Using mobile technology for reading assessment

Loukia Taxitari*, Claudia Cappa†, Marcello Ferro*, Claudia Marzi*, Andrea Nadalini*, Vito Pirrelli*

*Institute for Computational Linguistics "Antonio Zampolli", CNR, Pisa - Italy
email: name.surname@ilc.cnr.it

†Institute of Clinical Physiology, CNR, Pisa - Italy
email: claudia.cappa@cnr.it

2020 6th IEEE Congress on Information Science and Technology (CiSt) | 978-1-7281-6646-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/CIST49399.2021.9357173

***Abstract*—The enormous potential of Information and Communication Technologies (ICT) for addressing critical educational issues is generally acknowledged, but its use in the assessment of the complex skills of reading and understanding a text has been very limited to date. The paper contrasts traditional reading assessment protocols with *ReadLet*, an ICT platform with a tablet front-end, designed to support online monitoring of silent and oral reading abilities in early graders. ReadLet makes use of cloud computing and mobile technology for large-scale data collection and allows the time alignment of the child's reading behaviour with texts tagged using Natural Language Processing (NLP) tools. Initial findings replicate established benchmarks from the psycholinguistic literature on reading in both typically and atypically developing children, making the application a new ground-breaking approach in the evaluation of reading skills.**

***Index Terms*—reading assessment, reading research, mobile technology, NLP, cloud computing, special education needs.**

I. Introduction

Educational systems strive to teach children core reading skills as a way to lay down the basis for all subsequent learning, as students with early reading problems face serious difficulties with learning in general. In fact, reading is as such an extremely complex task to learn. In every two seconds of silent reading, a good English reader is demanded to accomplish a variety of nested sub-processes [1]: 1) decode, focus and access up to 10 words and their meanings; 2) parse an entire clause and form a complex meaning unit or *proposition*; 3) connect the new meaning unit to a growing *network* of propositions forming a *conceptual model* of the text being read; 4) check the text model against personal goals and background expectations; 5) monitor comprehension and make appropriate inferences.

Such a variety of concurrent processes taking place in a tiny fraction of time may interact and possibly go wrong in a number of ways. Over the last decades, education experts, scholars in language and cognition, psycholinguists and neuroscientists have been at the forefront of a huge effort to understand the mechanisms driving this interaction, with a view to providing evidence-based practices for improving reading competences. Aspects of reading have been investigated from a number of perspectives, with different methods and goals [2]–[4].

We gratefully acknowledge the financial support of PRIN grant 2017W8HFRX "ReadLet: reading to understand. An ICT-driven, large-scale investigation of early grade children's reading strategies" (2020-2022), from the Italian Ministry of University and Research.

On a functional level, research on reading has focused on the high-level processes taking place during reading and what is required for them to be smoothly integrated. This perspective has spawned top-down models of reading, among which the *simple view of reading* [5], [6] has probably been the most fertile and influential one. The model assumes that reading consists of *word decoding* and *language comprehension*, and focuses on how the two can be scored both independently and jointly, to evaluate reading proficiency and recommend personalized training practices.

On the psycholinguistic level, reading research investigates the explanatory mechanisms subserving reading, and accounts for how failures in one or more such mechanisms can affect reading performance. At the same time, upon observing reading errors, one can understand what mechanisms are performing sub-optimally. One such model, the *dual model of reading* [7] assumes that written-to-spoken-word conversion is mainly channelled through two routes. The first is triggered by familiar words, and involves access to representations in an orthographic input lexicon, associated with word meanings and output phonological representations. The second route involves grapheme-to-phoneme mapping rules that bypass the lexicon. A failure in activating the lexical route produces regularization errors in decoding words whose orthography is inconsistently mapped onto sounds (e.g. *one* is pronounced as in *bone*). If problems affect conversion rules, only words that are already stored in the orthographic input lexicon are read fluently.

Finally, on the neuro-biological level, neuro-functional models of reading strive to associate brain areas with critical reading processes [2]. Analysis of brain lesions, neuroimaging data, and correlation patterns with disturbed reading performance provide evidence for the involvement of both general-purpose and highly-specialized right- and left-hemisphere brain areas in reading and comprehension [8].

