The digital instructor for literacy learning

Catia Cucchiarini[1], *Ineke van de Craats*[2] , *Jan Deutekom*[3], *Helmer Strik*[1,2]

[1]Centre for Language and Speech Technology, Radboud University, Nijmegen, the Netherlands
[2]Department of Linguistics, Radboud University, Nijmegen, the Netherlands
[3]Department of Innovation, Friesland College, Leeuwarden, the Netherlands

c.cucchiarini@let.ru.nl, i.v.d.craats@let.ru.nl, J.Deutekom@fcroc.nl, w.strik@let.ru.nl

Abstract

The DigLIn project aims at providing concrete solutions for adult literacy students by developing and testing L2 literacy acquisition material in four different languages and by employing Automatic Speech Recognition (ASR) to analyse the learner's read speech output and provide feedback. We develop the technology, design sample exercises for different languages (Dutch, English, German and Finnish) and test them in literacy classes in adult education centres with adult L2 learners.
Existing language learning material for low-educated second language learners is augmented with an ASR module capable of recognizing what the learners say, of diagnosing possible errors in reading aloud or pronunciation and of providing practice and feedback in learning to read aloud in the L2.

Index Terms*: adult literacy learning, language and speech technology, second language acquisition*

1. Introduction

Europe has many immigrant and refugee adults with a low level of education, who lack basic skills such as reading and writing in both their native language and second language. However, the Common European Framework of Reference (CEFR) for Languages [5] departs from the basic level of primary school and implicitly assumes that adults are readers and writers. This does not correspond to reality. In addition to a high number of low-literate native-born adult residents, Europe counts many non-literate adults who need to learn to read and write for the first time, in a language other than their mother tongue. The numbers of non-literates and low-literates (unable to read and write well enough to use these skills in their daily lives) differ from country to country, but are between 10–15% of the population. Part of them are nonnatives who participate in (integration) courses with a literacy component. For instance, in Germany there were 65.000 of these immigrants between 2005 and 2012. 37,2 % of them are primarily illiterates. The DigLIn project aims to support this group of immigrant learners.

Being able to read and write is a prerequisite for active participation in society and employability. Poor oral and written proficiency in the second language (L2) leads to social exclusion [3], and prohibits social and economic integration [8]. Literacy – the ability to use reading and writing– is clearly a key factor in the integration and participation of immigrants in the society in which they live. Helping people acquire basic skills such as reading and writing is a crucial step in supporting social inclusion and citizen participation. Many European countries have programmes and initiatives aimed at promoting literacy acquisition and look for innovative methods that can boost the efficiency of literacy teaching by making it more flexible and more individualized.

In spite of these considerations, all kinds of financial cuts threaten adult education across the EU; in the UK for example, immigrant adults wishing to take ESOL classes exceed availability [4], and government funding continues to be reduced across the ESOL sector. This means that low-literate and non-literate adults are increasingly expected to be responsible for the costs of their individual learning. This implies that, more than ever before, speed of learning has become an essential factor for the least skilled L2 learners.

A Dutch study [14] aimed at assessing the learning load (in hours of instruction) for learning to read and write revealed two important findings. The time allotted to computer work – in which individuals had to work actively – correlated positively with reading scores, but the time allotted to whole group work correlated negatively. Moreover, there were substantial individual differences in pace across learners. The fast learners had additional opportunities to progress faster when there were facilities and materials to serve them. Facilities in this case means computer facilities and materials means ICT, CALL (Computer Assisted Language Learning) and multi-media.

Unfortunately, little attention is devoted to this group of learners. Publishers of course materials have little or no interest in this group because it is too small for gaining a reasonable profit and the development of computerized materials and multimedia adapted to this target group is expensive

2. Research background

Language and literacy development by first time (adult) readers in a second language is a new and underdeveloped domain of research [16] [18] [20]. Most research on reading concerns children who learn to read in their mother tongue. Adults who learn to read in an L2 with quite a different phonological system than that of their mother tongue face additional problems in mastering the phoneme-grapheme correspondences of the L2. These problems tend to mean that, in the Netherlands for example, many immigrants and refugee adults do not attain the level for reading and writing now required for integration and naturalisation [12] [13].

