to develop a set of descriptors for young learners.

A second limitation in the context of YL is that the framework is adult-centric and does not really take into account learners in primary and lower-secondary education. For example, many of the communicative acts performed by children at the lower primary level lie at or below A1, but the CEFR contains no descriptors below A1 and only a few at A1. Whilst the CEFR is widely accepted as the standard for adults, its usefulness to teach and assess young learners is limited and presents more challenges. We therefore regard the framework as not entirely suitable for describing young learners’ skills and the aim of this project is to develop a set of age-appropriate descriptors.

Thirdly the CEFR levels provide the means to describe achievement in general terms, but are too wide to track progress over limited periods of time within any learning context. Furthermore, the number of descriptors in the original CEFR framework is rather limited in three of the four modes or language use (listening, reading, and writing), particularly outside of the range from A2 to B2. In order to describe proficiency at the level of precision required to observe progress realistically achievable within, for example, a semester, a larger set of descriptors, covering all language modes, is needed.

Finally, the CEFR describes language skills from a language-neutral perspective and therefore it does not provide information on the linguistic components (grammar and vocabulary) needed to carry out the communicative functions in a particular language. We are currently working on developing English grammar and vocabulary graded inventories for different learning contexts (General Adult, Professional, Academic, and Young Learner) in order to complement the functional guidance offered in the CEFR. The YL learning objectives will also have an additional section dedicated to enabling skills, including phonemic skills.

3.2 A Numerical Scale of English Proficiency

Pearson’s inventory of learning objectives differs from the CEFR in a number of aspects, most importantly, in the use of a granular scale of English proficiency, the GSE. This scale was first used as the reporting scale of Pearson Test of English Academic -PTE Academic- (Pearson, 2010) and will be applied progressively to all Pearson’s English products, regardless of whether they target young or adult learners. The GSE is a numerical scale ranging from 10 to 90 covering the CEFR levels from below A1 to the lower part of C2. The scale is a linear transformation of the logit scale underlying the descriptors on which the CEFR level definitions are based (North, 2000). It was validated by aligning it to the CEFR and by correlating it to a number of other international proficiency scales such as IELTS and TOEFL (De Jong & Zheng, *forthcoming*; Pearson, 2010; Zheng & De Jong, 2011).

The GSE is a continuous scale which allows us to describe progress as a series of small gains. The learning objectives for young learners do not go beyond the B1+ level because communicative skills required at B2 level and beyond are generally outside of the cognitive reach of learners under 15 (Hasselgren & Moe, 2006). Below 10 on the GSE any communicative ability is essentially non-linguistic. Learners may know a few isolated words, but are unable to use the language for communication. Above 90 proficiency is defined as being likely to be able to realize any communication about anything successfully and therefore irrelevant on a language measurement scale.

The GSE breaks the wide CEFR levels into smaller numeric values along its 10–90 scale; it specifies 81 points as opposed to the six levels of the CEFR (see Fig. 1). For young children especially, who progress at a slower pace than adults, this is particularly crucial. The scale offers a consistent, granular, and actionable measurement of English proficiency. It provides an instrument for a detailed account of