The three levels (functional, psycholinguistic and neurobiological) are strongly reminiscent of David Marr's [9] hierarchy of levels for understanding vision, originally intended to focus on *what*, *how* and *where* (in the brain) characterizes a complex cognitive system. Of late, in many linguistic and cognitive domains there has been growing interest in the potential for between-level interaction, with a view to interdisciplinary synergy. Reading research can considerably benefit from this convergence. In fact, in spite of substantial

978-1-7281-6646-9/20/$31.00 ©2020 IEEE

progress in our understanding of the cognitive underpinnings of reading and the variety of factors that occasionally make reading very difficult for a child, there is still a long way to go before these factors are effectively used in everyday educational practices. We contend that ICT can play a fundamental role in this process. First, computer modeling of carefully collected and classified reading data provides the bridging layer between high-level reading functions (Marr's level 1) and their algorithmic interaction (Marr's level 2) [10]. Secondly, artificial neural networks can be used as a functional interface between psycholinguistic processes and their neuro-anatomical correlates (Marr's level 3), with a view to tighter interdisciplinary integration [11]. In this paper, we illustrate the fertility of this approach by constrasting traditional reading assessment protocols with *ReadLet* (https://www.readlet.it/), an ICT platform with a tablet front-end, designed to support online monitoring of silent and oral reading abilities [12], [13].

II. Reading Assessment Protocols

Current reading assessments aim to evaluate students' reading fluency (i.e. accuracy and speed of reading), reading comprehension (i.e. text understanding), or occasionally both. Different reading tests are created to serve specific purposes. Some are used by clinicians to assess whether a child has reading difficulties or learning disorders. Others are used by educators to assess children's reading level, often as a group. On the basis of the goal of each test, its parameters can change drastically, from the testing of single words and non-words or letter-to-phoneme exercises to comprehension tests through open-ended questions, or based on the combination of information from texts of different genres, as in the PISA tests[1]. To have as complete a reading profile of the child as possible, the main task used, e.g. reading of a passage or a cloze task, is complemented by other tasks, such as letter-to-sound correspondence, vocabulary or oral comprehension tests [14]. Variation of test parameters can result in different assessment outcomes for the same child [14]–[18]. For example, a cloze test is more dependent on basic word recognition skills and decoding than does a test with longer passages to read, as the latter provides more background context to the child in order to extract meaning [19], [20].

Traditionally, reading assessments are paper and pencil tests, but more recently computerized versions of them have been created. These are either identical to the original tests, or they are developed to include more technologically advanced functions, such as automated scoring (for example the RISE test in the USA[2]), or adjustments in the difficulty level of the test on the basis of the child's performance (for example the New Group Reading Test[3]). Available reading tests are aimed at children of a certain age range, e.g. primary school, and different test items are created for each grade. This makes comparisons across children of different ages or grades challenging, and requires demanding psychometric approaches in data analysis in order to be reliably computed [21]. It is noteworthy that the text passages and items (words or sublexical units) used in reading assessments are not controlled for complexity or difficulty, and their suitability for different ages/grades is assessed through pilot studies or test standardization with large numbers of children. Additionally, texts do not seem to be controlled for the various factors reported in the psycholinguistic literature to affect reading comprehension and fluency. For example, they are not controlled for the frequency or the morphological complexity of the words they are made up of, which are crucial factors affecting the performance of children while reading. Conversely, psycholinguistic research on reading and reading development seems to have followed a parallel route to the development of reading assessments, focusing mainly on the reading of individual words in isolation. This makes the results of psycholinguistic research, based on established linguistic and cognitive benchmarks such as frequency of words and *n*-grams, neighbourhood density, morphological complexity and predictability, difficult to compare with assessments of long passage reading.

III. *ReadLet*: an ICT platform for reading assessment

ReadLet is an Italian research project that aims to create the technological, infrastructural and educational conditions for continuous assessment of children reading in schooling activities, based on real-time screening of reading skills, and data-modelling and evaluation of their impact on learning. The approach is implemented through four activity strands (Figure 1): i) multimodal data collection, ii) time-text alignment, iii) evidence-based assessment and data modelling iv) reading profiles and personalized teaching recommendations.

ReadLet can record large streams of time-bound, multimodal reading data, capturing the real-time behaviour of a child that reads a short story on a tablet, and finger-points to the words displayed on the tablet touch screen while reading (Figure 1, purple bin). Data streams of the child's finger-tracking, video and audio recording are sent to a centralized server through an internet connection, where they are encrypted and anonymized for privacy protection, post-processed and time-aligned with the text (Figure 1, greenish bin). After this stage, ReadLet's modeling provides i) the estimated reading time for each letter, syllable, word and sentence in the text, ii) the time taken to answer each question, iii) the number of correct answers, iv) the number and type of decoding errors (Figure 1, grey bin).