In basic reading instruction, two main approaches can be distinguished: the sight-word approach and the phonics approach. Finnish with its excellent correspondence between graphemes and phonemes lends itself to phonics instruction, and

due to its regular syllable structure, the syllable is most often focused on as a unit. For Dutch and German with a relatively transparent orthography, phonics instruction is preferred in both L1 and L2 literacy programs. For English with its deep orthography, a sight-word approach is possible, yet the alphabetic code still has to be cracked for decoding the many regularly spelt words. Models of beginning reading development agree on a first stage of direct-word recognition using basic visual cues, a second stage of indirectly mediated word recognition through graphic cues (grapheme-phoneme correspondence) and a third stage of direct word recognition based on automatization [11].

The question arises as to why immigrant adults are ultimately less successful than children. Important reasons are that L2 adults receive fewer hours of reading instruction, the course material is of a lower quality and the ICT applications now readily available to children are not appropriate for adults. In a class, children often decode graphemes aloud and synthesize them into a word. In adult classes, this is often seen as childish and is therefore restricted to an absolute minimum. Moreover, in a class of 10 or 15 adults, individual differences are often too large to make it a useful activity. The same holds for reading texts aloud: only one student reads, the others succeed in passively following to various degrees.

More active practice in which literacy students can produce the sounds or words while a computer tells them whether they are correct is a much needed improvement. This becomes possible through the application of ASR technology because the computer recognizes the word uttered and can provide feedback to the learner on whether the word was correctly read or not. There is a long history of experimentation on children using ASR [7] [15] [17] [21], but this technique has not yet been applied in adult literacy education. In addition, most of the studies on reading support through ASR concern systems that can follow the learner while reading aloud, but which are not aimed at identifying errors at the phoneme level to diagnose grapheme-phoneme connections. For this latter kind of application, more advanced technology is required.

.

3. CALL for literacy development

Computer Assisted Language Learning (CALL) applications offer enormous advantages compared to teacher-fronted classes: learners can practice as much as they want at their own pace in a stress-free environment and can receive individualized, adaptive feedback from the computer. This is particularly important for adult language learners who lack basic skills such as literacy and who can use materials when they are able to take time off from family and other responsibilities. Becoming literate in a second language can be particularly challenging and requires much practice and patience. A reader not only identifies letters (graphemes) and words (the analyzing part of the reading process), s/he also makes a correspondence with the sound (phoneme) represented by the grapheme and the sounds that together form a word (synthesis). Perception is only one side of the reading process; the learning reader also has to translate graphemes into sounds, combine them into words and produce them. Feedback is traditionally given by the teacher or another proficient reader, but could be provided individually for a large group of learners, by the computer as proposed here.

4. Digital Literacy Instructor (DigLIn)

The Lifelong Learning Program (LLP) project 'Digital Literacy Instructor' aims at providing concrete solutions for adult literacy students by developing L2 literacy acquisition material in four different languages and by employing Automatic Speech Recognition (ASR) to analyse the learner's read speech output and provide feedback. We develop the technology and design sample exercises for different languages (Dutch, English, German and Finnish) and test them in literacy classes in adult education centres with adult L2 learners.

Existing language learning material for non-literate and low-literate L2 learners developed at Friesland College (the digital sources of the FC Sprint^2 [6]) is augmented with an ASR module capable of recognizing what the learners say, of diagnosing possible errors in reading aloud or pronunciation and of providing practice and feedback in learning to read aloud in the L2.

This is a considerable improvement in comparison to existing systems in which learners can listen to audio recordings and carry out receptive exercises of the sound-to-grapheme type. In our system learners have the possibility of engaging in production exercises to learn and practice grapheme-to-sound or graphemes- to-word correspondences in the L2, reading a sound, a word, or a sentence out loud and receiving corrective feedback from the computer.

In addition to the already mentioned advantages of ASR-based CALL, it is important to underline the importance of a private, stress-free environment in L2 beginning reading and speech production, because low-literate language learners often feel ashamed of their weak skills and then refrain from practicing in the presence of teachers and other students. After a short introductory period, the system we develop can be used at home so that learners can feel comfortable and can practice anytime for as long as they want.

5. The pedagogical approach in DigLIn

The pedagogical approach in FC-Sprint^2

As explained above, in this project we depart from a common framework (digital sources of FC-Sprint²), and develop content and exercises in keeping with the specific features and requirements of the language and the teachers in question.
The concept of FC-Sprint² [6] is based on two pillars:

a. A different approach to students by teachers: from control by the teacher to autonomy for the students.