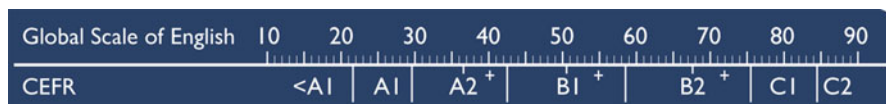


Fig. 1 The relation between the GSE and the CEFR

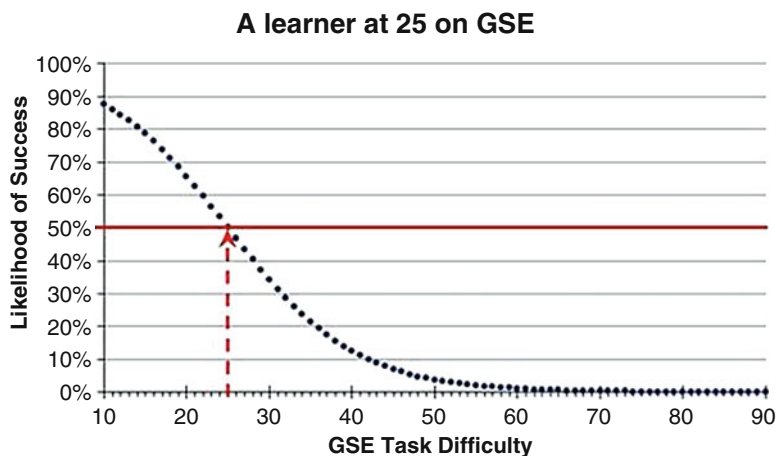


Fig. 2 Probability of success along the GSE of a learner with a score of 25

learners' levels and it offers the potential of more precise measurement of progress than is possible with the CEFR itself. The CEFR consists of six main levels to describe increasing proficiency and defines clear cut-offs between levels.

We should point out that learning a language is not a sequential process since learners might be strong in one area and weak in another. But what does it mean then to be, say, 25 on the GSE? It does not mean that learners have mastered every single learning objective for every skill up to that point. Neither does it mean that they have mastered no objectives at a higher GSE value. The definition of what it means to be at a given point of proficiency is based on probabilities. If learners are considered to be 25 on the GSE, they have a 50 % likelihood of being capable of performing all learning objectives of equal difficulty (25), a greater probability of being able to perform learning objectives at a lower GSE point, such as 10 or 15, and a lower probability of being able to cope with more complex learning objectives. The graphs below show the probability of success along the GSE of a learner at 25 and another learner at 61 (Figs. 2 and 3).

4 The Development of Learning Objectives

Pearson's learning objectives for young learners were created with the intention of describing what language tasks learners who are aged 6–14 can perform. Our inventory describes what learners can do at each level of proficiency in the same way as a framework, i.e. expressing communicative skills in terms of descriptors. In the next section we explain how we created YL descriptors sourcing them from different inputs. Then, we describe the rating exercise and the psychometric analysis carried out to validate and scale the descriptor set. Our work is overseen by a Technical Advisory Group (TAG) including academics, researchers, and practitioners

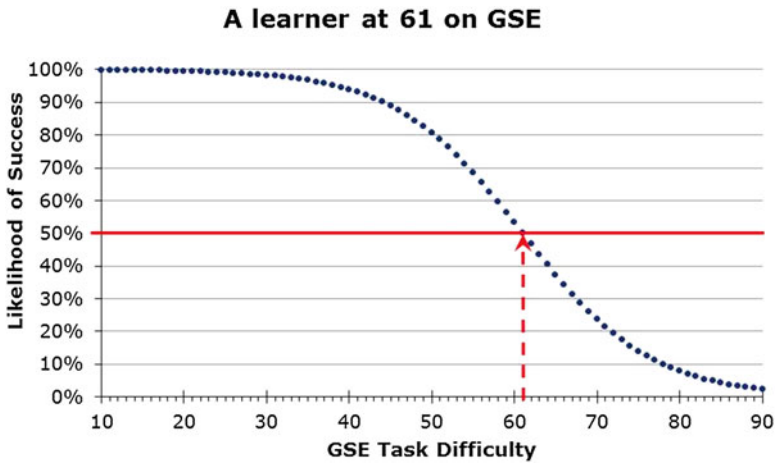


Fig. 3 Probability of success along the GSE of a learner with a score of 61

working with young learners who provide critical feedback on our methodology and evaluate the quality and appropriateness of our descriptor set and our rating and scaling exercises.

4.1 The Pool of Learning Objectives

The learning objectives were developed with the aim of describing early stages of developing ELT competencies. Accordingly, descriptors are intended to cover areas connected with personal identity such as the child’s family, home, animals, possessions, and free-time activities like computer games, sports and hobbies. Social interaction descriptors refer to the ‘here and now’ of interaction face to face with others. Descriptors also acknowledge that children are apprentice learners of social interaction; activities are in effect role-plays preparing for later real world interaction, such as ordering food from a menu at a restaurant. The present document is a report on the creation of the first batch: 120 learning objectives were created in two phases as described below: drawing learning objectives from various sources and editing them. In the next descriptor batches we are planning to refer to contexts of language use applicable particularly to the 6- to 9-year-old age range, including ludic language in songs, rhymes, fairy tales, and games.