Meanwhile, a battery of NLP tools annotate all words in the text for levels of linguistic analysis: from articulatory complexity and phonological transparency, to part-of-speech tagging, lexical typicality (in terms of density and entropy of a word's lexical neighbourhood), orthotactic probability (as a function of a word's bigram and trigram probabilities), morphological complexity, token and type frequency, token's

[1]Organization for Economic Co-operation and Development, "Programme for International Student Assessment", https://www.oecd.org/pisa/, accessed on June 28^{th}, 2020.

[2]Utah State Board of Education, "Readiness, Improvement, Success, Empowerment", https://utahrise.org/, accessed on June 28^{th}, 2020.

[3]GL Assessment, "New Group Reading Test", https://www.gl-assessment.co.uk/products/new-group-reading-test-ngrt/, accessed on June 28^{th}, 2020.

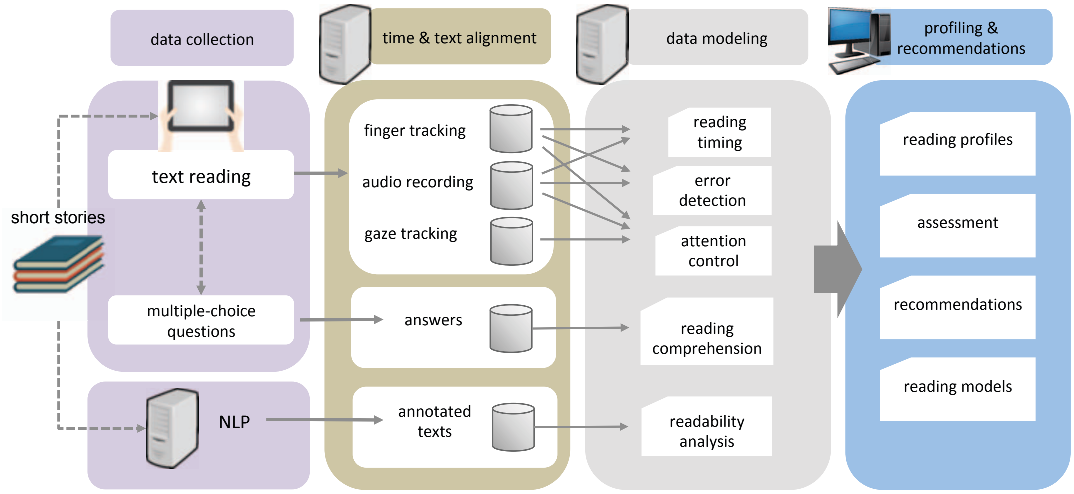

Fig. 1. An overview of ReadLet logical architecture.

position and syntactic role in a sentence. Linguistically annotated texts can be controlled and classified for levels of readability [22]. In addition, by relating children's reading performance to specific linguistic units/features in the text, ReadLet enables a better understanding of the factors accounting for children's reading strategy, suggesting ways to enhance their strengths and overcome their weaknesses. For example, reading speed is known to correlate with frequency. In analyzing developmental data, it is nonetheless important to understand whether a child's reading speed is more affected by the frequency distribution of letter chunks (e.g. bigrams or trigrams), word morphs or whole words. If the distribution of letter chunks is found to explain more of the child's reading speed than the distribution of morphs or words, the child is most likely using grapheme-to-phoneme decoding rules, which are known to slow down reading of longer words. Conversely, a dominant word frequency effect would indicate a lexical reading strategy, which makes the child more proficient in reading longer high-frequency words, but still slow in reading morphologically complex, rare words. Finally, sensitivity to (the frequency of) morpheme-sized units can make reading long, complex and rare words fluent.

IV. Evaluation of the Tool

A. Tablets in research

Tablets are ubiquitous devices, easy to use and fairly cheap to purchase. Recently, these characteristics have paved the way to their use in data collection during the study of cognitive development in toddlers and older children [23], [24]. These studies have reported high standards of data validity for tablet data collection, as well as a high correspondence between the results obtained with the tablet and similar results obtained through other methods (e.g. eye-tracking). A recent article published in *Nature Communications* [25] has shown that finger movements on a tablet screen (*finger-tracking*), can work as a proxy for eye movements and attention, with high correlations between finger-tracking and eye-tracking data in both healthy adults and adult autistic or brain-damaged patients. The studies above suggest that tablets are currently a very efficient solution in collecting large-scale data quickly and reliably, meeting both research and clinical objectives.

