Evaluating the Efficacy of a Serious Game in Enhancing Word Reading Speed

Chiara Piazzalunga[1], Linda Greta Dui[1], Stefania Fontolan[2], Sandro Franceschini[2], Marisa Bortolozzo[2], Cristiano Termine[2], Simona Ferrante[1]
[1]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
[2]Department of Medicine and Surgery, University of Insubria, Varese, Italy
chiara.piazzalunga@polimi.it
lindagreta.dui@polimi.it
stefania.fontolan@gmail.com
sandrofranceschini@gmail.com
marisa.bortolozzo@googlemail.com
cristiano.termine@uninsubria.it
simona.ferrante@polimi.it

Abstract: Dyslexia is a Specific Learning Disorder (SLD) characterized by subpar reading abilities in terms of accuracy and/or speed. Dyslexia can have a lifelong negative impact on academic and day-to-day life. However, early identification and subsequent training can help dyslexic children overcome their reading difficulties. For instance, tachistoscopic reading, a technique that involves presenting words for a brief period, has been shown to improve reading speed. Technological tools, such as serious games, can also be useful, and governments encourage their use to enhance the management of SLDs. In this study, we investigated the efficacy of a serious game called Tachistoscopio in improving reading speed in children. The game began with a short calibration process, followed by the presentation of a word for a brief duration. The user was then asked to write the word. We tested it on children, carrying out the pictorial system usability scale to assess usability and investigating the effect of class and gender through Mann-Whitney U tests. We determined it had excellent usability and could effectively adapt to users' abilities, as third graders performed significantly better than second graders ($p<0.001$). Following the success of this preliminary phase, 36 second-grade children, split into two groups, participated in a training study: 18 children underwent a three-week training phase with the game, while the remaining 18 children served as the control group. Before the training, we assessed all children's reading abilities using standardized word reading and vocabulary tests. The training involved four 20-item sessions per week. We then collected the reading performance of both groups again and used Mann-Whitney and interval estimates to test for statistically significant difference between the reading speed increments in the two groups. Their confidence intervals did not overlap (training: [0.11; 0.33], control: [-0.10; 0.05]), which suggests a significant difference between them. These findings suggest that Tachistoscopio may be an effective intervention tool for improving reading speed in dyslexic children.

Keywords: serious game, dyslexia, learning, reading speed, videogame

1. Introduction

Literacy is a social determinant of health: people with low literacy skills are up to three times more likely to experience poor health outcomes, ranging from having less health knowledge, lower general health, and shorter life spans (DeWalt *et al.*, 2004). Nonetheless the efforts to provide adequate education to children, 5% to 20% of children cannot reach their peers' reading level (Wagner *et al.*, 2020), thus being affected by developmental dyslexia (Snowling, Hulme and Nation, 2020). Dyslexia is characterized either by slowness in reading or by several errors, globally making text reading and comprehension very demanding. Its diagnosis is usually belated after reading is supposed to be mastered (end of 2nd grade), hence a pre-clinical intervention – starting from schools – could be fundamental to start helping these children to improve in such a fundamental task.
Among different remediation attempts, it was reported that tachistoscopic reading helps improving reading speed (Lorusso *et al.*, 2005). Tachistoscopic reading is a technique in which a word is shown for a very brief amount of time, and the subject is asked to repeat it after it has disappeared. It promotes the global visuo-lexical reading, as the brief presentation time forces words to be read as a unit, instead of decomposing them into phonological sub-units, as it happens in character-by-character reading.
The stimulus presentation time is an important parameter to be tuned. If it is too long, it is useless; if it is too short, it is frustrating; if it does not vary according to readers' performance, they do not improve. Relevant