Phase 1 started in September 2013 and lasted until February 2014. A number of materials were consulted to identify learning objectives for young learners: European Language Portfolio (ELP) documents, curriculum documents and exams (e.g., Pearson Test of English Young Learners, Cambridge Young Learners, Trinity exams, national exams), and Primary, Upper Primary and Lower Secondary course books. This database of learning objectives was our starting point to identify linguistic and communicative needs of young learners.

Phase 2 started in February 2014 and is still in progress: we are currently (summer 2014) working on two new batches of learning objectives (batch 2 and batch 3). With regard to batch 1, 120 new descriptors were created by qualified and experienced authors on the basis of the learning objectives previously identified. Authors followed specific guidelines and worked independently on developing their own learning objectives. Once a pool of learning objectives was finalised, they were validated for conformity to the guidelines and for how easy it was to evaluate their difficulty and to assign a proficiency level to them. We held in-house workshops to validate descriptors with editorial teams. Authors assessed one another's work. If learning objectives appeared to be unfit for purpose or no consensus was reached among the authors, they were amended or eliminated.

The set of 120 learning objectives included 30 for each of the four skills. Additionally, twelve learning objectives were used as anchor items with known values on the GSE, bringing the total number of learning objectives to 132. Among the anchors, eight learning objectives were descriptors taken verbatim from the CEFR (North, 2000) and four were adapted from the CEFR: they had been rewritten, rated and calibrated in a previous rating exercise for general English learning objectives. In our rating exercises for the GSE, the same anchors are used in different sets of learning objectives in order to link the data. The level of the anchors brackets the target CEFR level of the set of learning objectives to be rated: for example, if a set of learning objectives contains learning objectives targeted at the A1 to B2 levels, anchors are required from below A1 up to C1. A selection of the most YL-appropriate learning objectives from the CEFR was used as anchors.

A number of basic principles are applied in editing learning objectives. Learning objectives need to be relatively generic, describing performance in general, yet referring to a specific skill. In order to reflect the CEFR model, all learning objectives need to refer to the quantity dimension, i.e., what are the language actions a learner can perform, and to the quality dimension, i.e., how well (in terms of efficacy and efficiency) a learner is expected to perform these at a particular level. Each descriptor refers to one language action. The quantity dimension refers to the type and context of communicative activity (e.g., listening as a member of an audience), while the quality dimension typically refers to the linguistic competences determining efficiency and effectiveness in language use, and is frequently expressed as a condition or constraint (e.g., if the speech is slow and clear). Take, for example, one of our learning objectives for writing below:

- Can copy short familiar words presented in standard printed form (below A1 – GSE value 11).

The language operation itself is copying, the intrinsic quality of the performance is that words are short and familiar, and the extrinsic condition is that they are presented in standard printed form.

The same communicative act often occurs at different proficiency levels with a different level of quality.

See, for example, the progression in these two **listening** learning objectives developed by Pearson:

- Can recognise familiar words in short, clearly articulated utterances, with visual support. (below A1; GSE value 19)
- Can recognise familiar key words and phrases in short, basic descriptions (e.g., of objects, animals or people), when spoken slowly and clearly. (A1; GSE value 24)

The first descriptor outlines short inputs embedded in a visual context, provided that words are familiar to the listener and clearly articulated by the speaker. The listener needs to recognize only specific vocabulary items to get the meaning. The second descriptor shows that as children progress in their proficiency, they are gradually able to cope with descriptions that require more linguistic resources than isolated word recognition and the ability to hold a sequence in memory.

Similarly, for **speaking**, the earliest level of development is mastery of some vocabulary items and fixed expressions such as greetings. Social exchanges develop in predictable situations until the point where children can produce unscripted utterances. See, for example, the difference between a learner at below A1 and another learner at A1:

- Can use basic informal expressions for greeting and leave-taking, e.g., Hello, Hi, Bye. (below A1; GSE value 11).
- Can say how they feel at the moment, using a limited range of common adjectives, e.g., happy, cold. (A1; GSE value 22).