B. ReadLet Pilot Phase

A pilot collection campaign of reading sessions was conducted in Italy and Southern Switzerland.[4]

1) Participants and Procedure: Four hundred and thirty two children participated in the campaign, from grades 3 to 6, including typically developing children and 32 children with various learning difficulties (here referred to as *atypical*). During the study, children were presented with two short stories, which they had to read either silently or aloud. In both reading conditions, children were instructed to finger-track the text they were reading on the tablet screen, as is common practice for emergent readers to support directional movement and focus attention. After reading, children were prompted to answer a few multiple-choice questions related to the text content. Questions were presented on the tablet one at a time, with readers being allowed to go back to the short story and look for relevant passages. In Figure 1, this is shown by the two-pointed arrow linking "text reading" to "multiple-choice questions".

2) Measures: Finger-tracking *coverage*, corresponding to the percentage of text that children finger-tracked on each page of the story, was used as an index of children's engagement in the task. The *token tracking time* was calculated for each word in the text, corresponding to the total time taken by a child in finger-tracking a specific word token.

3) Results: Preliminary analyses of the pilot data confirmed a few behavioural findings reported in the reading literature. Figure 2 shows tracking coverage by page for typically and atypically developing children, in the two reading conditions

[4]Data collection was made possible thanks to the *AEREST* project, funded by the Department of Teaching and Learning of the University of Applied Sciences and Arts of Southern Switzerland (SUPSI).

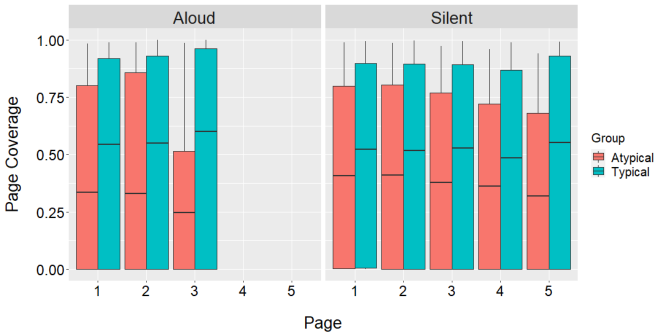

Fig. 2. Finger tracking coverage by document page for typical (light blue) and atypical (red) children in aloud (left) and silent (right) reading.

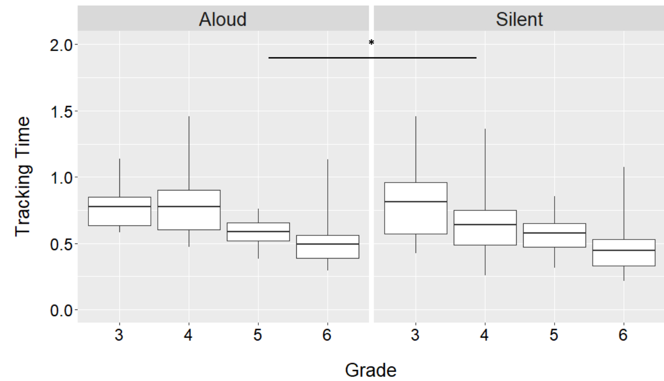

Fig. 3. Token tracking time by grade in the silent and aloud reading condition for typical children. The single-star bar indicates a .05 significance level.

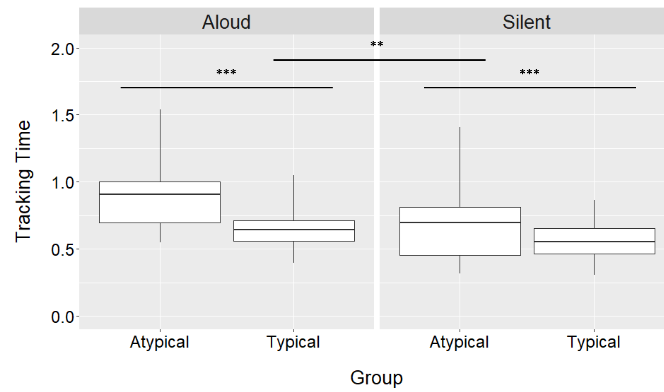

Fig. 4. Token tracking time by reading condition in typically and atypically developing children. A two-star bar indicates a .01 significance level, a three-star bar indicates a .001 significance level.