differences between letters presentation time may be small, hence requiring the support of a technological device. This point is also highly recommended by guidelines on the screening and characterization of specific learning disabilities in general, and on dyslexia in particular (Associazione Italiana di Psicologia *et al.*, 2021), as traditional paper-based assessment cannot provide sufficient information about the reasons for reading problems.
Moreover, considering the young age of its potential users, engagement is also a key element for effectively proposing this tool. Therefore, a suitable solution would be to implement the tachistoscope reading a serious game. In the past, a webapp was proposed to allow a remote assessment (Lorusso and Fumagalli, 2020). Even if webapps seem to make games widely available, in the case of training for reading difficulty their flexibility turns into a drawback: as it is not possible to control an important parameter, that is, words dimension, the game cannot be played on screens smaller than laptops, such as tablets. Instead, computers should be discouraged when performing this activity, as they cannot be compared to paper when used for learning (whilst tablet can) (Sage, Krebs and Grove, 2019) and, on the contrary, they cause more visual fatigue (Lee, Chiang and Hsiao, 2021). Moreover, data fruition is neither straightforward, nor targeted to teachers, as in the case of school-based training.
In the context of the ESSENCE H2020 project (www.essence2020.eu), several serious games to enhance reading, writing and calculation abilities were implemented, together with a narrative framework aimed at boosting engagement without affecting the "serious" purpose of the games (Piazzalunga, Donati and Ferrante, 2023). One of these games is the Tachistoscopio. The theoretical added value of such games is the possibility of gathering data on children's activity and to provide informative feedback to teachers, that could include them in a school-based training program. According to the training results, they can decide whether a clinical visit should be envisaged. To demonstrate the validity of the game in training children reading abilities, a validation process should be carried out. In particular, it should be confirmed that the game is capable of modelling children's abilities, it is perceived as useful and engaging by children, and it is effective in training them at school. Indeed, given the pre-clinical purpose of the game, one must investigate whether effective results could be achieved in short time spans, as an excessive number of game administration (e.g., 45 minutes, twice a week, for four months (Lorusso *et al.*, 2005)) would make school-level adoption not feasible.
Given these premises, this work has three aims: (1) to assess whether the Tachistoscopio game is sensitive to children writing abilities; (2) to assess the usability of the game; (3) to assess short term training effectiveness.

2. Materials and Methods

To answer the aims of the work, the first step was to develop the Tachistoscopio game. Then, two testing phases were organized: first an early testing involved second and third graders, to understand if different reading proficiency was captured by the game (aim 1) and to understand if it was usable from their perspective (aim 2); then, a field testing was performed with second graders, that were trained with the Tachistoscopio for a three-week period, to understand if any effect was evident with respect to a control group (aim 3). Subjects were recruited in the framework of the ESSENCE project, according to the protocol approved by Politecnico di Milano's Ethical Committee (n. 04/2021 for early testing, n. 36/2022 for training), in primary schools in the province of Varese. To assure that a training effect could be observed, a minimum of 9 subjects needed to be recruited. The sample size was computed based on previous studies on similar protocols (He, Legge and Yu, 2013), that assessed statistically significant differences between pre- and post-reading training, by considering a 95% level of confidence. Given the longitudinal design of the study, this figure was doubled to account for potential dropouts.
Statistical analysis was carried out in Python 3.8 and in Matlab R2020b. For all statistical analyses, the significance level α was set to 0.05.

2.1 Gamification of the tachistoscope

What differentiates serious games from the mere digitalization of tests is gamification, a set of techniques aimed at exploiting human competitiveness and ambition as strategies to make the user accomplish the task in a more enjoyable way (Bedwell *et al.*, 2012). However, when developing serious games to train learning abilities, one must be careful not to excessively alter the rationale and the mechanics, as this may lead to a loss of effectiveness. To preserve the efficacy of the training, gamification elements should be external to the core mechanics of the games. In the ESSENCE context, serious games incorporate a guide character, unique settings for each learning domain, a narrative framework, and a system of coin collection that rewards accurate execution of exercises. Tachistoscopio, the game described in this work, shares these key elements. The game's

difficulty dynamically adjusts based on the child's performance, ensuring a constant sense of challenge that encourages the player to commit to the task. Moreover, it takes place in a forest, the designated setting for reading games, and features Falchetto, the forest's guide character, who provides tutorials. Falchetto is shown in Figure 1.
In particular, the story behind Tachistoscopio is the following: Falchetto, which is responsible for the collection and safeguard of the letters in the game world, is worried because a messy anteater has gone into his nest and has caused them to flee. The player is thus required to read words before they disappear, “capturing” the fugitive letters. It is worthy to note that these characters appear in other ESSENCE games as well.

