

1 Introduction

Szilvia Papp

Consultant, Cambridge Assessment English

The context for the volume

Examining Young Learners: Research and Practice in Assessing the English of School-age Learners stands as the latest addition in a series of Studies in Language Testing (SiLT) volumes designed to explore the ability constructs underpinning the testing of English language skills. The previous four ‘construct volumes’ in the series have explored the testing of the skills of listening (Geranpayeh and Taylor (Eds) 2013), speaking (Taylor (Ed) 2011), reading (Khalifa and Weir 2009) and writing (Shaw and Weir 2007) in adult learners of English. The focus of the present volume shifts to an investigation of the ability constructs underlying tests for young language learners (YLLs) whose age range is 6 to 16.

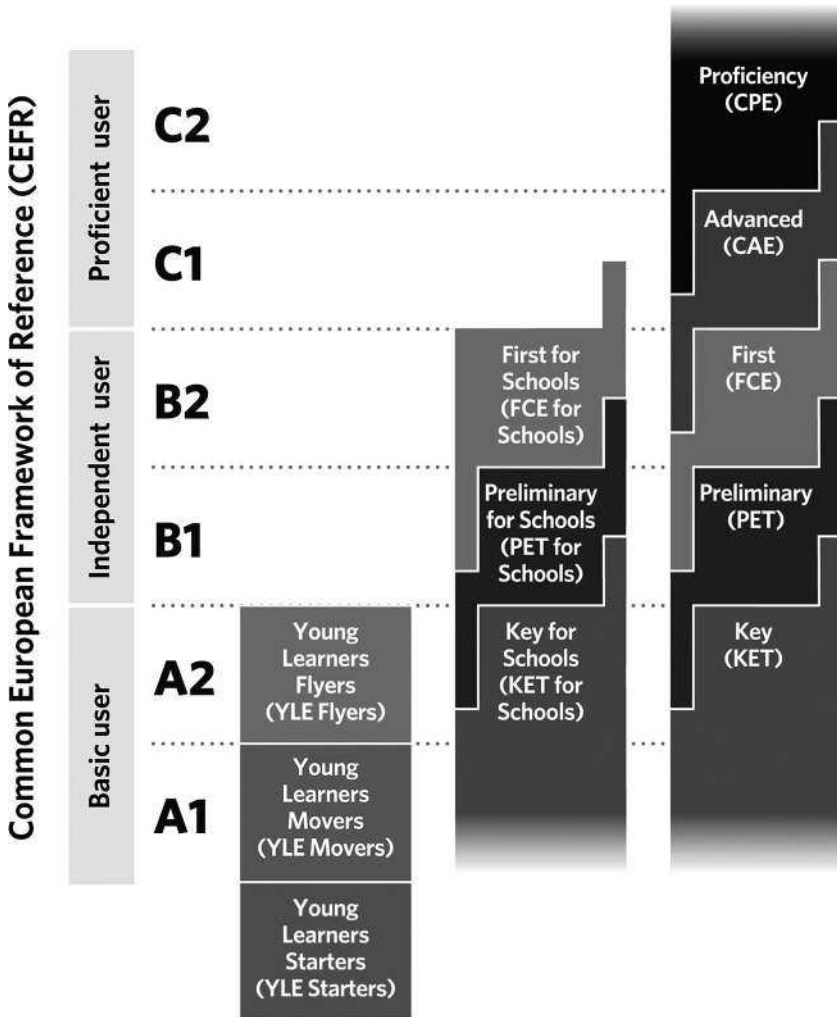
The focus of the volume

Young learners, defined as children and young people between the ages of 6 and 16, are the focus of this volume. This age range was chosen as Cambridge Assessment English offers a suite of examinations to suit school learners with various Common European Framework of Reference for languages (CEFR, Council of Europe 2001) proficiency levels (pre-A1 to B2), as seen in Figure 1.1. The age range is to cover primary and lower secondary school learners. This range covers compulsory schooling stages in most educational systems. It can be divided into subgroups in different ways. For instance, target age ranges of European Language Portfolios tend to fall into two larger categories: ages 7 to 10 and 11 to 15 (Goodier and Szabo 2017). Hasselgreen and Caudwell (2016) have categories for younger children (5/6 to 8/9 years), older children (8/9 to 12/13 years) and teenagers (12/13 to 17 years). Morgan (2013) subdivides adolescents into early teens (roughly 13/14), mid-teens (roughly 14/15) and late teens (15 to 17 years).