(silent and aloud). A Univariate ANOVA with coverage as a dependent variable, and group, reading condition and page number as fixed factors showed a significant main effect of group (i.e. typical vs. atypical children, $F(1, 2746) = 70.67$, $p < .001$, $\eta_p^2 = .025$), as well as an interaction between group and reading condition ($F(1, 2746) = 6.01$, $p = .014$, $\eta_p^2 = .002$), but no other effects or interactions. A separate between-group comparison in each condition confirmed that, in both oral and silent reading, typically developing children finger-tracked more text than atypically developing children (aloud: $t(93) = 3.88$, $p < .001$, silent: $t(96) = 3.05$, $p = .001$). Although further analyses are needed, the evidence appears to support a connection between attention problems and reading and learning difficulties. Difficult readers are more likely to experience fatigue and more prone to lapses in focused attention, particularly when concurrent working memory demands increase [26]. This is confirmed by the difference in coverage between silent and oral reading, due to the extra cognitive load of concurrent overt articulation.

Figure 3 shows data from 112 typically developing children (Grade 3: N = 13, Mean Age: 8.4 (SD: 0.5), Grade 4: N = 29, Mean Age: 9.2 (SD: 0.4), Grade 5: N = 21, Mean age: 10.1 (SD: 0.4), Grade 6: N = 49, Mean Age = 11.3 (SD: 0.5)) without any reported learning or physical difficulties and after the exclusion of sessions with technical problems, Additionally a conservative document finger coverage criterion (75%) was applied as an index of children's attention and level of engagement in the task. Children with lower coverage were not included in the analysis. The data were entered in a Univariate ANOVA with tracking time as a dependent variable, and grade and reading type as fixed factors. Figure 3 shows the mean finger-tracking time by grade in the silent and the reading aloud conditions, with overall shorter tracking times for the silent condition ($F(1, 213) = 4.49$, $p = .035$, $\eta_p^2 = .021$), as expected from previous findings reporting longer reading times when the child is required to read a text orally due to increased processing demands [27]. Figure 3 also shows how tracking time is reduced with grade in both reading conditions, with shorter times as children grow ($F(3, 213) = 29.84$, $p < .001$, $\eta_p^2 = .296$), replicating previous findings reporting faster reading in older children [28]. Additionally, post-hoc LSD tests compared silent reading to reading aloud for each grade separately and showed that in the reading aloud condition the 3^{rd} and 4^{th} grades did not differ significantly, as well as the 5^{th} and 6^{th} grade, but all other comparisons were significantly different ($p < .01$). In the silent condition, all grades significantly differed from each other ($p < .01$), except the 4^{th} from the 5^{th} grade.

In Figure 4, we compared data from 32 typically developing children (Mean Age = 9.9, SD = 0.9) and 32 children with learning difficulties (Mean Age = 10.0, SD = 1.1), as reported by teachers. A Univariate ANOVA with tracking time as a dependent variable, and group and reading type as fixed factors confirmed that reading aloud takes longer to track than silent reading ($F(1, 81) = 7.88$, $p = .006$, $\eta_p^2 = .089$), as the analysis of the whole group above, but also showed that children with learning difficulties have longer tracking times than typically developing children ($F(1, 81) = 13.12$, $p < .001$, $\eta_p^2 = .139$). The difference between the two groups is larger in the reading aloud condition, with typical children tracking faster than atypical ones ($t(23) = 2.93$, $p < .01$), but no such difference is present in the silent condition ($t(23) = 1.86$, $p = .08$). This suggests that the extra load of articulating the text is posing a bigger challenge for children with learning difficulties

TABLE I
SPEARMAN *rho* CORRELATIONS BETWEEN TRACKING TIME AND WORD LENGTH/FREQUENCY FOR TYPICAL CHILDREN BY GRADE (CORRELATIONS ARE SIGNIFICANT AT THE .001 LEVEL).

