Dot-to-Dot: Pre-reading Assessment of Literacy Risk via a Visual-Motor Mechanism on Touchscreen Devices

Wonjeong Park[1], Paulo Revés[2], Augusto Esteves[2,3](✉), Jon M. Kerridge[4], Dongsun Yim[1], and Uran Oh[1]

[1] Ewha Womans University, Seoul, Republic of Korea
{iamwj,sunyim,uran.oh}@ewha.ac.kr
[2] Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
{paulo.reves,augusto.esteves}@tecnico.ulisboa.pt
[3] ITI/LARSyS, Lisbon, Portugal
[4] Edinburgh Napier University, Edinburgh, UK
j.kerridge@napier.ac.uk

Abstract. While early identification of children with dyslexia is crucial, it is difficult to assess literacy risks of these children early on before they learn to read. In this study, we expand early work on Dot-to-Dot (DtD), a non-linguistic visual-motor mechanism aimed at facilitating the detection of the potential reading difficulties of children at pre-reading age. To investigate the effectiveness of this approach on touchscreen devices, we conducted a user study with 33 children and examined their task performance logs as well as language test results. Our findings suggest that there is a significant correlation among DtD task and series of language tests. We conclude the work by suggesting different ways in which DtD could be seamlessly embedded into everyday mobile use cases.

Keywords: Developmental dyslexia · Pre-reading assessment · Literacy risk · Visual-motor tracing

1 Introduction

The ability to read fluently is considered as a prerequisite for children to be successful in academic settings [59]. Thus, children with reading difficulties have challenging times following class materials at school. It is also reported that children who have trouble reading are more likely to have emotional and behavioral problems compared to the children who do not have reading disorders, and these problems may continue after adolescence [11]. Fortunately, previous studies on children with reading difficulties report that early screening and intervention can prevent from resulting in significant problems in school-age learning [12, 14, 33, 52]. Therefore, it is especially important to screen the children with reading difficulties and to conduct appropriate interventions in educational setting as early as possible.

Published by Springer Nature Switzerland AG 2021
C. Ardito et al. (Eds.): INTERACT 2021, LNCS 12932, pp. 192–212, 2021.
https://doi.org/10.1007/978-3-030-85623-6_13

To identify children who have trouble reading, language-based tests such as TOWL (Test of Written Language) [19], DIBELS (Dynamic Indicators of Basic Literacy Skills) [51], and CTOPP (Comprehensive Test of Phonological Processing) [57] have been used by professionals. In these tests, children's reading abilities such as phonological awareness, decoding skills, reading fluency, and rapid naming are evaluated through certain methods. However, since most children with reading difficulties are screened during periods after the school year, following the implementation of literacy education, it is a major issue to carry out accurate screening for children at an earlier age before they learn to read for effective interventions.

For this reason, numerous previous studies have been conducted to designing tasks for efficient screening of the children with reading difficulties in their early childhood [7,15,26]. Recently, Phonological Awareness (PA) test and Rapid Automatized Naming (RAN) have been used as representative tasks to identify children with reading difficulties before they learn to read. PA test measures an individual's meta-linguistic abilities to phonetic structures of oral sound to estimate the child's reading development [28]. RAN, on the other hand, is a task consisted with the series of stimuli such as colors, numbers or familiar objects where a participant is required to speak the stimuli as quickly as possible to predicting reading fluency [36]. These two tests have been reported to be strongly linked to current literacy skills of children in early literacy stage, as well as to future reading and academic achievements, and thus commonly used in clinical settings [24]. Yet, these methods are time-consuming and expensive as these require language specialists. Also, it is difficult to get tested unless referred to a clinic by a parent or a teacher.

Meanwhile, there are screening tools which can save time and costs for requiring specialists. For instance, most of these attempted to identify children with reading difficulties by assessing their game performances on smart devices [3,29,44]. However, as language-dependent tools, they cannot be used for children who speak different languages other than the supported ones. Even if the tool supports the very language that the child speaks, a proper evaluation of the child's underlying abilities is not possible if she/he has not learned to read yet.

Inspired by studies on literacy and dyslexia suggesting that reading difficulty is not only a language surface problem but also related to visual, sensory and motor skills [9,41,54,58], we further investigated a non-linguistic approach described as "Dot-to-Dot (DtD)" [4,37] which originally is a desktop application designed to help identifying at-risk behaviors related to reading difficulties for pre-schoolers. This utilizes a simple tracing mechanism, connecting a series of dots, that can potentially facilitate the passive and implicit screening of children at pre-reading age. Our contributions are threefold:

- We expand on the original DtD task which supported indirect, stylus-based input to a separate display with the goal of significantly reducing the manual scoring time and potential bias of experts via an automated screening process (see Fig. 1). We explore instead direct, touch-based input using a single mobile device (see Fig. 3). Our motivation in the context of human-computer

interaction (HCI) is the possibility of embedding DtD into everyday touchscreen applications such as games or UI manipulations (*e.g.*, dragging) to enable not only automated but also implicit screening.
- We explore our approach via a user study with 33 pre-school and school-aged Korean-speaking children, widening the participant pool of the original work which reported on the performance of English-speaking children. The empirical data collected from the study revealed that the completion time for DtD task can be used to assess the potential risk of having reading difficulties given the children's age regardless of the input conditions. Direct input condition, in particular, can be used to distinguish between children with low and high literacy group.
- We analyze participant tracing performance using a wider but systematic set of trace characteristics including amplitude, angle, or direction. Our goal is to enable HCI designers and developers to embed DtD into their own applications and systems, e.g., the pervasive SwiftKey keyboard. This with the goal of enabling the implicit assessment of not only children of pre-reading age, but teenagers and adults during they everyday interactions with touchscreen devices.