For **writing**, the following learning objectives show a progression from very simple (below A1) to elaborate writing involving personal opinions (B1):

- Can copy the letters of the alphabet in lower case (below A1; GSE value 10).
- Can write a few basic sentences introducing themselves and giving basic personal information, with support (A1; GSE value 26).
- Can link two simple sentences using “but” to express basic contrast, with support. (A2; GSE value 33).
- Can write short, simple personal emails describing future plans, with support. (B1; GSE value 43).

The third example above shows that ‘support’ (from interlocutor, e.g., the teacher) is recognized in the learning objectives as a facilitating condition. Support can be realized in the form of a speaker’s gestures or facial expressions or from pictures, as well as through the use of adapted language (by the teacher or an adult interlocutor).

Similarly, the following **reading** descriptors show the progression from basic written receptive skills to the ability to read simple texts with support:

- Can recognise the letters of the Latin alphabet in upper and lower case. (below A1; GSE value 10).
- Can recognise some very familiar words by sight-reading. (A1; GSE value 21)
- Can understand some details in short, simple formulaic dialogues on familiar everyday topics, with visual support. (A2; GSE value 29)

A number of other secondary criteria were applied. North (2000, pp. 386–389) lists five criteria learning objectives should meet in order to be scalable.

- Positiveness: Learning objectives should be positive, referring to abilities rather than inabilities.
- Definiteness: Learning objectives should describe concrete features of performance, concrete tasks and/or concrete degrees of skill in performing tasks. North (2000, p. 387) points out that this means that learning objectives should avoid vagueness (“a range of”, “some degree of”) and in addition should not be dependent for their scaling on replacement of words (“a few” by “many”; “moderate” by “good”).
- Clarity: Learning objectives should be transparent, not dense, verbose or jargon-ridden.
- Brevity: North (2000, p. 389) reports that teachers in his rating workshops tended to reject or split learning objectives longer than about 20 words and refers to Oppenheim (1966/1992, p. 128) who recommended up to approximately 20 words for opinion polling and market research. We have used the criterion of approximately 10–20 words.
- Independence: Learning objectives should be interpretable without reference to other learning objectives on the scale.

Based on our experience in previous rating projects (Pearson, 2015b), we added the following requirements.

- Each descriptor needs to have a functional focus, i.e., be action-oriented, refer to the real-world language skills (not to grammar or vocabulary), refer to classes of real life tasks (not to discrete assessment tasks), and be applicable to a variety of everyday situations. E.g. “Can describe their daily routines in a basic way” (A1, GSE 29).
- Learning objectives need to refer to gradable “families” of tasks, i.e., allow for qualitative or level differentiations of similar tasks (basic/simple, adequate/standard, etc.), e.g., “Can follow short, basic classroom instructions, if supported by gestures” (Listening, below A1, GSE 14).

To ensure that this does not conflict with North’s (2000) ‘Definiteness’ requirement, we have added two further stipulations:

- Learning objectives should use qualifiers such as “short”, “simple”, etc. in a sparing and consistent way as defined in an accompanying glossary.
- Learning objectives must have a single focus so as to avoid multiple tasks which might each require different performance levels.

In order to reduce interdependency between learning objectives we produced a glossary defining commonly used terms such as “identify” (i.e., pick out specific information or relevant details even when never seen or heard before), “recognize” (i.e., pick out specific information or relevant details when previously seen or heard), “follow” (i.e., understand sufficiently to carry out instructions or directions, or to keep up with a conversation, etc. without getting lost). The glossary also provides definitions of qualifiers such as “short”, “basic”, and “simple”.

4.2 *The Rating of Learning Objectives*

Once the pool of new learning objectives was signed off internally, they were validated and scaled through rating exercises similar to the methodology used in the original CEFR work by North (2000). The ratings had three goals: (1) to establish whether the learning objectives were sufficiently clear and unambiguous to be interpretable by teachers and language experts worldwide; (2) to determine their position on the CEFR and the GSE scales; and (3) to determine the degree of agreement reached by teachers and experts in assigning a position on the GSE to learning objectives.