Students have to work with their resources, the teacher is the last resort.

b. Providing students with resources so that they can become more autonomous learners.

We try to build small programs so that student can find out themselves instead of being told by a teacher how it all works.

The principles underlying FC-Sprint^2 [6] can be summarized as:

- Start with high expectations as teachers who expect more get students who perform better.
- Students should carry responsibility (prevent passive behaviour).
- Learning efficiency grows if the student carries responsibility.
- Learning is doing what you cannot do yet.
- Students need to make mistakes in order to learn.
- Learning is more effective when students feel the need to learn.
- Students should first employ their own resources and ask for help when they need it.

When a student has been struggling with a certain subject the effect of instruction likely is much stronger than when a topic is completely new. So first we try to let students work with their own resources before a teacher explains.

- Talent is always an observation afterwards.

There actually is quite a lot of evidence that "talent" is at least a highly overrated concept and that achievement takes a lot of time and effort. Relying on talent can slow down development. (see [2] [9] [10])

- A student can learn everything until (s)he proves otherwise.
- The student is addicted to learning efficiency.
- Motivation is the result of a process.

Teachers can have a lot of influence on a student's motivation (negative and positive).

The DigLIn approach to literacy instruction

The pedagogical approach of FC-Sprint2 is translated in DigLIn by giving non-literate learners the materials for cracking the alphabetical code and providing them all necessary feedback. The teacher makes clear that (s)he is confident that learners will manage to read these words in a few days and will be able to show that to the whole class. As soon as learners have found out what the system can do for them, they will have the feeling of success and will be more and more motivated to continue.

The underlying method for a system like FC-Sprint^2 [6] and the one to be used in DigLIn is in fact a phonics-based method: the structure method. The primary aim of the structure method is grasping the structure of the spelling system or associating specific sounds (phonemes) with specific letters (graphemes). This is done on the basis of a whole word which is visually and auditorily structured in smaller units (analysis). In this way the student learns to consider a written word as a composite unit of separate elements and to make use of the systematic nature of letter-sound associations for autonomously decoding new words.

The basis of this method is a restricted number of concrete basic words the meaning of which is clear. In classes of 6- and 7-year-old children, those words are presented in a context of a story or a picture story and learnt by heart. In DigLIn those words can be made clear by pressing a button. Basic words should have a 'one-on-one grapheme-phoneme correspondence', that is to say that the sounds are not influenced in their pronunciation by preceding or following sounds or by the fact that they are in word-final or syllable-final position, as is the case in Dutch. We use the label "pure sound".

Examples for:

- English: dad, map, mop, jump, bin, big, yes
- Dutch: mat, kap, kip, boom
- German: Rat, Hut, Oma
- Finnish: eno, iso, akka

Ideally, there is a one-to-one relationship between phoneme- and grapheme Many languages have too few graphemes for the repertoire of phonemes, which is the case for Dutch, but more particularly for English with one and the same grapheme representing different phonemes.

As soon as a couple of basic words are recognized, the analysis and synthesis exercises can start. The spoken word is analyzed in sounds, the written word in letters. Next, the sounds are blended to a spoken word. Many analysis and blending exercises are needed for establishing a tight association between sound and letter. Software can help to automatize this phase of the reading process. For this stage, FC-Sprint^2 has found many challenging exercises with feedback (e.g., a letter dragged to an incorrect position, does not stay, but jumps away, back to its original position). An example of such a drag-and-drop task for a Danish version of the system is given in Figure 1:

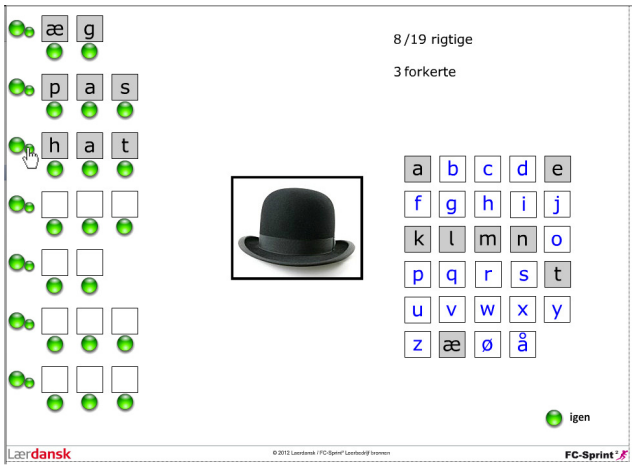

Figure 1. From letters to words, a drag-and-drop exercise for Danish

In this example a student can drag letters to the right square. On the left hand side students can hear the complete word by clicking on the big green button. By hovering over the little green button a student can see what the word means. By clicking on the button below the square he/she can hear the individual sounds of the letters. If a student drags a wrong letter to the square the letter jumps back to its original position and a "mistake" sound is heard.

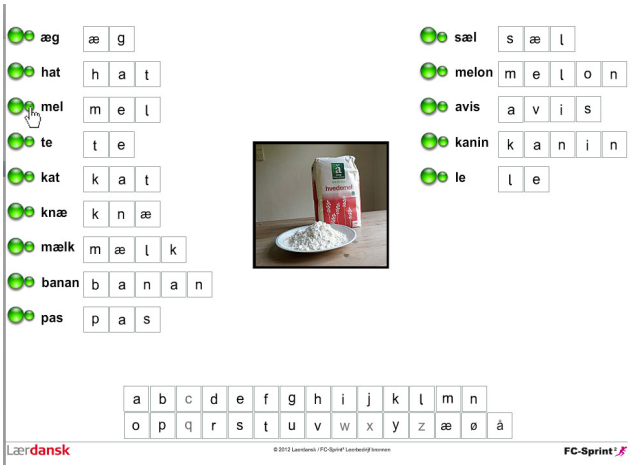

Figure 2. Presentation of the words for Danish with sound bar

In Figure 2 students have a "sound bar" with all phonemes at their disposal at the bottom of the screen. By clicking on a letter in this sound bar students can hear the individual sounds of the complete alphabet. By clicking on the big green buttons students can hear the sound of the word. Hovering over the little green buttons gives students a photo displaying the word. By clicking on the letter squares behind the words students can hear the individual sounds corresponding to the letters. (See for the working examples: http://diglin.eu/?page_id=222.)

In DigLIn we provide feedback on reading aloud the blended words. The step to reading new words and the transition from spelling words to a more automatized stage is supported first by exercises in which either the onset or the rime is kept constant, as illustrated in Table 1, and later, by a mix of these words.

Table 1. Example of building up complexity of the word structure for Dutch

CVC with same rime	CCVC with same onset	CCVC with same onset
p-ak	st-ok	sch-ool
z-ak	st-ak	sch-aap
b-ak	st-op	sch-ep
t-ak	st-ip	sch-ip
l-ak	st-ik	sch-uur
v-ak	st-ap	sch-oen

FC-Sprint[2] materials offer possibilities for analyzing and blending and for training automatization. Feedback on reading aloud the words has to be built in by applying ASR.
In addition to the autonomy stimulating approach adopted in FC-Sprint[2], we provide a limited set of video-taped instructions, so as to allow students to work quite alone with the program, if they wish to do so, without the support of classmates and/or teachers.

5.1.1. Criteria for selecting words

Words are selected taking into account an order of increasing complexity. Words also target adult LESLLA (Low-Educated adult Second Language and Literacy) readers who are still at an early stage of reading, e.g., the glance and guess stage. Because preferably photographs, but also pictures are used for explaining the meaning, words should be concrete content words so that pictures can be attached. Frequency is also a selection criterion, but is applied with caution since many frequent words are function words that cannot be represented by images (and are more difficult to understand). Frequency lists are also of less importance because they are not based on what low-literate immigrants are likely to encounter. Systematic stress variation in polysyllabic words is also taken into account. Together with the criteria for usefulness for literacy instruction we come to the following criteria for building up complexity.

5.1.2. Criteria for building up complexity

Start with:
1. CV(C) words
2. "Pure sound'' words
3. Maximal difference: first cardinal vowels: /i/, /u/, /a/ occurring in most languages of the target group of learners (so, not /y/). Followed by consonants that are maximally different on the basis of other features.
4. No minimal consonant pairs in one word or series of words for reasons of auditory similarity (not: **p**ak and **b**ak) or visual similarity (not: **d**ak and **b**ak).