Figure 1: Falchetto, the reading domain's guide character.

2.2 Game design and description

The game has been developed in Unity 2020.1.6f1 and built for Android on a Samsung S6 Lite. It is a gamified version of the tachistoscope, an instrument used to show a word for very brief intervals.
The game was developed through a co-design process (J.E. *et al.*, 2011), during which all relevant stakeholders were involved both to identify the functional specifications and the parameters regulating the appearance of the words, and to iteratively test and refine the game. The list of presented words was provided by clinicians, who selected them based on their use frequency, length and complexity.
The first part of the game is calibration. In this phase, a word appears on the screen and the player must push a button when they are done reading it. The time needed to read each character is then computed as the time needed to read the word divided by its length. This value constitutes the reading delay, i.e., the time that needs to elapse before a character is read, and will influence the rate with which words disappear, to modulate the difficulty of the game. This procedure is repeated 10 times, and the final reading delay is computed as the average of the 10 individual values.
After the calibration, the game scene is loaded. In this scene, a word appears on the screen and remains visible for an amount of time equal to the reading delay times the number of characters in the word. Regarding the word’s disappearance, the game mechanics vary slightly depending on the value of the reading speed: if it is under 3 syllables/second, the word is cancelled letter-by-letter, otherwise it disappears all at once. Some noise appears on the screen to avoid the imprinting of the afterimage of the word on the retina. The two different cases are shown in Figure 2a and Figure 2b.
Then, the player must write the word in a pop-up. If the typed word is correct, positive feedback appears. After three consecutive successful attempts, the reading delay is decreased by 6 ms/character, producing an increment in the speed of cancellation. Conversely, after three errors the reading delay is increased by 6 ms/character.

(a) (b)

Figure 2: (a) Letter-by-letter and (b) global cancellation of words.

2.3 Early testing

2.3.1 Protocol

The preliminary testing phase for the Tachistoscope game was designed to ensure that the game was not only suitable for children but could also effectively adapt to their abilities. The testing phase involved second and third graders who played the game in a 10-item session after the calibration phase. During this session, the game recorded the initial and final reading speed achieved by the children in syllables/second and their accuracy.
To gather more information about the children, they were asked to complete a characterization questionnaire. They were asked about their dominant hand, mother tongue, and familiarity and habits with tablets and video games.
Finally, the children were administered the Pictorial System Usability Scale (P-SUS) (Baumgartner *et al.*, 2019). This was done to assess their overall satisfaction with the game. Moreover, a custom questionnaire, investigating how easy, boring and fun the game was, has been administered to understand the level of appreciation. Answers to this questionnaire were encoded on a 5-point Likert scale.

2.3.2 Statistical analysis

Firstly, to determine if the data was normally distributed, the Lilliefors test was utilized. Following the normality test, the Mann Whitney U test was conducted to examine the impact of gender and grade on game variables.

2.4 Training

2.4.1 Protocol

Children were firstly administered a series of tests to assess their reading capabilities and phonological memory. These tests are the De.Co.Ne. for words reading [13].
To be eligible for the study, reading abilities should not be affected by potential cognitive impairment. Hence, exclusion criteria were: scored below the 25th percentile in the CPM Raven matrices test (Raven, Raven and Court, 1998; Belacchi *et al.*, 2008) or scored less than or equal to 6 in the vocabulary test of the WISC-IV test (Wechsler, 2003).
Children who met the inclusion criteria were randomly assigned to a train group or to a control group. Children assigned to the control group did not receive specific training. Children assigned to the train group underwent a training phase with the Tachistoscopio game. They played the game in 20-item sessions four days a week for three consecutive weeks. Each session was conducted under the guidance of a clinician, who ensured that the children understood the rules of the game and were playing it correctly.
After the training period, both children groups repeated the reading tests.