2 Related Work

Our work was inspired by prior work on dyslexia particularly for children and digital assessment tools for cognitive abilities.

2.1 Possible Causes for Reading Difficulties

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines reading difficulties such as developmental dyslexia within the category of learning difficulties. According to this criteria, reading difficulties refer to children who have deficits in decoding and reading comprehension compared to their peers. These children show significant difficulties in learning academic skills related to reading accuracy, fluency, and recognition [2]; as is the definition we use in our work when talking about developmental dyslexia. Previous studies have shown that children with reading difficulties have lower phonological processing abilities than peers without reading difficulties [6,8,10]. However, it is argued that the reading difficulties cannot be explained only by phonological processing skills. Previous studies have revealed that children with reading difficulties also show a broader range of deficits such as sensory, cognitive, and visual-spatial problems. In particular, many studies have linked the poor visual processing skills of children with reading difficulties to their literacy problems [13,16,53]. These problems of visual processing in reading difficulties also affect the sequential processing of specific stimuli. In addition, children with reading difficulties showed lower performance in tasks using motor skills such as motor coordination [60] and motor sequential learning [56] due to their visual processing deficits. We investigated DtD for assessing literacy of children which involves hand-and-eye coordination on a touchscreen motivated by these prior findings that poor visual-motor skills are closely related to the reading problem.

2.2 Early Identification of Dyslexia

Although the cause of dyslexia is still under debate, there is a widespread consensus that early identification and intervention is crucial in both language remediation and in limiting the low self-esteem and behavioral difficulties often reported in unrecognized dyslexia [48,52]. However, identifying dyslexia is often difficult, time consuming, and expensive as it requires assessments by clinical professionals. As such, its formal identification in many schools typically occurs long after children have failed to learn to read, and interventions are provided only after [49]. Identifications of dyslexia may be further delayed when children do not use their native language at school. Several screening tools have been developed in recent years to estimate the risk of reading difficulties at an early stage in primary-aged children (*e.g.*, [34,50]) for offering early intervention. While effective, drawbacks exists since these often require specialists to administer the procedure and interpret results. Also, these tools lack engagements. While a number of game-based tools have developed both in academia [3,17,18,29,42–44] as well as commercial markets (*e.g.*, *Lexercise Screener*[1], *Nessy*[2]), children who have not acquired reading and writing skills (or who speak different languages) cannot use these tools as they are language-dependent.

2.3 Dyslexia Screening Tools for Pre-schoolers

Others have proposed various approaches on early identification of dyslexia for pre-school-aged children who have not learned to read yet. For example, *DIESEL-X* [18] is a tablet game designed to predict the likelihood of having dyslexia for pre-schoolers by measuring dyslexia-related indicators such as letter knowledge and end-phoneme recognition while a child is playing the game although the effectiveness of their prediction model is unknown. *MusVis* [43] and *DGames* [42] utilize musical-visual cues instead of relying on existing knowledge of literacy or phonological awareness, which pre-school children may not have. While promising, limitations still exist because they can only be applied to children who use specific languages, and they are difficult to use for young children as these approaches screen reading difficulties in a way that assesses children's phonological knowledge involving letters. On the other hand, Bannach-Brown proposed Dot-to-Dot (DtD) [4], a stylus-based dot-connecting task on a computer as a screening tool for young children with potential reading difficulties once they learn to read. Her approach is different from *DIESEL-X* or *DGames* as it is entirely non-linguistic; no prior phonological knowledge is required and thus language-independent. In addition, she showed that DtD could successfully differentiate between children at 'low' and 'high' risks of developing dyslexia

[1] https://www.nessy.com/uk/product/dyslexia-screening.
[2] https://www.lexercise.com/tests/dyslexia-test.

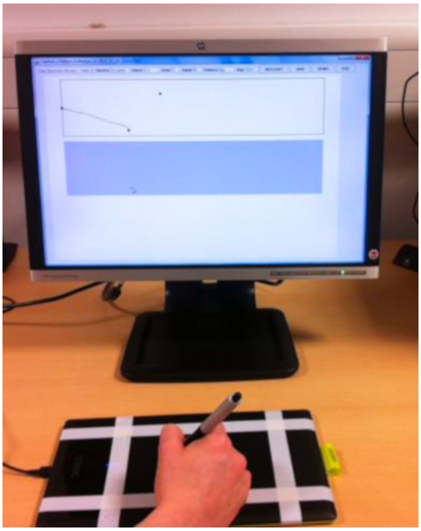

Fig. 1. Photo from the original DtD experiment where participants interacted indirectly with an external display via a stylus [37].

(classified using LUCID Rapid [49]) based on an empirical evidence collected from a user study with 68 English-speaking participants (4 to 8 years old).

We expand this work by exploring the effectiveness of DtD for assessing the potential literacy risks of pre-schoolers via a user study with Korean-speaking children. Moreover, we implemented the task for touchscreen devices to investigate the effects of input modes (direct *vs.* indirect) on detecting potential reading difficulties since direct hand-eye coordination involving one's visual-motor skills of standard touchscreen input is known to outperform indirect input performance [21,45] – as employed in the original DtD work [4,37].