The *Cambridge English: Young Learners* tests (now known as ‘Young Learners exams’) were developed for candidates between ages 7 and 12 to build a bridge to take young learners of English as a foreign language from beginner to A2/*Waystage* level (van Ek and Trim 1998a). They were designed, during the test development phase between 1992 and 1997, to be linked to the CEFR via the Council of Europe’s T-series, to cover three key levels of

Examining Young Learners

Figure 1.1 Young Learners and for Schools exams from Cambridge English



assessment: pre-A1 (*Cambridge English: Starters*), A1 (*Cambridge English: Movers*) and A2 (*Cambridge English: Flyers*) on the CEFR. (These are now known as Pre A1 Starters, A1 Movers, and A2 Flyers, respectively.)

The designers of the *Young Learners* tests were aware of the need for the test to be motivational while not sacrificing the need to measure accurately and fairly (Saville 1996). The main aim of these tests is not to offer certification, in sharp contrast with the purpose of test use in the Cambridge English General English exams, including the *for Schools* exams. Test specifications

Introduction

for *Young Learners (Starters, Movers, Flyers)* were carefully developed to offer candidates the first steps on a learning ladder, presenting a positive first experience of international tests and leading to the following exams: *Cambridge English: Key (KET)*, *Cambridge English: Preliminary (PET)*, and *Cambridge English: First (FCE)*. The *for Schools* versions of *Cambridge English: Key* and *Cambridge English: Preliminary* were designed for 11- to 14-year-olds and launched in 2009, and *Cambridge English: First for Schools* for 12- to 16-year-olds was first administered in 2011.

The purpose of the volume

The principle aim of the volume is firstly to explore how language constructs in the context of young learners have been conceptualised and operationalised by the language testing and assessment community, including Cambridge English, over the last 35 years in language tests designed for children and teenagers. The second aim is to offer a critical reflection on practices adopted by Cambridge English.

The field of assessment of YLLs is rapidly changing, as testing agencies and examination boards have engaged in either developing or revising their exams for YLLs in the face of the growing status of English worldwide and increasing demand in the school sector. There are several large-scale international YLL examinations with varying underlying constructs. While national tests of general English language proficiency have existed for a long time as secondary school leaving exams, large-scale international assessments for younger English language learners have appeared only in the last 35 years. The Pearson Test of English (PTE) Young Learners (for 6- to 13-year-olds, pre-A1 to A2) formerly known as London Test of English for Children has been in existence since 1982. The PTE Young Learners and Cambridge English tests measure general English language achievement with explicit exam syllabuses specifying what vocabulary and grammatical structures learners need to have mastered for successful achievement. Pearson has more recently developed a set of Young Learner Learning Objectives aligned to the Global Scale of English, the reporting scale that is used across a number of Pearson products, and the CEFR, intended to support both the teaching and assessment of YLLs.

To extend their family of assessments, Educational Testing Service (ETS) developed *TOEFL Junior*® in 2011 and *TOEFL Primary*® in 2012 and has implemented a comprehensive research agenda for their validation. With no clearly specified exam syllabus, *TOEFL Primary*® provides ‘information about the English proficiency of young English learners in countries where English is not typically used in daily life’ (Educational Testing Service 2015a, 2015b), whereas *TOEFL Junior*® measures English language proficiency needed in English language instructional contexts (So, Wolf, Hauck,

Examining Young Learners

Mollaun, Tumposky and Wang 2015) which is reflected in the construct and content of the tests. The computer-based British Council's Aptis for Teens (13- to 17-year-olds, A1 to C levels) was launched in 2014, measuring the general proficiency of English as a Foreign Language/English as a Second Language (EFL/ESL) learners who learn English 'as a subject of study or as a medium of instruction to study other subjects' (Zheng and Berry 2015).

There are other assessments for young learners for which there is less information provided in the public domain: Trinity's GESE grades suitable for children, i.e. Grades 1–4; Anglia's General English test levels suitable for children and teenagers, i.e. First Step (suitable for pre-schoolers), Junior, Primary, Preliminary, Elementary and Pre Intermediate; telc's Certificate in English A1 level exam and telc for Schools exams offered at Levels A2 and B1 of the CEFR, to name but a few. For very young candidates, there is Trinity Stars (a group performance award for children between the ages of 3 and 12). A comprehensive in-depth exploration of key issues is, therefore, timely and necessary. The present volume presents a critical reflection on theories and practices of testing younger learners to underpin and enhance the increased activities surrounding YLL assessment.

The contribution of this volume

YLL examining needs to be set into a broader context. Therefore, this volume goes beyond investigating internationally available exams and looks at educational contexts in which assessment (as a broader concept) of young learners takes place (see Chapter 2 and Chapters 13 and 14). This is firstly because large-scale international examinations (including those offered by Cambridge English) need to fit into existing assessment cultures as well as being sources of influence on existing assessment cultures.