	Length	**Frequency**
Grade 3	.45	-.38
Grade 4	.49	-.43
Grade 5	.50	-.42
Grade 6	.46	-.36

TABLE II
SPEARMAN *rho* CORRELATIONS BETWEEN TRACKING TIME AND WORD LENGTH/FREQUENCY FOR ALL GRADES BY ATYPICAL/TYPICAL CHILDREN (CORRELATIONS ARE SIGNIFICANT AT THE .001 LEVEL).

	Length	**Frequency**
Atypical	.41	-.34
Typical	.47	-.39

than for typically developing children, as also found in the coverage analysis above. Finally Tables I and II show the results of Spearman *rho* correlations between tracking times and word length/frequency, only for typical children in the four grades, and for both typical and atypical children (for all grades) respectively. All correlations were highly significant, indicating that tracking times increase as the length of a word increases, but they decrease as the frequency of a word increases. Both of these effects have been reported in the reading literature, although their interaction and developmental interplay is more intriguing than simple correlations suggest [28], [29]. Importantly, however, the novel tablet methodology is sensitive enough to capture these well-established effects in the psycholinguistic literature and allow them to manifest themselves through the behavioural measure of finger-tracking.

V. IMPLICATIONS FOR READING RESEARCH AND TEACHING

In all those cases where educational results depend on the orchestration of numerous skills, conclusions and recommendations for practice can only be built upon assessment of long, longitudinal screenings of a large population of graders. In reading research and teaching, the poor availability of ICT tools for large data collection has made it prohibitive to a) replicate lab results at scale, b) assess educational results of research findings with pre-test and post-test designs, and c) make recommendations for long duration education practices addressing specific groups of struggling readers, as lamented by major international organizations [30], [31]. In this section we summarise the impact that ICT technology can have on current reading assessment and research.

NLP technology provides a large array of tools for annotating orthographic, phonological, sublexical, lexical and supralexical units and features known to raise potential problems in reading natural texts. Automatic linguistic annotation makes it possible to increase the ecological validity of protocols for reading assessment. Reading sessions can be based on connected texts rather than words or sentences in isolation, while controlling for the linguistic variables and the cognitive factors that may affect text readability. This has a few important implications. First, not only can we ascertain *if* a child is a difficult reader, but also *what* in the text may cause the reported difficulties, focusing on the linguistic nature of reading problems. Reading assessment is broken down into language levels and key reading skills, resulting in a fine-grained profile (unlike most existing protocols, which offer a single score). This is particularly relevant for children with learning disorders who often "mask" their difficulties by using compensating strategies to overcome their problems. For example, dyslexic readers might use morphological parsing to identify sublexical units without the need to decode words lexically, thus accelerating their reading rate, when in fact they do not master whole-word processing [32]. ReadLet can work around the difficulty of assessing these children; by monitoring tracking coverage, pauses or regressions, it can in fact produce a dynamic profile of the child's reading process across linguistic levels, rather than a result-oriented score. Process-based evidence has in fact been shown to indicate potential reading difficulties even in the absence of recoding errors [33], and can raise a red flag to the teacher or clinician. Finally, ReadLet can also address the reader's Vygotskian zone of *proximal development*, by training the child on an increasingly more complex set of the linguistic phenomena that were shown to cause problems. This is likely to help children overcome specific reading problems through personalized training, while building self-confidence and a positive approach to reading.

Our mobile technology can be used in almost any environment, with no need for a data-acquisition specialist or special equipment bound to provoke anxiety, particularly for a child who is learning to read. We know that reading probes are a commonly used instrument in the early grades for monitoring progress in reading fluency and text comprehension [34]. However, the time and effort to collect such data are huge. The use of a tablet for extended reading enables deriving this information unobtrusively and continuously, wherever the child fancies reading, even when she/he is at home engaged in studying a textbook or enjoying a fictional narrative.

Finally, ReadLet cloud computing architecture supports highly parallelized and distributed processes of multimodal data acquisition, which can be delivered in real time to education centers, research institutions and clinical centers as terminals for massive data harvesting and quantitative analysis. Time-aligned multimodal data streams (text, audio and video recording and finger-tracking) can push data analysis beyond state-of-the-art accuracy. In addition, since processing is carried out remotely, computing performance ceilings are only set by internet connection capacity. Large-scale studies can thus be conducted, paving the way to more generalizable results than ever in the past. In addition, the concrete possibility to take single-subject measurements on more occasions and in different environments makes ReadLet data usable not only in group studies but also for individual profiling purposes, as required in typical schooling settings.