2.4 Digital Tools for Cognitive Assessments

A number of digital tools for assessing one's cognitive abilities have been proposed [5,25,38–40] as an alternative to traditional pen and paper based approach where a trained expert (clinician) is required to examine the test results manually. For instance, Prange *et al.* [39] proposed a multimodal speech-based dialogue system for a questionnaire-based cognitive test called the Mini-Mental State Examination (MMSE). The system could automatically run and evaluate the test with high usability. On the other hand, drawing-based assessments were investigated. Prange *et al.* also implemented a system for detecting signs of dementia or monitoring the stroke recovery progress using the clock drawing test [40]. They showed that their system can reduce the scoring time automating the process and that the its results are clinically reliable. Similarly, Kuosmanen *et al.* [25] developed a smartphone-based clinical tool to detect symptoms of Parkinson's disease with the spiral drawing test. Researchers also focused on Trail

Making Test (TMT) for assessing one's cognitive performance. Barz *et al.* [5], for instance, used TMT performance to predicted the task difficulties in terms of cognitive load. They collected TMT data with six drawing patterns from the Snijders-Oomen Non-verbal intelligence test (SON) with children from elementary school and showed high prediction accuracy. Moreover, a recent study on TMT [38] proposed an automated system that monitors various pen features on a tablet and provides a structured analysis report of the test with explanations for clinicians.

Inspired by the approach for digitizing the assessment process of cognitive performance, we designed an application that utilizes DtD test, which is similar to TMT test but simpler for children, to assess the likelihood of having a dyslexia using off-the-shelf device (*e.g.*, a tablet, a smartphone) so that any sign of reading difficulties can be detected without having to visit a trained expert.

3 The Design of Dot-to-Dot Task

Motivated by prior works on designing digital tools for assessing one's cognitive performance from pen-based tests such as spiral drawings [25] or trail making tests [5,38], our work expands on "Dot-to-Dot" (DtD), a simple dot-connecting task that involves visual-spatial, attentional, and motor mechanisms that often occur in dyslexia [20,22,35]. DtD was initially developed following a series of observations on children tracing their names that highlighted potential visual and motor problems [37]. It was argued by the authors that the difficulty to fixate one's eyes on the screen while moving the stylus on an adjacent trackpad distinguished this task from standard drawing where the eyes gaze just ahead of the hand holding the pen. This ability to dissociate the gaze and the hand may be related to divided attention, and was argued to be likely to reveal developmental delay in control of sensorimotor processing (which may be compromised in individuals with dyslexia).

While the original work by Bannach-Brown [4] and Piotrowska [37] focused on DtD as proof-of-concept for an automated screening process using indirect, pen-based input; we focus instead on the use of the index finger of the dominant hand – a more prevalent form of input with modern touchscreen devices. By exploring how different touch-based strokes relate to at-risk behaviors, we aim to provide interaction designers working on mobile platforms with the knowledge required to pick-up on such at-risk behaviors via already enabled user actions in their mobile apps (or re-think new interaction techniques altogether around these strokes). Regardless of how successful the original DtD task is, a single-purpose application (i.e., a DtD app) will always have a limited reach. Our goal instead is to facilitate the embedding of an implicit DtD screening process into everyday apps and games, which would greatly expand the reach and impact of this approach.

With this in mind, the DtD task we designed for our study requires the user to connect a series of targets (dots) that consecutively appear on a tablet screen as quickly and accurately as possible (without lifting their fingers from the display). At the start of a new task two targets are displayed: the first target in the sequence in red, and the next target in green. Once the user successfully drags his/her finger from the red to the green targets, the next target in the sequence is displayed in green and the previous targets disappear. The user continues to drag its finger towards the latest green target until the sequence is complete. If, for whatever reason, the user lifts the finger from the display during the task, a red target is displayed at the location where this took place. Users can resume the task by moving their fingers from this red target to the green target they were previously pursuing. Five sequences (or patterns) of growing difficulty were designed using eight sequential targets, or seven unique traces, varying index of difficulty and angle to the next target with fixed target width of 0.6 cm:

- **Index of difficulty (approx. amplitude):** 1.5 (1.10 cm), 2 (1.80 cm), 2.5 (2.79 cm), 3 (4.20 cm), 3.5 (6.19 cm)
- **Angle to the next target:** 0, 30, 60, 90, 270, 300, 330 degrees

Index of difficulty (ID) was calculated using Shannon's formulation [30,47]): $ID = \log_2(\frac{A}{W} + 1)$, where A is the amplitude of movement, and W the target width. The five patterns designed can be seen in Fig. 2, accounting for 35 unique traces.

4 User Study

To evaluate the effectiveness of the DtD task in touchscreens, we conducted a user study with 33 children where they were asked to perform a series of dot-connecting tasks. The study was approved by the institutional review board (IRB) of the university where the researchers are affiliated with.

4.1 Participants

Thirty-eight preschool and school-aged children (18 males, 20 females) living in South Korea were initially recruited for this study. We excluded five children from the analysis because two scored less than 85 standard scores on the non-verbal intelligence test, and three had difficulty in handling the tablet PC and thus were severe outliers in the data. As a result, we report on the data from 33 participants in analysis below (13 males, 20 females, mean age of 89.33 months, $SD = 24.29$). Participants were recruited through online and offline advertising on bulletin boards at a private child development center and daycare center; and all children met the following criteria: (1) chronological age was between 5 to 12; (2) the standard score of non-verbal IQ test [31] was higher than 85 ($-1SD$); (3) they speak only Korean; and (4) did not show sensory impairments or psychological problems, as reported by parent and the nursery teacher. Note that we recruited both pre-school (aged 5) and school-aged children (aged 6 to 12) to capture a wide set of data on children with different reading skills and at different points of their development.