2.4.2 Statistical analysis

Firstly, the Lilliefors test was used to check whether data about the reading speeds, both pre- and post-training, achieved in the word reading tests were normally distributed. Then, the increment in word reading speed was computed as the difference between the reading speed measured in the second time point and the initial

reading speed, and t-test for normally distributed data or a Mann-Whitney test for non-normal data was performed to investigate if there were any differences between the trained group and the control group.
In case of uncertainty, confidence intervals were computed as follows:

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

where $\bar{x}$ is the sample mean, z is the confidence level value, s is the sample standard deviation, and n is the sample size. In the case the variable is not normally distributed, the sample mean was replaced by the median and the standard deviation by the inter-quartile range. To obtain a confidence interval of 95%, which corresponds to the chosen significance level α = 0.05, z was set to 1.96. If confidence intervals of the two samples do not overlap, it suggests that there is a statistically significant difference between the two populations (Cumming, 2013).
To check whether the performance of children who played with the Tachistoscopio increased over time, a Spearman's rank correlation coefficient test was performed on the time needed to read a character, which should be decreasing as the performance increases. The Spearman's rank correlation coefficient test can verify whether there is a monotonic relationship between two variables, in this case, time and milliseconds per character.

3. Results

3.1 Early testing

3.1.1 Sample

28 children participated in the study: 13 (7 females, 6 males) attended second grade, while 15 (8 females, 7 males) attended third grade.

3.1.2 Statistical analysis

Variables were not normally distributed, so the Mann-Whitney U test was used.
The results of the Mann-Whitney U test show a significant difference between the accuracy reached by second and third graders (p=0.005), with older children achieving better performances. The same trend is observed in reading speed, which is significantly lower in second graders than in third graders (second grade: 1.30±0.62 syllables/second, third grade: 2.24±0.67 syllables/second, p<0.001). There was no significant difference between males and females, neither in accuracy (p=0.740) nor in reading speed (p=0.052). Finally, the increment in reading speed at the beginning and at the end of the game is statistically different between grades as well (p=0.003), but not between genders (p=0.079). Boxplots for the three variables are shown in Figure 3a, 3b and 3c.

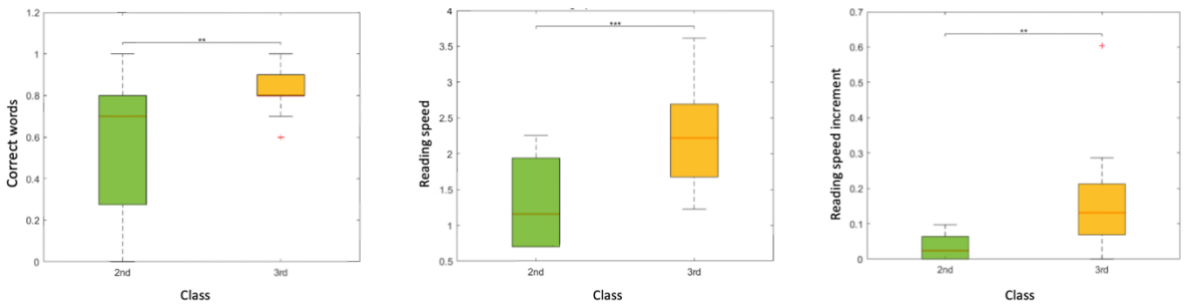

Figure 3: (a) Boxplot of the accuracy in second and third grade, (b) boxplot of the reading speed in second and third grade, (c) boxplot of the reading speed increment in second and third grade.

3.1.3 Usability and appreciation

The SUS score obtained by the game was 86.2±14.4, which is considered an excellent usability score, as the minimum value to consider a tool usable is 68, and the excellence threshold is set at 85 ('Determining what individual SUS scores mean: adding an adjective rating scale', 2009). Regarding the appreciation, children were asked to express on a scale how easy (1 = very difficult, 5 = very easy), fun (1 = not fun at all, 5 = very fun), and

boring (1 = not boring at all, 5 = very boring) they found the game. In Table 1, median values and lower and upper quartiles are reported for each question.