The Council of Europe (2009) states that to align materials (tests, items, and learning objectives) to the CEFR, people are required to have knowledge of (be familiar with) policy definitions, learning objectives, and test scores. As it is difficult to find people with knowledge of all three, multiple sources are required (Figueras & Noijons, 2009, p. 14). The setting of the rating exercise for each group was a workshop, a survey or a combination of both workshop and online survey for teachers. Training sessions for Batch 1 were held between March and April 2014 for two groups accounting for a total of 1,460 raters: (1) A group of 58 expert raters who were knowledgeable about the CEFR, curricula, writing materials, etc. This group included Pearson English editorial staff and ELT teachers. (2) A group of 1,402 YL teachers worldwide who claimed to have some familiarity with the CEFR. The first group took part in a face-to-face webinar where they were given information about the CEFR, the GSE, and the YL project and then trained to rate individual learning objectives. They were asked to rate the learning objectives, first according to CEFR levels, and then, to decide if they thought the descriptor would be taught at the top, middle or bottom of the level. Based on this estimate, they were asked to select a GSE value corresponding to a sub-section of the CEFR level. The second group participated in online surveys, in which teachers were asked to rate the learning objectives according to CEFR levels only (without being trained on the GSE).

All raters were asked to provide information about their knowledge of the CEFR, the number of years of teaching experience and the age groups of learners they taught (choosing from a range of options between lower primary and young adult/adult – academic English). We did not ask the teachers to provide information on their own level of English, as the survey was self-selecting; if they were familiar with the CEFR and able to complete the familiarisation training, we assumed their level of English was high enough to be able to perform the rating task. They answered the following questions:

- How familiar are you with the CEFR levels and descriptors?
- How long have you been teaching?
- Which of the following students do you mostly teach? If you teach more than one group, please select the one you have most experience with – and think about this group of students when completing the ratings exercise.
- What is your first language?
- In which country do you currently teach?

Appendixes A and B comprise the summary statistics of survey answers by selected teachers and by selected expert raters respectively. They report data of only 274 raters ($n=37$ expert raters and $n=237$ teachers out of the total of 1,460 raters) who passed the filtering criteria after psychometric analysis.

The total of 120 new learning objectives was then subjected to rating together with twelve anchors (a total of 132 learning objectives) by the two groups. For the online ratings by the 1,402 teachers, the learning objectives were divided into six batches each containing 20 new learning objectives and four anchors. Each new descriptor occurred in two batches and each anchor occurred in four batches. The teachers were divided into six groups of about 230 teachers. Each group of teachers were given two batches to rate in an overlapping design: Group 1 rated Batches 1 and 2, Group 2 Batches 2 and 3, etc., so each new descriptor was presented to a total of about 460 teachers, whereas the anchors occurred in twice as many batches and were rated by close to a thousand teachers, producing a total of over 61,000 data points. The descriptor set was structured according to specific criteria. Similar learning objectives were kept in separate batches to make sure each descriptor was seen as completely independent in meaning. Moreover, each batch was balanced proportionally, so that each contained approximately the same proportion of learning objectives across the skills and levels in relation to the overall set. The 58 experts each were given all 120 learning objectives and the twelve anchors to rate, resulting in a total data set of more than 6,500 data points.

4.3 *The Psychometric Analysis*

After all ratings were gathered, they were analysed and were assigned a CEFR/GSE value. The data consisted of ratings assigned to a total of 132 learning objectives by 58 language experts and 1,402 teachers. Below we describe the steps we followed to assign a GSE value to each descriptor and to measure certainty values of the individuals' ratings.

As the GSE is a linear transformation of North's (2000) original logit-based reporting scale, the GSE values obtained for the anchor learning objectives can be used as evidence for the alignment of the new set of learning objectives with the original CEFR scale. Three anchor learning objectives were removed from the data set. One anchor descriptor had accidentally been used as an example (with a GSE value assigned to it) in the expert training. Independence of the expert ratings could therefore not be ascertained. Another anchor did not obtain a GSE value in alignment with the North (2000) reported logit value. For the third descriptor no original logit value was available in North (2000), although it was used as an illustrative descriptor in the CEFR (Council of Europe, 2001). Therefore, the number of valid anchors was nine and the total number of rated learning objectives was 129.