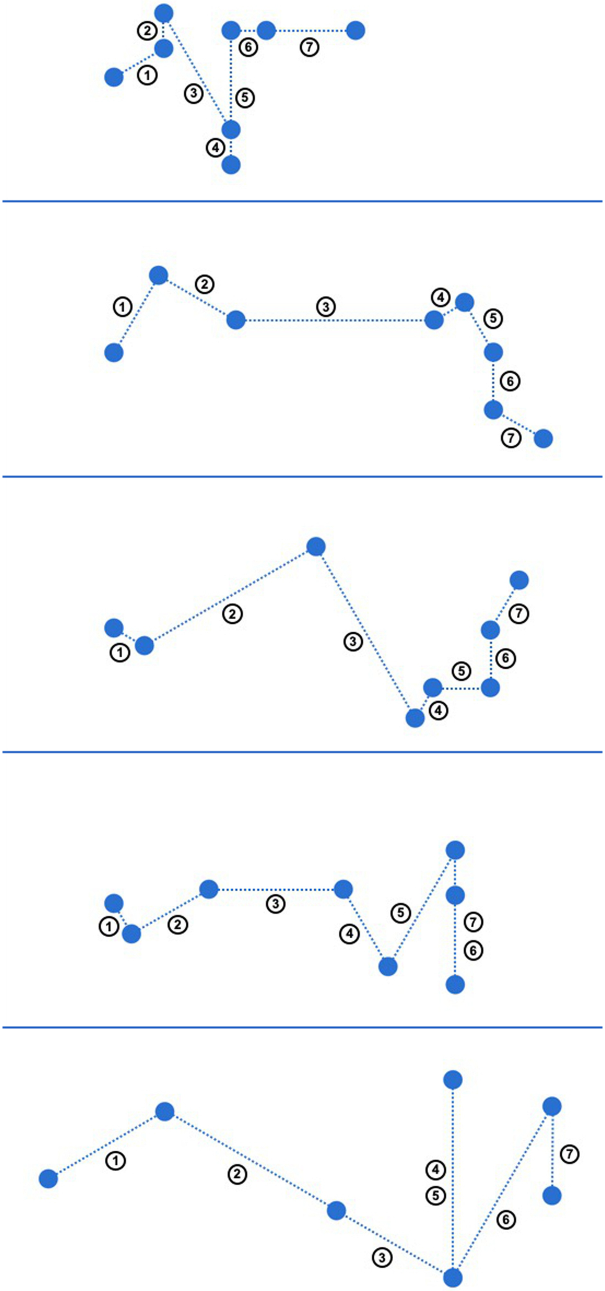

Fig. 2. Patterns 1 to 5 from top to bottom, representing 35 unique traces (5 IDs × 7 angles). These are incrementally harder, represented by traces with average amplitudes of: 2.33 cm, 2.61 cm, 2.85 cm, 3.15 cm, and 5.14 cm. Numbers indicate the tracing order.

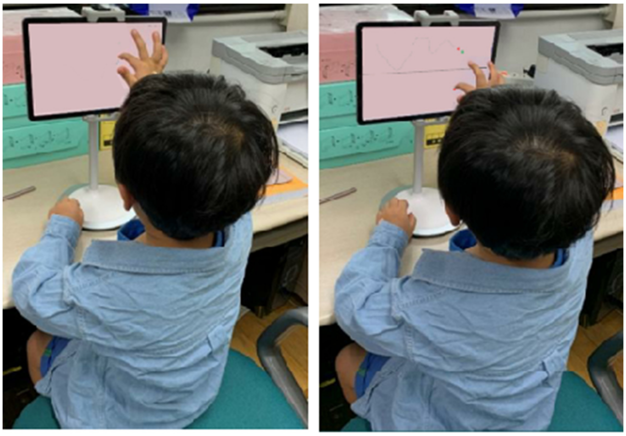

Fig. 3. Our experimental setting for direct (left) and indirect (right) conditions; the latter emulating the indirect approach of the original study, now via touch input.

4.2 Apparatus

DtD was implemented using the Processing programming language. To make the task more appealing, one out of three recordings of children cheering (<1 s) would play every time a green target was reached. User performance was captured locally in the form of x- and y-coordinates representing finger positions at approximately 60 Hz.

4.3 Input Conditions

In addition to direct manipulation mode where participants could connect the dots by directly touching the dots that appear on the entire screen as described in Sect. 3 (**direct mode**), we investigated indirect input mode where the participants were only able to use the bottom half of the screen to manipulate a cursor on the top half of the screen; an additional red target were displayed at the bottom half of the screen to indirectly manipulate the cursor on the top (**indirect mode**); see Figs. 4 and 3. We investigated these two input modes to examine if and how DtD performance vary depending on which input mode is used assuming that different levels of visual-motor skills would be required depending on the input mode for performing the DtD task [37].

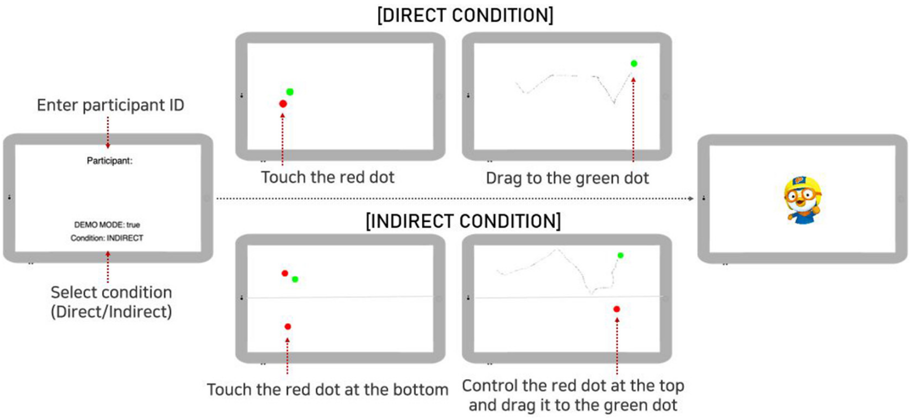

Fig. 4. Procedures of direct (top) and indirect (bottom) conditions. Starting dot was colored as red and the target dot was colored green. (Color figure online)