Table 1: Median scores and IQRs of appreciation questions.

	Median [Q1; Q3]
How easy was the game?	3 [3; 4]
How fun was the game?	5 [5; 5]
How boring was the game?	1 [1; 1]

3.2 Training

3.2.1 Sample

36 second-grade children participated in the study. 18 children were administered the Tachistoscopio training (9 females, 9 males), while the remaining 18 composed the control group (12 females, 8 males).

3.2.2 Statistical analysis

According to the Lilliefors test, the word reading speed variable did not result to be normally distributed. Therefore, the Mann-Whitney U test was carried out to compare the performance of the trained and control groups (independent variable) in the De.Co.Ne. test, separately for the initial word reading speed (before training, dependent variable) and the final word reading speed (after training, dependent variable). None of these two dependent variables were statistically different between the trained group and the control group. In Table 2, we report their median values, along with their lower and upper quartiles and the p-value of the Mann-Whitney U test.

Table 2: Median values and IQRs of reading speeds and reading speed increment, with the p-values of the Mann-Whitney U test performed between trained and control groups.

	Median [Q1; Q3] trained group	Median [Q1; Q3] control group	p-value
Initial word reading speed (syllables/second)	2.03 [1.53; 2.35]	2.06 [1.53; 2.41]	0.612
Final word reading speed (syllables/second)	1.95 [1.66; 2.64]	1.85 [1.75; 2.16]	0.788

The increment between initial and final word reading speed (dependent variable) was normally distributed, and thus compared through a t-test between the trained and control groups (independent variable), with a result of $t(36) = 1.84$, $p=0.074$.
Considering that the speed increment was close to statistical significance, the variable was further studied by means of its confidence interval of the mean. It was found that the confidence interval of the mean speed increment was [0.11; 0.33] for the train group and [-0.10; 0.05] for the control group. As the confidence intervals did not overlap, the trend suggested by the t-test was confirmed. It means that the reading speed increment is significantly higher in the trained group.

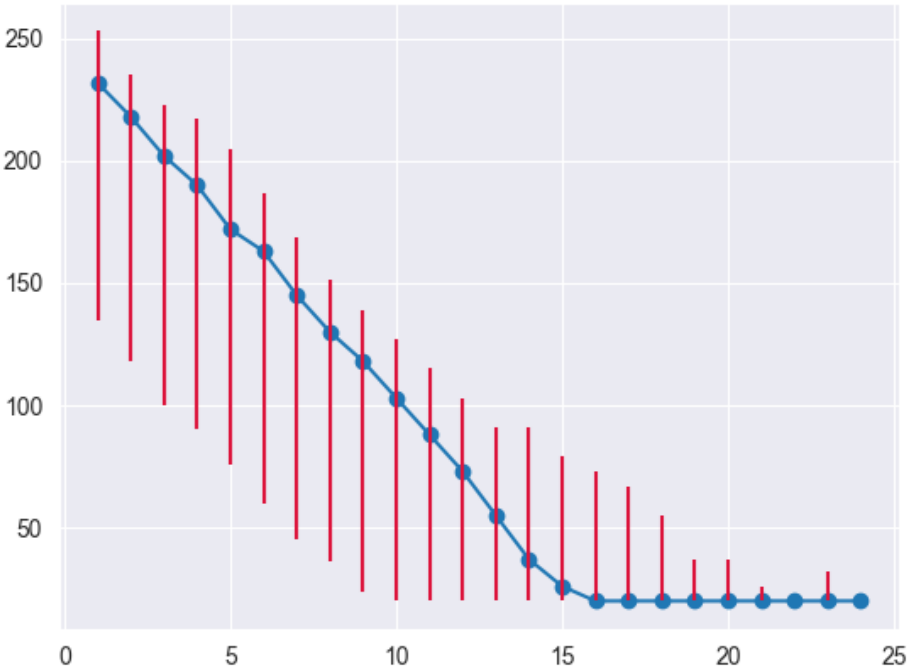

Figure 4: Plot of the median time needed to read a character during the whole training period. In red, the IQR of each measurement.