The values of the anchors found in the current project were compared to those obtained for the same anchors used in preceding research rating adult oriented learning objectives (Pearson, 2015b). The correlation (Pearson's r) between ratings

assigned to anchors in the two research projects was 0.95. The anchors had a correlation of 0.94 (Pearson’s *r*) with the logit values reported by North (2000), indicating a remarkable stability of these original estimates, especially when taking into account that the North data were gathered from teachers in Switzerland more than 15 years ago.

The rating quality of each rater was evaluated according to a number of criteria. As previously explained, the original number of 1,460 raters (recruited at the start of the project) reduced to only 274 raters after running psychometric analysis of all data. Raters were removed if (1) the standard deviation of their ratings was close to zero as this was an index of lack of variety in their ratings; (2) they rated less than 50 % of the learning objectives; (3) the correlation between their ratings on the set of learning objectives and the average rating from all raters was lower than 0.7; and (4) if they showed a deviant mean rating (*z* mean beyond $p < 0.05$). As a result, from the total of 1,460 (37 of 58 expert raters and 237 of 1,402 teachers) only 274 raters passed these filtering criteria. The selected teachers came from 42 different countries worldwide.

Table 1 shows the distribution of the learning objectives along CEFR levels according to the combined ratings of the two groups. It was found to peak at the A2 and B1 levels, indicating the need to focus more on low level learning objectives in the following batches.

Table 2 shows the certainty index distribution based on the two groups’ ratings. Certainty is computed as the proportion of ratings within two adjacent most often

Table 1 Learning objectives’ distribution across CEFR levels

GSE	CEFR	n	%
<22	<A1	4	3
22–29	A1	20	16
30–42	A2	66	51
43–58	B1	37	29
59–75	B2	2	2
76–84	C1	0	0
≥85	C2	0	0
Total		129	100

Table 2 Certainty index distribution of ratings

Certainty	n	%
>.90	29	22
.80–.90	66	51
.75–.79	25	19
.70–.74	4	3
.60–.69	5	4
<.60	0	0
Total	129	100

selected levels of the CEFR. Let us take, for example, a descriptor which is rated as A1 by a proportion of .26 of the raters, as A2, by .65 of the raters, and by .09 as B1. In this case the degree of certainty in rating this descriptor is the sum of the proportions observed with the two largest adjacent categories, i.e., A1 and A2 with .26 and .65 respectively. The sum of these yields a value of .91. This is taken as the degree of certainty in rating this descriptor. Only 4 % of the data set showed certainty values below .70, while only 7 % of the learning objectives showed certainty below .75. At this stage we take the low certainty as an indication of possible issues with the descriptor, but will not reject any descriptor. At a later stage, we will combine the set reported on here with all other available descriptor sets and evaluate the resulting total set using the one-parameter Rasch model (Rasch, 1960/1980) to estimate final GSE values. This will increase the precision of the GSE estimates and reduce the dependency on the raters involved in the individual projects. At that time the certainty of ratings will be re-evaluated and learning objectives with certainty below a certain threshold will be removed.

5 Final Considerations

In this paper we described work in progress to develop a CEFR-based descriptor set targeting young learners. In Sect. 3 we discussed limitations of the CEFR, with a special focus on its restricted suitability to describe what young learners can do in their new language. The system of levels provided by the CEFR has widely spread among practitioners and the framework has been the theme of international conferences such as EALTA and LTRC 2014. The CEFR has been validated by numerous follow-up initiatives since its publication in 2001. Since the principle of a qualitative and a quantitative dimension of language development of the CEFR is applicable to learners of all age groups, we believe the framework provides firm guidance and is suitable to be adapted to young learners. Although the present paper reports on the initial stage of the project, the analysis of the first batch of 120 learning objectives has allowed us to review our methodology to inform the next phases of the project. The current batch has shown high reliability and methodological rigour.

Next steps will include the calibration of more sets of learning objectives and the inclusion of these sets in a larger set including data on general academic and professional English learning objectives to be analysed using the Rasch (1960/1980) model for final scaling. In the near future, we hope to be able to report on the development of these additional batches of learning objectives as well as the standardisation of Pearson teaching and testing materials based on the same learning objectives and the same proficiency scale.