4.4 Procedure

In accord with IRB regulations, researchers explained the procedure and benefits of the study and received consent from the children and their guardians. All children participated in a language development test, literacy-related tasks, and DtD test. Additionally, we conducted the Korean Receptive and Expressive Vocabulary Test [23] to assess children's language development through evaluation of vocabulary skills. REVT is divided into two tests: the receptive vocabulary test (REVT-R) and the expressive vocabulary test (REVT-E); in which we presented a series of pictures corresponding to different vocabulary (ranked by difficulty level), and evaluated whether the child could vocalize and understand each picture. The total raw scores were kept for further statistical analysis.

The literacy-related tasks included the Rapid Automated Naming (RAN) and Phonological Awareness (PA) test. The RAN test [46] presents a 6×6 matrix on a computer monitor and randomly placed red, yellow, blue, and green colors inside each cell. Children were asked to speak the colors in the cells as quickly as possible, and the time required for the children to speak all the name of colors was measured and used for analysis. The PA is a test that evaluates children's abilities to recognize and modify sound during speech, and consists of two subtests: phonological segmenting and phonological blending. These subtests were conducted at phonemic and syllabic levels, and the total score obtained from this process was recorded. Both RAN and PA have been shown to be strong predictors of a child's literacy skills [15, 55].

In our DtD task, participants were asked to trace various sequence of dots on a tablet PC (LG-X760, 10.1") as accurately as possible with their index fingers (without lifting their fingers from the display). The dots in each pattern and condition were presented from left to right as shown in Fig. 2. Motivated by our goal of embedding DtD into everyday mobile devices and by the findings in [3], where children were reported to prefer direct touch (i.e., finger) over indirect touch (i.e., mouse, stylus) due to the intuitiveness of the physical interaction, we examined two input modes. In the direct input condition children were task to place their fingers on the red dot and move this to the green dots that are presented sequentially. In the indirect input condition this red dot was presented with a vertical offset to the sequential green dots, and children had to perform this tracing task indirectly. The particulars of each experimental condition are presented in Fig. 4.

We used a tablet stand to minimize any occlusion that can occur from children's hands covering the next green dot to follow. The tablet was placed at the height of participants heads, at a distance of approximately 30 cm (see Fig. 3). All parents received a report from specialists (licensed speech-language pathologists in South Korea) on the child's language and cognitive abilities after the study as a benefit to participation.

4.5 Data and Analysis

The DtD task performance of all children was analyzed using task completion time and the frequencies of finger lifts per condition and pattern. Each participant's finger lift pattern and trace was represented using x-, y-coordinates. Based on this data, we conducted a series of statistical tests to: (1) examine whether the children's DtD performance correlated with the language development level literacy-related skills; (2) identify whether there is a significant correlation between specific DtD patterns with language and literacy task results; and (3) determine whether there is a significant group difference on DtD performance between high and low literacy groups. For the study purpose, Pearson's correlation and Mann-Whitney U test were conducted using SPSS (v. 26.0).

4.6 Findings

In our study, we found that DtD performance was highly associated with children's reading related variables where the degrees of association between children's age, literacy skills, and the input condition and patterns of the task varied.

Overall Performance in DtD and Literacy-Related Tasks. Prior to the correlation analysis, we examined descriptive statistics of children's literacy-related tasks and DtD performance (see Table 1). Overall, language ability and literacy-related variables, as well as the performance of DtD tend to improve with age. Paired t-tests showed that the time required for performing the direct condition was significantly shorter ($t = -5.481$, $p < .001$), and that the number of finger lifts was less frequent in the direct condition than in the indirect condition ($t = -3.099$, $p = .004$). These indicate that the overall difficulty of DtD was lower in the direct condition as expected [21,45].

Correlation Analysis Across Age Groups. To identify correlations between DtD performance and children's language development and literacy skills, we computed Pearson's correlation coefficients between these. The results

Table 1. Descriptive statistics in age subgroups where DtD-D and DtD-I refer to direct and indirect conditions respectively. Overall, participants who have relatively higher literacy skills took a shorter time to perform the DtD task with fewer finger lifts for both direct and indirect conditions.

	Preschool-age (n=16)	Lower grade (n=10)	Upper grade (n=7)
Age (month)	71.06 (5.99)	90.90 (10.94)	128.86 (12.66)
REVT-R	64.13 (10.65)	84.50 (20.00)	136.00 (23.97)
REVT-E	68.19 (15.33)	90.10 (19.50)	144.43 (23.49)
RAN (sec)	42.31 (8.87)	39.80 (25.29)	23.43 (3.64)
PA	35.00 (17.81)	73.00 (26.83)	94.29 (6.86)
DtD-D Time (ms)	60151.31 (19902.08)	51084.60 (12911.64)	35072.57 (9866.21)
DtD-I Time (ms)	91266.56 (25610.86)	78160.20 (35102.15)	45597.43 (5967.74)
DtD-D Finger Lift	9.50 (6.70)	4.80 (5.92)	0.14 (0.38)
DtD-I Finger Lift	16.69 (11.74)	11.40 (10.08)	1.14 (2.27)

Values are presented as mean(SD)

showed that receptive vocabulary scores (REVT-R), expressive vocabulary scores (REVT-E), and PA showed significant negative correlation with the task duration and the number of finger lifts both in direct and indirect conditions ($p <$ .050). The only exception was the correlation between RAN and the number of finger lifts in both conditions ($p >$ .050). Additionally, these results also show that children's age has a significant effect on overall performance on DtD task as in the literacy-related tests. This suggests DtD has the potential for estimating one's literacy risk when compared to his/her peers in the same age group.