And proceed with:
1. Vowels and consonants from maximally different (/a/-/u/-/i/) to minimally different (/i/-/I/ or /u/-/Y/) and from very common in other languages to language-specific sounds (e.g., for Dutch ui in huis ('house').
2. From CVC to CCVC or CVCC and more extensive consonant clusters
3. From monosyllabic to disyllabic words then polysyllabic words
4. From concrete to abstract words
5. From noun to adjective and verb
6. From pure sound to spelling conventions (e.g. in Dutch for open and closed syllables: raam-ramen)
7. From word to sentence

6. Language and speech technology

Innovative in the DigLIn project is that within the CALL system for literacy training use is made of language and speech technology' (LST), and especially 'Automatic Speech Recognition' (ASR). As is well known, developing ASR-based applications for L2 learners implies having to deal with non-native speech which, for many reasons, is more challenging than native speech [1] [19].

Therefore, exercises are developed such that the possible answers by the users are restricted (see e.g. the screen shot of FC-Sprint[2]). For every item, a list of correct and incorrect responses is used to limit the recognition task. This can be achieved in different ways: by using confidence measures to identify an utterance in the list of possible responses, or by using the list of responses to train constrained language models. In

doing so, care is taken to also include a number of possible meta-responses, such as the equivalents of "I don't understand" or "what?".

The DigLIn system is intended to be web-based, and should run in different browsers. Since practical, technical details can be important for a good performance, we carefully look at issues such as head-sets, audio recording settings (for different browsers), audio file formats, signal-to-noise ratio (SNR), and noise cancelling (techniques).

When an error has been identified in the learner's response, feedback is provided to signal this to the learner. With respect to the spoken responses, feedback is provided on two levels: (1) on the utterance level, and (2) on the error level. Regarding the former, the speech recognition module determines which utterance was spoken, and before proceeding to error identification the learner is given feedback on the recognized utterance. After all, it would be highly confusing if the learner gets feedback on (parts of) an utterance that was not spoken at all by the learner. It is also more confusing if the system signals an error while the response was correct (false alarm), than v.v. (false accept). Therefore, in tuning the system we try to keep the number of false alarms smaller than the number of false accepts.

Feedback is gradual in the sense that it indicates the degree of correctness. A student can repeat again and again and a slider indicates in real time whether there is any improvement so that the student can try again immediately and see whether the new attempt is better or worse. The feedback should be simple, intuitive, and easy to interpret, such as a score presented visually (e.g. a bar, possibly with colors).

While for many languages databases of native speech are available, corresponding databases of non-native speech are in general lacking, especially non-native speech for the target groups of the application. This makes it even more challenging to develop ASR technology for this application. In DigLIn, we cope with this issue in the following way. We start with an ASR trained on native material, using native resources (lexica, speech corpora, etc.). Later we study whether using extra information can improve the system's performance. Possibilities are to use non-native resources (lexica, speech corpora, etc.), and to use information on errors made by the target group (annotations of errors). Available non-native audio recordings and error annotations are first used, while interactions of users with (initial versions of) the system, and annotations of (part of) these recordings will be employed at a later stage.

Learners can also listen to correct examples in stored audio recordings. Students can repeat the speech they listen to in the program as often as they want. We carefully considered criteria for these audio recordings, such as normal speed, careful speech (no or limited amount of reduction), sounds natural, limited amount of silence, whether or not carrier sentences should be used, good selection of speakers (male and female, amount of dialect, etc.), recording environment and conditions (studio, 'silent office'), technical specifications (e.g. file format (wav/mp3), signal-to-noise ratio (SNR), etc.). The reason for presenting the speech in the program at normal speed is to provide a "jump" from the slow speech usually spoken by teachers to real world speech.

At SLaTE 2013 we intend to show a preliminary version of the system, illustrating the feasibility of the exercises, the type of practice the learners receive and the corrective feedback provided by the system. Possible additional features and their pedagogical relevance are discussed.

7. Conclusions

ASR seems to constitute a valuable add-on to current computer-based adult literacy programs for various reasons. The nature of the language tasks involved is such that constrained ASR tasks can be designed, which in turn guarantees adequate ASR performance. For the first time, this makes it possible for learners to receive automatic, immediate feedback on their reading performance, without learners having to make comparisons themselves between what they heard and they produced themselves. This is an important improvement for L2 reading instruction, which paves the way to more autonomous learning conditions.