Appendices

Appendix A: Summary Statistics of Survey Answers by Selected Teachers (Tables 3, 4, and 5)

Table 3 Familiarity with CEFR descriptors

How familiar are you with the CEFR levels and descriptors?	n	%
I have a detailed knowledge of them	37	16
I have a general understanding of them	200	84
Total	237	100

Table 4 Age groups taught

Which of the following students do you mostly teach? ^a		
Age group	n	Percentage
6–9	55	23
9–12	71	30
10–14	90	38
12–15	92	39
15–18	82	35

^aMost teachers responded they were teaching in more than one age group

Table 5 First language

What is your first language?	n	%
Spanish	51	22
Russian	38	16
English	28	12
Italian	15	6
Greek	12	5
Polish	13	5
Romanian	11	5
Serbian	10	4
Portuguese	10	4
21 Other languages	39	16
No answer	10	4
Total	237	100

Appendix B: Summary Statistics of Survey Answers by Selected Expert Raters (Tables 6, 7, 8, 9, and 10)

Table 6 Teaching experience

How long have you been teaching?	n	%
Less than 2 years	5	2
2–5 years	5	12
More than 5 years	31	86
Total	37	100

Table 7 Country of teaching

In which country do you currently teach?	n	%
Russia	28	12
Argentina	22	9
Greece	19	8
Italy	20	8
Poland	13	5
Spain	12	5
42 other countries	109	46
No answer	14	6
Total	237	100

Table 8 Age groups taught

Which of the following students do you <u>mostly</u> teach?	n	%
<i>Lower Primary (age 6–9)</i>	10	27
Upper Primary (age 9–12)	12	32
Upper Primary/Lower Secondary (age 10–14)	3	8
Lower Secondary (age 12–15)	8	22
Upper Secondary (age 15/16–18)	2	5
Young adult/adult students – general English	2	5
Total	37	100

Table 9 First language

What is your first language?	n	%
English	10	27
Cantonese	9	24
Polish	7	19
Portuguese	3	8
Catalan	2	5
Hungarian	2	5
Spanish	3	8
Italian	1	3
Total	37	100

Table 10 Teaching experience

How long have you been teaching?	n	%
2–5 years	5	14
More than 5 years	32	86
Total	37	100

References

- Cameron, L. (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.
- Council of Europe. (2001). *The common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2009). *Relating examinations to the Common European framework of reference for languages: Learning, teaching, assessment (CEFR): A Manual*. Strasbourg: Council of Europe.
- Council of Europe. (2014). *ELP checklists for young learners: Some principles and proposals. European Language Portfolio templates and resources language biography*. Retrieved June 20, 2014, from http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Templates/ELP_Language_Biography_Checklists_for_young_learners_EN.pdf.
- Cummins, J. (2001). *Negotiating identities: Education for empowerment in a diverse society* (2nd ed.). Los Angeles: California Association for Bilingual Education.
- De Jong, J. H. A. L. (2009). *Unwarranted claims about CEF alignment of some international English language tests*. Paper presented at EALTA Conference, June 2009. Retrieved May 25, 2015, from http://www.ealta.eu.org/conference/2009/docs/friday/John_deJong.pdf.
- De Jong, J. H. A. L., & Zheng, Y. (forthcoming). Linking to the CEFR: Validation using a priori and a posteriori evidence. In J. Banerjee & D. Tsagari (Eds.), *Contemporary second language assessment*. London/New York: Continuum.
- European Commission. (2011). *Commission staff working paper. European strategic framework for education and training (ET 2020). Language learning at pre-primary school level: making it efficient and sustainable a policy handbook*. Retrieved July 10, 2014, from http://ec.europa.eu/languages/policy/language-policy/documents/early-language-learning-handbook_en.pdf.
- Figueras, N., & Noijons, J. (Eds.). (2009). *Linking to the CEFR levels: Research perspectives*. Arnhem: Cito/EALTA.
- Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly*, 1(4), 253–266.
- Glabionat, M., Müller, M., Rusch, P., Schmitz, H., & Wertenschlag, L. (2005). *Profile deutsch A1-C2 (Lernzielbestimmungen, Kannbeschreibungen, Kommunikative Mittel)*. München: Langenscheidt.
- Hasselgren, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337–354.
- Hasselgren, A., Kaledaitė, V., Maldonado-Martin, N., & Pizorn, K. (2011). *Assessment of young learner literacy linked to the common European framework for languages*. European Centre of Modern languages/Council of Europe publishing. Retrieved July 15, 2014, from http://srvcs-npbs.xtec.cat/cirel/cirel/docs/pdf/2011_08_09_Ayllit_web.pdf.
- Hasselgren, A., & Moe, E. (2006). *Young learners' writing and the CEFR: Where practice tests theory*. Paper presented at the Third Annual Conference of EALTA, Kraków. Retrieved July 15, 2014, from http://www.ealta.eu.org/conference/2006/docs/Hasselgren&Moe_ealta2006.pdf.
- Inbar-Lourie, O., & Shohamy, E. (2009). Assessing young language learners: What is the construct? In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 83–96). Berlin, Germany: Mouton de Gruyter.