On the other hand, in line with the publication of *The Cambridge Approach* (Cambridge Assessment 2009), *Cambridge English Principles of Good Practice* (Cambridge English 2016) and the four 'construct volumes', this volume is another instance of the professional responsibility of test developers and examinations providers to be transparent and accountable in what they do and to provide explicit evidence to support validity claims made about their exams. The present volume places in the public domain the rationale and evidence in support of Cambridge English testing policy and practice for YLL examinations. This volume, following the ethos of the other construct volumes, represents an exercise in public accountability and scrutiny and hopes to offer a critical evaluation of and reflection on the Cambridge English approach to testing children's and teenagers' English language abilities. It is hoped that other institutions and testing bodies will consider doing the same public exercise in reflecting on their current practice in testing children and teenagers, so that stakeholders can thus be enabled to make informed choices.

Introduction

Such a self-reflective exercise about theory and practice would contribute to the growing professionalisation of the field and with it meet an increasing need for the development of assessment literacy in different constituencies involved directly or indirectly with language assessment of young learners.

It is worth noting that the process of professionalisation of language testers was much invigorated after Messick's (1989a, 1996) re-conceptualisation of construct validity to include the notion of social consequence as an essential facet of validity. The idea of test consequence is now widely seen as essential to a validity argument. As Saville (2009) argues, an awareness of test consequence as impacting on all elements of validity dictates an 'impact by design' approach, where a consideration of consequence forms a guiding principle for the test development process from the very beginning of the cycle.

The development of standards of good practice, codes of conduct and ethics has also been a strong indication of the growing professionalisation of the field. This professional ethic continues to evolve, as noted by Kunnan (2008, 2014), as there is increased demand for accountability and responsiveness to stakeholders and a wider use of language tests for decision-making purposes. Professional bodies and associations (e.g. ALTE, EALTA, ILTA) advocate an evidence-driven approach to establishing the quality of language tests.

It is interesting to reflect that the move towards a broader and more socially responsible conceptualisation of validity started to evolve at around the same time – in the early 1990s – as the conception of the need for English language tests for children emerged at Cambridge English (then known as Cambridge ESOL). Not surprisingly, ethical and fairness considerations have always been at the heart of the development of language tests for young learners at Cambridge English. Quality, transparency and accountability are just as important in the case of tests for younger learners as they are in the case of much higher-stakes tests. The Council of Europe, which produced the most influential interpretative language educational framework in Europe, the CEFR (Council of Europe 2001), has recently started a project to incorporate children's and young people's needs in learning foreign/second languages (L2) in the formal school contexts in which most young learners learn their L2, as advocated by Papp and Salamoura (2009); see Goodier and Szabo (2017).

Surprisingly, there are still unresolved issues, uncertainties and misguided myths around children's L2 learning and assessment. Examples are the debates surrounding the question of the optimal age to start learning an L2, the issue of what language should be the medium of instruction in schools in multilingual contexts, what proportion of target language or children's first language (L1) should be used to conduct L2 lessons in schools, even discussions on the benefits of bilingualism. Some authorities have expressed

Examining Young Learners

caution, even misgivings, over the formal testing and grading of young children. Smith (1995), in an influential article, expressed the following views:

[...] the assessment and testing of young learners have to reflect the approaches used in teaching this age group. Traditional paper and pencil tests cannot incorporate the wide repertoire of activities found in the primary EFL classroom (Smith 1995:1).

She adds later:

Several of the activities the young language learner is engaged in cannot be simplified into test situations; attitude goals cannot be tested, neither can most learning-to-learn and educational goals (Smith 1995:3).

However, the evidence for common classroom practice provided by research such as that by Rea-Dickins and Rixon (1999) and Brumen, Cagran and Rixon (2009) suggests that for many teachers of children the formal testing model predominates, even for the purposes of tracking progress. There is also the administrative requirement in some contexts for children to be summaratively judged by formal means for school transition purposes.

As a response to this state of affairs, the present volume's exploration of theory and practice within the context of the assessment of young learners aims to argue that assessing YLLs is beneficial *if it is done well* and if the methods, content, and delivery fit the original purpose of the assessment. YLL tests need to have a firm grounding in relevant theories of second language learning and development by children and teenagers, and be supported by a clear validation programme. A validity argument needs to be constructed for test use and interpretation of results, based on a principled theoretical framework. Such an approach can be found in Weir's (2005a) socio-cognitive framework, which advocates an evidence-based approach to test validation and provides a framework which offers the potential to bridge the theory practice divide (see further discussion in the 'methodological approach' section of this chapter). The present volume aims to argue that investment in young learners' language education will bring future benefits, and tests for younger learners (including the ones offered by Cambridge English) can have a positive impact on all stakeholders, including candidates, their teachers and the wider society (as will be discussed in more detail in Chapter 2 and Chapter 13).