VI. Concluding remarks and future directions

Ensuring that all children achieve a core level of literacy skills irrespective of socio-economic, cultural, religious, cognitive and gender differences, is an ambitious goal, calling for a broad range of coordinated actions. Undoubtedly, targeted interventions have to include both evidence-based novel education practices, and evidence-informed educational policies addressing equity in education as their main focus. Integration of different ICT strands proves to offer an interesting infrastructure for harvesting and modeling reading data. The approach can effectively be extended beyond reading assessment, to address issues of personalized training at home and circumvent the problems of culturally/socially impoverished environments, as well as special education needs.

References

[1] W. Grabe and F. L. Stoller, *Teaching and researching reading*, 3rd ed. Routledge, 2019.

[2] S. Dehaene, *Reading in the brain. The science and evolution of a human invention*. Viking Penguin, 2009.

[3] L. Maffei, *Elogio della parola*. Il Mulino, 2018.

[4] M. Wolf, *Reader, come home. The reading brain in a digital world*. Harper Collins, 2018.

[5] W. A. Hoover and P. B. Gough, "The simple view of reading," *Reading and Writing: An Interdisciplinary Journal*, vol. 2, pp. 127–160, 1990.

[6] K. Nation, "Children's reading difficulties, language, and reflections on the simple view of reading," *Australian Journal of Learning Difficulties*, vol. 24, no. 1, pp. 47–73, 2019. [Online]. Available: https://doi.org/10.1080/19404158.2019.1609272

[7] M. Coltheart, K. Rastle, C. Perry, R. Langdon, and J. Ziegler, "DRC: a dual route cascaded model of visual word recognition and reading aloud." *Psychological review*, vol. 108, no. 1, p. 204, 2001.

[8] C. J. Price, "A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading," *Neuroimage*, vol. 62, pp. 816–847, 2012.

[9] D. Marr, *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman and Company, 1982.

[10] V. Pirrelli, C. Marzi, M. Ferro, F. A. Cardillo, H. Baayen, and P. Milin, "Psycho-computational modelling of the mental lexicon," in *Word knowledge and word usage: a cross-disciplinary guide to the mental lexicon*, V. Pirrelli, I. Plag, and W. U. Dressler, Eds. Mouton De Gruyter, 2020, vol. 337, pp. 23–82.

[11] C. Marzi and V. Pirrelli, "A neuro-computational approach to understanding the mental lexicon," *Journal of Cognitive Science*, vol. 16, no. 4, pp. 493–535, 2015.

[12] M. Ferro, C. Cappa, S. Giulivi, C. Marzi, F. A. Cardillo, and V. Pirrelli, "ReadLet: an ICT platform for the assessment of reading efficiency in early graders." Edmonton, Alberta (Canada): 11th International Conference on the Mental Lexicon, 25-29 September, 2018 2018, p. 61.

[13] M. Ferro, C. Cappa, S. Giulivi, C. Marzi, O. Nahli, F. A. Cardillo, and V. Pirrelli, "Readlet: Reading for understanding," in *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 2018, pp. 1–6.

[14] M. Westerveld, "Measuring reading comprehension ability in children: Factors influencing test performance," *Acquiring Knowledge in Speech, Language and Hearing*, vol. 11, pp. 81–84, 11 2009.

[15] L. E. Cutting and H. S. Scarborough, "Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured," *Scientific Studies of Reading*, vol. 10, no. 3, pp. 277–299, 2006. [Online]. Available: https://doi.org/10.1207/s1532799xssr1003_5

[16] J. M. Keenan, R. S. Betjemann, and R. K. Olson, "Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension," *Scientific Studies of Reading*, vol. 12, no. 3, pp. 281–300, 2008. [Online]. Available: https://doi.org/10.1080/10888430802132279

[17] S. H. Eason, L. F. Goldberg, K. M. Young, M. C. Geist, and L. E. Cutting, "Reader-Text Interactions: How Differential Text and Question Types Influence Cognitive Skills Needed for Reading Comprehension," *Journal of educational psychology*, vol. 104, no. 3, pp. 515–528, aug 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26566295https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640191/

[18] T. O'Reilly, J. Weeks, J. Sabatini, L. Halderman, and J. Steinberg, "Designing Reading Comprehension Assessments for Reading Interventions: How a Theoretically Motivated Assessment Can Serve as an Outcome Measure," *Educational Psychology Review*, vol. 26, no. 3, pp. 403–424, 2014. [Online]. Available: https://doi.org/10.1007/s10648-014-9269-z

[19] K. Nation and M. Snowling, "Assessing reading difficulties: the validity and utility of current measures of reading skill," *British Journal of Educational Psychology*, vol. 67, no. 3, 1997.