High Literacy *vs.* Low Literacy Group. We further examined the performance in terms of the literacy skills of participants to explore the potential for our DtD mechanism in detecting children with reading difficulties. For this analysis we rely on the RAN and PA subtests for evaluating children's literacy skills, and to group them based on this performance. Thus, the children who performed below 0.5 standard deviations (*SDs*) from the mean in both RAN and PA tests we considered part of the low literacy group, while children performing above 0.5 *SDs* where considered part of the high literacy group.

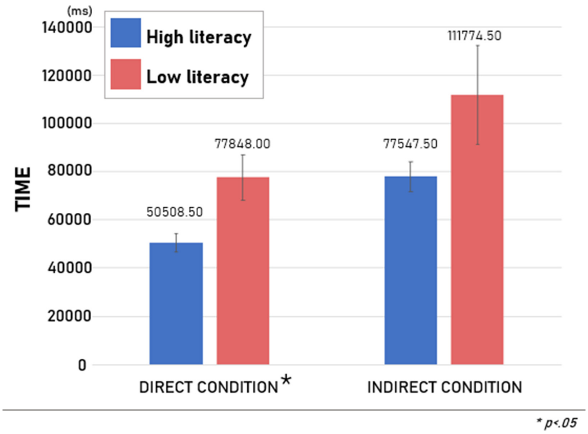

Fig. 5. The average time for both direct and indirect conditions between high literacy group and low literacy group with Mann-Whitney U test results.

Following this criteria, eight similarly aged children were selected: four from each literacy group. As shown in Fig. 5, Mann-Whitney U tests showed that there was a significant difference between the two groups in terms of time only for the direct condition ($U = .000$, $p = .021$); while the number of finger lifts was not found to be significant for either of the two conditions. Still, the number of finger lifts during direct input could also potentially be used to distinguish the two groups with a larger sample size ($U = 1.500$, $p = .056$). This result implies that DtD performance during direct input has potential to identify children with low literacy skills (see Fig. 6 for trace examples produced by children in high and low literacy groups).

Correlation Analysis by Patterns. To examine if the index of difficulty (ID) of each pattern (i.e., trace amplitude) can be a relevant predictor of literacy risk, we conducted a secondary analysis focusing on the direct condition results which were significantly correlated with RAN – one of the representative literacy tests we used for the main analysis above (age-controlled). The results of the correlation analysis suggest *Pattern 5*, with the highest ID, was positively correlated with RAN results ($r = .381$, $p = .031$). Children's pattern-specific performance in the direct input condition and the results of correlation analysis between patterns and RAN performance are presented in Fig. 7.

Analysis by Trace. Continuing our top-down approach, our last analysis focuses on developing a deeper understanding of which individual trace properties can be helpful in predicting literary risk. In order to achieve this, we consider four performance metrics (see Fig. 8): (1) the time to complete a trace in *ms* (*TT*), (2) the distance between the farthest point of a trace to the best fit in *px* (*MD*; the shortest path between two successive points in a pattern), (3) the average distance between each point in the trace to the best fit in *px* (*AD*),

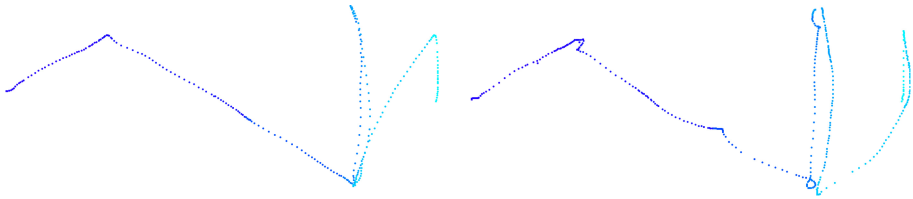

Fig. 6. Example traces from two children during *Pattern 5* (P5): high literacy group on the left and low literacy on the right. Noticeable differences include direction corrections at almost every target for the child with low literacy skills. Further, the pattern on the left was completed in a little over 7 *s* with no finger lifts; while the pattern on the right was completed in over 25 *s* and included 6 finger lifts. The input direction is encoded as a blue gradient towards a lighter shade of blue from left to right. (Color figure online)

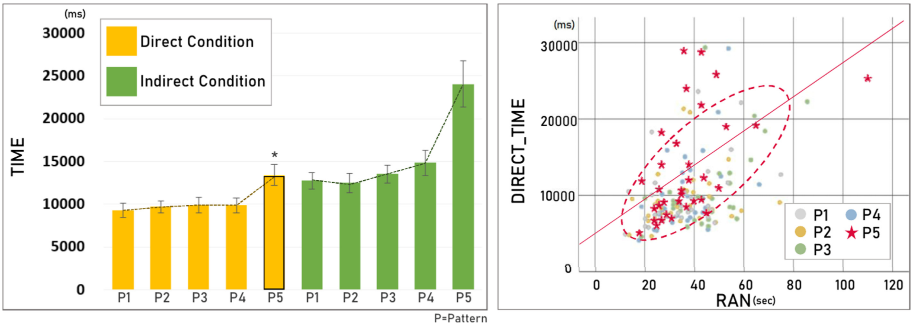

Fig. 7. Left: pattern-specific average time performance per pattern and per input condition. Right: the correlation between time performance and RAN for each of the five patterns. The order of the presentation was the same starting with the first pattern (P1) to the last (P5), whose index of difficulty (ID) is the highest.

and (4) the time to correct the tracing direction at the start of a new trace in *ms* (*TD*). While MD can be used to detect mistakes such as overshooting or sudden changes in direction, AD reflects how well a participant were able to stay close to the shortest path when connecting two dots during the task. TD is the time it takes a participant to start a new trace in the direction of the latest point; this is measured from the time a new point is displayed until the direction of the participant's trace is within 30° s of the target point to quantify visual-motor challenges associated with each new trace if any.