The intended audience for the volume

In a world where language assessment is growing in importance, the number and type of stakeholders in the language testing field are increasing as growing

Introduction

numbers of people are directly involved in selecting and developing tests, in using test scores for decision-making purposes, and in preparing learners for tests, often in high-stakes contexts (Taylor 2000b). These include, firstly, the constituency of *professionals* in the language testing, educational assessment, and educational policy sectors who are involved in policy-making, curriculum development, materials design, teaching and assessment of young learners' second language development in a formal school context. The *academic* community working in child development, language education, child second language acquisition and second language assessment of young learners is another important group of stakeholders. Finally, *parents* form an important constituency of stakeholders through their involvement with their children's language education and qualifications. It is hoped that this volume will reach many of these constituencies and contribute to the language testing community's 'moral obligation' (Taylor 2009a, 2013) to share core knowledge, skills and understanding of good practice in assessment in a clear, comprehensive and accessible way. Let us consider these constituencies in turn and see in what ways they may find the volume and its contents useful.

The primary audience for this volume is professional language testers who are directly involved in the practical assessment of young learners' second language abilities. These include staff at examination boards and testing agencies at international, national, regional or local levels. These language testing professionals may find new insights into the knowledge, skills, expertise and experience that Cambridge English has accrued over the last 20 years in the field of YLL assessment, which are brought together in this volume. The collected information on research, development and validation activities underpinning Cambridge English Qualifications for YLLs could be a valuable source of information for similar validation activities that other examination boards and testing agencies may wish to consider carrying out in the future. Similarly, this body of information is meant as a stimulus for a critical evaluation of Cambridge English YLL examinations. Of course, the volume is not intended as a final authoritative account of children's and teenagers' second language ability constructs for the purposes of testing as currently conceptualised and operationalised by Cambridge English Qualifications for younger learners. On the contrary, our primary intention in bringing together and critically reflecting on state-of-the art practice and knowledge in the field is to offer the professional community a platform for discussion, critical thinking and constructive reconceptualisations of the issues that remain still under-researched or insufficiently understood regarding the assessment of young learners.

The volume is also directly relevant for the academic community, especially those with a specific interest in investigating young learners' second language development, including postgraduate students and academic researchers specialising in assessment issues. As Taylor (Ed) (2011) explains, the bridge between research and practice is not an easy one to build, largely due to the

Examining Young Learners

small-scale nature of most academic research carried out on the development of language abilities on the one hand, and the constraints that surround the large-scale industrialised assessment of the same language abilities on the other. Perhaps this is the main reason why there have been so few successful attempts so far to link academic research on language learning/acquisition with the practice of formally measuring the same language development in children and teenagers. A notable exception is the attempt by Second Language Acquisition and Testing in Europe (SLATE), a group combining academics and professionals with SLA expertise and language testing experience, albeit not specifically in the context of younger learners (Bartning, Martin and Vedder 2010). The questions, methodologies and instruments used and insights gained by academic researchers of language development and by practitioners in the language testing and assessment fields should inform each other's work and offer a much-needed cross-fertilisation of ideas and new questions for exploration in interdisciplinary research. We share Taylor's (2013:4) view that the explication of theory and practice in language assessment and second language development will lead to a broader and deeper understanding of the issues involved and will serve to strengthen the relationship between the academic and practitioner communities. We also hope this volume will help build a similar bridge between child and adolescent second language development and assessment research.

The intended audience for this volume includes staff at ministries of education and other educational authorities responsible for advising on the development of assessment policy and the implementation of examination reform programmes. In an era during which education has undergone radical reforms in the past quarter of a century in numerous countries and where the status of English has become decidedly established as the language of international communication in a globalised world, it is increasingly important that ministries of education are aware of the choices available to them and ensure that their decisions are well informed by up-to-date research and current theories of educational assessment, young learners' second language development and assessment of English as a second language.