[20] D. J. Francis, J. M. Fletcher, H. W. Catts, and J. B. Tomblin, *Dimensions Affecting the Assessment of Reading Comprehension.*, ser. Center for improvement of early reading achievement (CIERA). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2005.

[21] J. Sabatini, J. Weeks, T. O'Reilly, K. Bruce, J. Steinberg, and S.-F. Chao, "Sara reading components tests, rise forms: Technical adequacy and test design, 3rd edition," *ETS Research Report Series*, vol. 2019, no. 1, pp. 1–30, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ets2.12269

[22] F. Dell'Orletta, S. Montemagni, and G. Venturi, "READ–IT: Assessing readability of Italian texts with a view to text simplification," in *Proceedings of the second workshop on speech and language processing for assistive technologies*, 2011, pp. 73–83.

[23] M. C. Frank, E. Sugarman, A. C. Horowitz, M. L. Lewis, and D. Yurovsky, "Using tablets to collect data from young children," *Journal of Cognition and Development*, vol. 17, no. 1, pp. 1–17, 2016. [Online]. Available: https://doi.org/10.1080/15248372.2015.1061528

[24] K. Semmelmann, M. Nordt, K. Sommer, R. Röhnke, L. Mount, H. Prüfer, S. Terwiel, T. W. Meissner, K. Koldewyn, and S. Weigelt, "U Can Touch This: How Tablets Can Be Used to Study Cognitive Development," *Frontiers in psychology*, vol. 7, p. 1021, jul 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27458414https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4935681/

[25] G. Lio, R. Fadda, G. Doneddu, J.-R. Duhamel, and A. Sirigu, "Digit-tracking as a new tactile interface for visual perception analysis," *Nature Communications*, vol. 10, no. 5392, pp. 1–13, 2019.

[26] E. Commodari, "Novice readers: the role of focused, selective, distributed and alternating attention at the first year of the academic curriculum," *i-Perception*, vol. 8, no. 4, 2017. [Online]. Available: https://doi.org/10.1177/2041669517718557

[27] C. Vorstius, R. Radach, and C. J. Lonigan, "Eye movements in developing readers: A comparison of silent and oral sentence reading," *Visual Cognition*, vol. 22, no. 3-4, pp. 458–485, 2014. [Online]. Available: https://doi.org/10.1080/13506285.2014.881445

[28] P. Zoccolotti, M. De Luca, G. Di Filippo, A. Judica, and M. Martelli, "Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability," *Reading and Writing*, vol. 22, no. 9, pp. 965–992, 2009.

[29] M. De Luca, L. Barca, C. Burani, and P. Zoccolotti, "The effect of word length and other sublexical, lexical, and semantic variables on developmental reading deficits," *Cognitive and Behavioral Neurology*, vol. 21, no. 4, pp. 227–235, 2008.

[30] Y. Chzhen, A. Gromada, G. Rees, J. Cuesta, and Z. Bruckauf, "An unfair start: Inequality in children's education in rich countries," UNICEF Office of Research, Innocenti, Florence, Tech. Rep. 15, 2018.

[31] OECD, "Equity in education: Breaking down barriers to social mobility," OECD, OECD Publishing, Paris, https://doi.org/10.1787/9789264073234-en, 2018.

[32] C. Burani, S. Marcolini, M. De Luca, and P. Zoccolotti, "Morpheme-based reading aloud: evidence from dyslexic and skilled Italian readers." *Cognition*, vol. 108, no. 1, pp. 243–262, jul 2008.

[33] M. Nilsson Benfatto, G. Öqvist Seimyr, J. Ygge, T. Pansell, A. Rydberg, and C. Jacobson, "Screening for dyslexia using eye tracking during reading," *PLOS ONE*, vol. 11, no. 12, pp. 1–16, 12 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0165508

[34] M. Miura Wayman, T. Wallace, H. I. Wiley, R. Tichá, and C. A. Espin, "Literature synthesis on curriculum-based measurement in reading," *The Journal of Special Education*, vol. 41, no. 2, pp. 85–120, 2007.