As such, our analysis focused on 48 items: 4 performance metrics × (5 IDs + 7 angles). Following the earlier analysis, we examined solely at the direct condition and RAN performances. We began by filtering these features via a Pearson's correlation matrix between the 48 items and the RAN results, ignoring all items with a correlation coefficient below 0.50. This resulted in six features: 30°-MD (0.50), 90°-TT (0.59), 1.5-TT (0.61), 3.5-TT (0.51), 3.5-MD (0.55), and 3.5-TD (0.58) – the latter three features were not surprising as they played a prominent

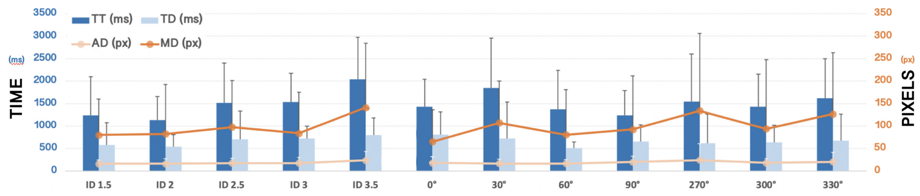

Fig. 8. Participant performance by trace (IDs and angles), measured as: time to complete a trace in ms (TT); maximum distance to best fit in px (MD); average distance to best fit in px (AD); and time to correct the tracing direction at the start of a new trace in ms (TD).

role in *Pattern 5*. We then computed another correlation matrix between these. When two features were correlated (>0.50), the one highest correlated with RAN was selected. This resulted in three final features: 30°-MD, 90°-TT, and 1.5-TT (see Fig. 9). That is, the farthest distance to the best fit between two points at 30° s; the time it takes to drawn an upwards trace (90°); and the time it takes to draw a trace with ID 1.5 (approx. 1.10 cm).

Finally, we computed a Decision Tree Regression using NumPy and Scikit-learn (CART algorithm, default parameters). This reported a mean absolute error (MAE) of 5.71 and root mean squared error ($RMSE$) of 8.50. The importance of each feature was reported as follows: 1.5-TT (0.84), 90°-TT (0.12), and 30°-MD (0.04). With mean RAN results of 37.55 ms (SD = 16.61), a MAE of 5.71 ms represents an error 15.21% and is below 0.5 standard deviations representing the RAN results of our low literacy group (i.e., 8.31 ms). In sum, these results start to illustrate the potential of DtD as an implicit tracing mechanism that can assess low literacy risk factors by simply measuring characteristic slow and short traces, particularly if these are performed upwards.

5 Discussion

The purpose of this work was to determine whether DtD on touchscreen devices is a suitable mechanism to assess literary-related factors.

5.1 Relevance to Literacy-Related Factors

Overall, children's performance during the direct condition of DtD was highly correlated with literacy-related variables. These results showed a consistent tendency regardless of age effect, especially with rapid automatized naming (RAN) performance. In other words, children who performed RAN quicker were also were quicker at connecting the dots in the direct condition of the DtD task. As such, it is expected that the properties represented by RAN – children's vocabulary development level, rapid naming, and phonological awareness skills – can also be represented via the DtD task. Unlike vocabulary development tests that recall and produce the lexical knowledge stored in long-term memory, and phonological recognition tasks that manipulate speech sounds

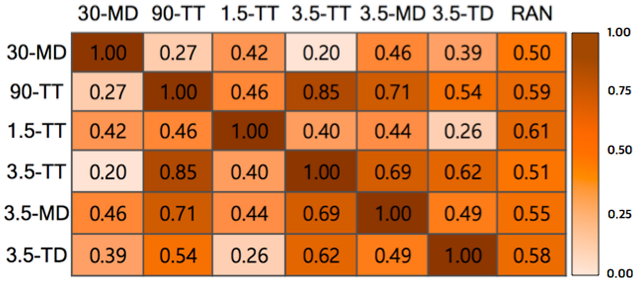

Fig. 9. The Pearson's correlation matrix between the most relevant features. When two features are correlated (> 0.50), the one highest correlated with RAN was selected for analysis.

based on phonemic knowledge and representation, the RAN task requires effective processing skills of visual information. Similarly, in the DtD task users need to be constantly paying attention to moving target points and processing visual information. Common features in performing these tasks can explain this correlation between RAN and DtD performance.

While we found a significant correlation between DtD and RAN, the same cannot be said of phonological awareness (PA) performance. This result is supported by previous work that demonstrates the literacy predictability of PA decreases from the early states of reading education (while the predictability of RAN gradually increases [27,61]). Most of the participants in this study had indeed started to learn how to read, which further explains these observations on RAN and PA performance. In sum, the performance in the direct input condition of the DtD task can be considered as an assessment tool that reflects children's literacy ability; including visual information processing, linguistic features, and the developmental characteristics necessary for reading.