No doubt, other types of readers may benefit from the volume. Among them are those who are directly involved in the teaching and assessment of young learners within the classroom, or in the preparation of learners for tests designed for younger candidates. This group of readers includes curriculum developers, textbook publishers, materials writers, teacher educators, language teachers and heads of language departments or schools with some influence of how assessment is carried out in their institutions. They may find certain sections of the book informative and some of the arguments, highlighted links, issues or practical ideas offered directly relevant to their own working contexts. Although the primary focus of this volume is the assessment of young second language learners, the child and pedagogical issues

Introduction

explored and the theory, principles and practice behind them will undoubtedly prove useful for the large and growing constituency of professionals involved in some way with the education of young learners. Similarly, even though the focus of the volume is on language assessment, and especially on English, the growing international influence of the content and language integrated learning (CLIL) approach, especially at the primary school level, means that a large constituency of teachers working within different language contexts may benefit from sections of this book.

The vast community of *parents* and carers of children and teenagers between the ages of 6 and 16 around the world are very much among the intended audiences, who may benefit from the discussions either directly or indirectly through channels such as *Cambridge English Parent News*. They may find some of the sections of the volume useful as a source of information to support their decision-making about the language education and qualifications that they choose for their children. Parents are now likely to expect more explicit evidence from test developers supporting claims about the quality, validity and usefulness of their exams. We would draw their attention particularly to Chapter 2 on the contexts and roles of assessment, and Chapter 3 on test taker characteristics.

Lastly, a group who may benefit from this volume is Cambridge English personnel and test material writers involved in developing and administering the *Young Learners* and *for Schools* exams. The discussion of issues fundamental to the assessment of YLLs will position their immediate responsibilities within the broader context of YLL assessment. The explication of theory and practice presented in this volume will hopefully deepen their understanding of the theoretical basis of Cambridge English tests for young learners and inform decisions made about specific task-contextual parameters which are fundamental in test development and play a crucial role in the dynamic relationship between task characteristics, test takers' cognitive abilities and scoring decisions. The present volume can also inform current and future development and revision projects carried out at Cambridge English.

In summary then, this volume is offered as a rich source of information for a wide variety of audiences on multiple aspects of examining YLL ability. For all audiences, then, the intention of this volume is to critically reflect on the practices adopted by a large examinations board, and as a result to inform and educate members of the wider testing community and related stakeholder constituencies engaged in the enterprise of testing children's and adolescents' language abilities. In doing so, the main argument developed in this volume will focus on the beneficial aspects of high-quality tests for young learners. Let us discuss each of these aims in more detail.

The volume's aim of *informing* readers is fulfilled by information gathering and sharing, including a discussion about the cognitive development underlying second language learning in young learners, the history of the

Examining Young Learners

development of Cambridge English tests for younger learners, a survey of alternative examinations for young learners, and a review of teaching materials and teacher qualifications for teaching young learners. Questions and issues such as why Cambridge English has been involved in the assessment of YLLs, what is tested in Cambridge English tests for young learners, how exams at different proficiency levels and for different age ranges differ and why, are explored, as well as the revisions which Cambridge English tests for children and teenagers have gone through in the light of historical, political, social, educational and environmental changes in the past 20 years. Case studies are presented of educational systems around the world and how Cambridge English exams have been incorporated into existing or new curricula, teaching, learning and assessment systems in those parts of the world. In addition, information on the technical qualities of Cambridge English tests for children and teenagers is offered, along with information on external and internal research activities surrounding these tests. Some of these aspects will include previously unpublished research documents, records and data contributing to the aim of information gathering, sharing, and critical evaluation.

The volume's aim of *educating* is seen in its explicit attempt to increase assessment literacy among the stakeholder community, mainly by including discussions on the development, research and validation activities surrounding Cambridge English YLL tests, explicitly reiterating the intended purposes and uses of results of these tests. In addition, for all engaged in the assessment and testing of YLLs, we provide in Chapter 14 a guide to good practice for the assessment of young learners. This is intended as a framework of the positive qualities that tests for children and teenagers should aspire to, i.e. setting out aspects of a test's validity within a principled framework. In addition, a clear position statement is offered on Cambridge English constructs defining YLL abilities that the Cambridge English tests for children and teenagers reflect, and which covers essential areas such as the relationship between construct and operationalisation, a test's fitness for purpose, assessment criteria and rating scales, as well as its impact and washback.

Finally, fulfilling the volume's aim of *persuading* is done through developing the argument that assessing YLLs is beneficial *if it is done well*, and that investment in young learners' language education will bring future benefits. We hope to demonstrate in the volume that Cambridge English tests for young learners aged between 6 and 16 have the desired test qualities and the intended positive impact on all stakeholders.

At present there is clearly a need for a volume based on current research to share information among the wider testing community, educate stakeholders and introduce more transparency about the development and validation of Cambridge English tests for YLLs. For instance, Cambridge English recently engaged in an evaluation of the *Progetto Lingue 2000* project in