To verify whether the performance in the Tachistoscopio game increased over time, we considered the mean time needed to read a character. Its trend can be observed in Figure 4. The number of sessions resulted to be significantly correlated to the time needed to read a character (milliseconds per character), with rho = -0.974, $p<0.001$, thus implying that the game produces a significant increase on the reading ability.

4. Discussion.

In this work, a tablet app for training reading abilities was developed and preliminary validation was performed. The co-design process was successful both in terms of appropriateness of game design and in terms of users' experience. As for game design, early testing results on second and third graders revealed that it is sensitive to different reading proficiency levels expected in two successive years of school education (aim 1). Specifically, third graders were significantly more accurate, faster, and improved more than their younger colleagues during game usage. Sensitivity to proficiency is a key element for the game to be effective, but longitudinal assessment to the same children is necessary to understand its efficacy in training. Hence, to promote repeated assessment, the second aim of the work was to understand if the game was usable. Usability was excellent, and the game was found fun and not boring at all. Both these results (appropriateness and usability) paved the way to the second phase of the study, i.e., the training phase.

Given these encouraging results, a training phase was organized to understand if a short-term training could improve reading abilities (aim 3). As the confidence interval of the mean increment of speed in the trained group and in the control group did not overlap, a group effect was found. This means that the Tachistoscopio game was effective in improving reading speed in children, even if training lasted three weeks only. This is a time frame way shorter than the usual training time found in literature (Lorusso *et al.*, 2005), that makes the game even more suitable for a school-level adoption, as it allows an intervention without excessive burden on other activities. However, the result was not supported by statistical tests. Even if significance was not reached on this parameter, there is an evident trend that confirms an important role of the app in improving reading speed. Probably, a slightly longer training period or an increase in the sample size would have confirmed the trend also by leveraging classical statistics.

Given the promising results, the ESSENCE H2020 project (www.essence2020.eu) ended with a strong request from teachers to have the developed serious games deployed to continue performing reading training for children in need. This is an important achievement, as schools are the first place where learning difficulty should be addressed, before referring children to clinicians, as early identification and training of learning weaknesses

assure better efficacy of the training (Kadar *et al.*, 2020). Moreover, objective data coming from the game on reading speed and speed increment after treatment are useful parameters to evaluate children and better characterize their weakness, as fostered by national guidelines on specific learning disabilities diagnosis (Associazione Italiana di Psicologia *et al.*, 2021), as could be smart pens for dysgraphia (Dui *et al.*, 2021), eye tracking for dyslexia (Piazzalunga *et al.*, 2023), and so on.
As future work, the same training could be performed on dyslexic children to understand their reaction to training, and if further refinements are necessary to adapt also to this population.
In conclusion, the co-design and validation of the Tachistoscopio serious game was successful and paves the way towards a school-based training of weak abilities.

Acknowledgements

This research was funded by European Union's Horizon 2020 research and innovation programme under the call SC1-PHE-CORONAVIRUS-2020-2B, grant number 101016112.

References

Associazione Italiana di Psicologia *et al.* (2021) 'Linea Guida sulla gestione dei Disturbi Specifici dell'Apprendimento'. Roma.

Baumgartner, J. *et al.* (2019) 'Pictorial System Usability Scale (P-SUS): Developing an instrument for measuring perceived usability', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/3290605.3300299.

Belacchi, C. *et al.* (2008) *CPM – Coloured Progressive Matrices. Standardizzazione Italiana*. Edited by Giunti Organizzazioni Speciali. Firenza.

Cumming, G. (2013) *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge. doi: 10.4324/9780203807002.

Bangor, A., Kortum, P., & Miller, J. (2009) "Determining what individual SUS scores mean: adding an adjective rating scale." Journal of Usability Studies, Vol 4, No. 3, pp 114-123.

Dewalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N., & Pignone, M. P. (2004). Literacy and health outcomes: a systematic review of the literature. Journal of general internal medicine, Vol 19, No. 12, pp 1228-1239. https://doi.org/10.1111/j.1525-1497.2004.40153.x

L. G. Dui, E. Calogero, M. Malavolti, C. Termine, M. Matteucci and S. Ferrante, "Digital Tools for Handwriting Proficiency Evaluation in Children," 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece, 2021, pp. 1-4, doi: 10.1109/BHI50953.2021.9508539.

Franceschini, S. *et al.* (2016) *Batteria De. Co. Ne. per la lettura. Strumenti per la valutazione delle abilità di lettura nelle scuole primarie*. Edited by Centro Studi Erickson.

He, Y., Legge, G. E., & Yu, D. (2013). Sensory and cognitive influences on the training-related improvement of reading speed in peripheral vision. Journal of vision, Vol 13, No. 7, p 14. https://doi.org/10.1167/13.7.14.

van Gemert-Pijnen, J. E., Nijland, N., van Limburg, M., Ossebaard, H. C., Kelders, S. M., Eysenbach, G., & Seydel, E. R. (2011). A holistic framework to improve the uptake and impact of eHealth technologies. Journal of medical Internet research, Vol 13, No. 4. https://doi.org/10.2196/jmir.1672.

Kadar, M. *et al.* (2020) "A systematic review of occupational therapy intervention for handwriting skills in 4–6 year old children", Australian Occupational Therapy Journal, Vol 7, No. 1, pp 3-12. doi: 10.1111/1440-1630.12626.

Lee, C. C., Chiang, H. Sen and Hsiao, M. H. (2021) 'Effects of screen size and visual presentation on visual fatigue based on regional brain wave activity', Journal of Supercomputing, Vol 77, No. 5, pp 4831–4851. doi: 10.1007/s11227-020-03458-w.

Lorusso, M. L. *et al.* (2005) 'Tachistoscopic treatment of dyslexia changes the distribution of visual-spatial attention', Brain and Cognition, Vol 57, No. 2, pp 135-142. doi: 10.1016/j.bandc.2004.08.057.

Lorusso, M. L. and Fumagalli, V. (2020) *La piattaforma Tachidino per la riabilitazione in remoto dei disturbi della lettura e della scrittura [The Tachidino platform for remote rehabilitation of reading and spelling disorders]*,

Teleriabilitazione Nei Disturbi Di Apprendimento. Edited by Erickson. Trento, Italy.

Piazzalunga, C. *et al.* (2023) 'Investigating Visual Perception Impairments through Serious Games and Eye Tracking to Anticipate Handwriting Difficulties', Sensors, Vol 23, No.4, 1765. doi: 10.3390/s23041765.

Piazzalunga, C., Donati, A. and Ferrante, S. (2023) 'Co-design of a points-based reward system to boost motivation in children and improve adherence to learning serious games', in *GNB Atti*.

Raven, J., Raven, J. C. and Court, J. H. (1998) 'Section 2 (Coloured Progressive Matrices)', in Oxford Psychologist Press (ed.) *Raven Manual*. Oxford, UK.

Sage, K., Krebs, B. and Grove, R. (2019) 'Flip, Slide, or Swipe? Learning Outcomes from Paper, Computer, and Tablet Flashcards', *Technology, Knowledge and Learning*, Vol 24, No.3, pp 461–482. doi: 10.1007/s10758-017-9345-9.

Snowling, M. J., Hulme, C. and Nation, K. (2020) 'Defining and understanding dyslexia: past, present and future', *Oxford Review of Education*, Vol 46, No. 4, pp 501-513. doi: 10.1080/03054985.2020.1765756.

Wagner, R. K. *et al.* (2020) 'The Prevalence of Dyslexia: A New Approach to Its Estimation', *Journal of Learning Disabilities*, Vol 53, No. 5, pp 354–365. doi: 10.1177/0022219420920377.

Wechsler, D. (2003) *WISC-IV. Wechsler Intelligence Scale for Children – Fourth UK Edition. Administration and Scoring Manual*. Edited by Harcourt Assessment. London, UK.