8. Acknowledgements

This project has been funded with support from the European Commission under project number 527536-LLP-1-2012-NL-GRUNDTVIG-GMP. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

We are indebted to the other members of the DigLIn team for their contributions, in alphabetical order: Marta Dawidowicz, Vanja de Lint, Maisa Martin, Jan-Willem Overal, Karen Schramm, Taina Tammelin-Laine, Joost van Doremalen and Martha Young-Scholten.

9. References

[1] Benzeghiba, M., R. D. Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens, Automatic speech recognition and speech variability: a review, Speech Communication, vol. 49, no. 10-11, pp. 763-786, 2007.

[2] Bloom, B. (1985) Developing talent in young People, Ballantine Books.

[3] Bynner, J. (2001). *Outline of the research, exploratory analysis and summary of the main results*. London: Centre for Longitudinal Studies.

[4] Cooke, M. (2010) ESOL in the United Kingdom. Paper at the inaugural EU-Speak workshop, Newcastle, 6 November.

[5] Council of Europe (2001). *A Common European Framwork of Reference for Languages: Learning, teaching,, assessment.*. Cambridge: Cambridge University Press.

[6] Deutekom, J. FC-Sprint², Grenzeloos Leren, Boom 2008.

[7] Duchateau J. Kong, Y. Cleuren, L. Latacz, L., Roelens, J., Samir, A., Demuynck, K., Ghesquière, P., Verhelst, W., Van hamme, H. (2009) Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules. Speech Communication 51(10): 985-994.

[8] Dustmann, C. and F. Fabbri (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal*, 113, 695-717.

[9] Dweck, C. (2006) Mindset: The New Psychology of Success, Random House Publishing Group,

[10] Ericsson, A. (2006) The Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press.

[11] Kurvers, J. (2007). Development of word recognition skills of adult L2 beginning readers. In N. Faux (ed.) Low-Educated Second Language and Literacy Acquisition. Richmond, Virginia: The Literacy Institute at Virginia Commonwealth University, 23-43.

[12] Kurvers, J. and I. van de Craats (2009). Het Haalbaarheidsonderzoek van De Voortwijzer [Feasability Assessment]. Tilburg: Universiteit van Tilburg.
[13] Kurvers, J., W.Stockmann, & I. van de Craats (2010). Predictors of success in adult L2 literacy acquisition. In Th. Wall & M. Leong (eds.). Low-educated Adult Second Language and Literacy Acquisition. Banff: Bow Valley College: 64-79.
[14] Kurvers, J. and W. Stockmann (2009). Alfabetisering in beeld. Leerlast en succesfactoren. [Literacy in the picture. Learning load and success factors.] Tilburg: Universiteit van Tilburg.
[15] Mostow, J., Roth, S., Hauptmann, A.G., Kane, M., A Prototype Reading Coach that Listens. Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-1994), pp. 785-792, 1994.
[16] Roll, H., Schramm, K. (Hrsg.), Alphabetisierung in der Zweitsprache Deutsch, Osnabrücker Beiträge zur Sprachtheorie, 77, Duisburg: Gilles & Francke, pp. 5-10, 2010.
[17] Russell, M., D'Arcy, S. (2007) Challenges for computer recognition of children²s speech, In proc. Of SLaTE-2007, pp. 108-111.
[18] Tarone, E., M. Bigelow, & K. Hansen (2009). Literacy and Second Language Oracy. Oxford: Oxford University Press.
[19] van Doremalen, J., Cucchiarini, C. and Strik, H., Optimizing automatic speech recognition for low-proficient non-native speakers. EURASIP Journal on Audio, Speech, and Music Processing, Article ID 973954, 13 pages, 2010.
[20] Wagner, D.A. (2004). Literacy (ies), culture(s), and development(s): The ethnographic challenge. *Reading Research Quarterly,* 39, 234-241.
[21] Li, Y., and Mostow, J. (2012). Evaluating and improving real-time tracking of children's oral reading. In Proceedings of the 25th Florida Artificial Intelligence Research Society Conference (FLAIRS-25), Marco Island, Florida.

Lifelong
Learning
Programme