- Lightbown, P. M., & Spada, N. (2008). *How languages are learned*. New York: Oxford University Press.
- Little, D. (2006). The common European framework of reference for languages: Content, purpose, origin, reception and impact. *Language Teaching*, 39(3), 167–190.
- Little, D. (2007). The common European framework of references for languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4), 645–655.
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- Mihaljević Djigunović, J. M. (2016). Individual differences and young learners' performance on L2 speaking tests. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives*. New York: Springer.
- Negishi, M., Takada, T., & Tono, Y. (2012). *A progress report on the development of the CEFR-J* (Studies of Language Testing, 36, pp. 137–165). Cambridge: Cambridge University Press.
- Nikolov, M. (2016). A framework for young EFL learners' diagnostic assessment: Can do statements and task types. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives*. New York: Springer.
- Nikolov, M., & Curtain, H. (Eds.). (2000). *An early start: Young learners and modern languages in Europe and beyond*. Strasbourg: Council of Europe.
- Nikolov, M., & Mihaljević Djigunović, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review Applied Linguistics*, 26, 234–260.
- Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics*, 31, 95–119.
- Nikolov, M., & Szabó, G. (2012). Developing diagnostic tests for young learners of EFL in grades 1 to 6. In Galaczi E. D. & Weir C. J. (Eds.), *Voices in language assessment: Exploring the impact of language frameworks on learning, teaching and assessment: Policies, procedures and challenges* (pp. 347–363). Proceedings of the ALTE Krakow Conference, July 2011. Cambridge: UCLES/Cambridge University Press.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- Oppenheim, A., N. (1966/1992) (2nd ed.) *Questionnaire design, interviewing and attitude measurement*. London: Pinter Publishers.
- Pearson. (2010). *Aligning PTE Academic Test Scores to the common European framework of reference for languages*. Retrieved June 2, 2014, from http://pearsonpte.com/research/Documents/Aligning_PTEA_Scores_CEF.pdf.
- Pearson. (2015a). *The Global Scale of English*. Retrieved May 25, 2015, from <http://www.english.com/gse>.
- Pearson. (2015b). *The Global Scale of English Learning Objectives for Adults*. Retrieved May 25, 2015, from <http://www.english.com/blog/gse-learning-objectives-for-adults>.
- Philp, J., Oliver, R., & Mackey, A. (Eds.). (2008). *Second language acquisition and the young learner: Child's play?* Amsterdam: John Benjamins.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. Expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
- Singleton, D. (1989). *Language acquisition. The age factor*. Clevedon: Multilingual Matters.
- Speitz, H. (2012). Experiences with an earlier start to modern foreign languages other than English in Norway. In A. Hasselgren, I. Drew, & S. Bjørn (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 11–22). Bergen: Fagbokforlaget.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Oxford, UK: Palgrave.
- Zheng, Y., & De Jong, J. (2011). *Establishing construct and concurrent validity of Pearson Test of English Academic* (1–47). Retrieved May 20, 2014, from <http://pearsonpte.com/research/Pages/ResearchSummaries.aspx>.