5.2 Feasibility as a Pre-reading Assessment Mechanism

In this paper we have used RAN and PA performances to categorize children's literacy skills as either high or low (above or below 30% from the average performance of their peers). A comparative analysis of the two subgroups showed that children with low literacy had low performance in the direct input condition of the DtD task. Thus, this can be seen as a result of predicting a high probability of diagnosis for children with literacy-related issues such as reading difficulties and dyslexia. Our findings build on the preliminary DtD work showing similar results with British children using indirect input (i.e., stylus pen) [37]. Taken together, this highlights the language-independent features of the DtD task as a potential literacy screening tool.

However, what further sets this work apart from the state-of-the-art is the analysis that shows the potential for the DtD mechanism to work as an abstract layer that can be embedded into every day touch-screen applications. This would ultimately lead to continuous and unsupervised monitoring of literacy problems

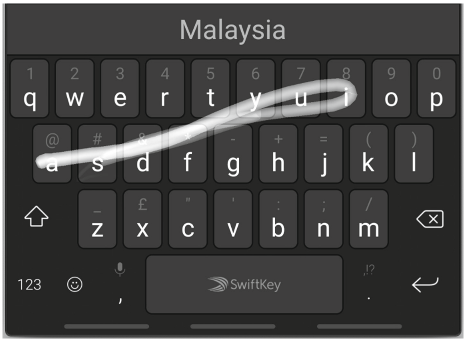

Fig. 10. Input via the popular keyboard SwiftKey could be interpreted as a DtD pattern. Future work should investigate if such everyday actions on mobile devices could be used as a quick and seamlessly way to flag for signs of dyslexia in teenagers and adults.

in preschool-age children – shown to interact with tablets as early as 2.5 years old [32]. To do so, applications simply need to measure user performance matching the three features identified in our study. These are highly viable in touch-screen applications as they require short traces of approx. 1.10 cm, combined with upward traces.

5.3 Limitations and Future Work

Although our data showed the potential of using the DtD task as a screening tool for detecting reading difficulties in children, a long-term investigation is needed to confirm if DtD is indeed an accurate predictor. This will be done by following up with children with low DtD performance and assess if they are diagnosed with reading difficulties or dyslexia later on. Additionally, a future study on reliability and feasibility tests for larger clinical groups will also be needed to find out whether the patterns and methods used in DtD itself can be used as a tool to screen children with reading difficulties. For example, a larger clinical group would allow us to do a trace analysis by preschool-age, or by lower and higher grades; allowing us to further validate and fine-tune our findings.

Finally, an exciting future direction for this work is to look at the validity of the DtD task while embedded into various touch-screen applications. This can be explored in children's applications, but could also be used to flag for signs of dyslexia in teenagers and adults. One such way would be to identify applications that already require users to provide direct input in manner resembling a DtD pattern, i.e., applications that require users to sequentially swipe between interface targets without lifting their fingers. One pervasive example would be SwiftKey[3], a popular software keyboard for Android where users write text not by typing, but by swiping between characters. Figure 10 illustrates how writing *Malaysia* with SwiftKey would produce an input with a similar ID to Pattern 5

[3] https://www.microsoft.com/en-us/swiftkey.

(depending on screen size and resolution). A machine learning model could also be trained to identify the trace features described earlier (e.g., 1.5-TT, 90°-TT, 30°-MD), potentially allowing the assessment of low literacy factors when users perform slow and short upward traces. These are used often, not only between characters in SwiftKey (e.g., between 'x' and 'e'), but unlocking an iPhone, dragging App icons on them main screen of most touchscreen devices, or various videogames. Further studies would require us to test IDs that vary not only length but target size, backwards traces, and traces that might engage muscle memory in addition to visual-motor coordination (such as the ones performed in SwiftKey). If successful, such approaches could quickly and seamlessly support millions of undiagnosed adults [1] that go through life dealing with academic failure, low self-esteem, and behavioral and motivational difficulties.

6 Conclusion

We have proposed a simple, non-linguistic touchscreen-based task with the goal of predicting a child's likelihood for experiencing reading difficulties based on their visual-motor skills. The effectiveness of this approach was verified through an empirical study conducted with children aged from preschool to elementary school. Results show that children's literacy-related variables have strongly correlated to DtD performance. These results were significant not only in the correlation analysis but also in the comparison analysis between low and high literacy groups. Moreover, we identified specific DtD pattern and trace characteristics that could be effective in screening future learning difficulties. Thus, DtD can contribute to not only an early identification and intervention of children's literacy problems but can also be embedded into any frequently used touchscreen-based applications for implicit detection.

Acknowledgements. We thank all the children and educators who have made this research possible. This was supported by LARSyS (Projeto – UIDB/50009/2020).

References

1. Alexander-Passe, N.: The Successful Dyslexic: Identify the Keys to Unlock Your Potential. Springer, Rotterdam (2017)
2. American Psychiatric Association: Diagnostic and Statistical Manual Of Mental Disorders (DSM-5®). American Psychiatric Pub, Arlington (2013)
3. Van den Audenaeren, L., et al.: DYSL-X: design of a tablet game for early risk detection of dyslexia in preschoolers. In: Games for health, pp. 257–266. Springer, Wiesbaden (2013)
4. Bannach-Brown, A.: Visual-motor integration in developmental dyslexia (2014)
5. Barz, M., Altmeyer, K., Malone, S., Lauer, L., Sonntag, D.: Digital pen features predict task difficulty and user performance of cognitive tests. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, pp. 23–32, UMAP 2020. ACM, New York (2020). https://doi.org/10.1145/3340631.3394839

6. Boets, B., De Smedt, B., Cleuren, L., Vandewalle, E., Wouters, J., Ghesquiere, P.: Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. Br. J. Dev. Psychol. **28**(1), 5–31 (2010)
7. Catts, H.W.: The relationship between speech-language impairments and reading disabilities. J. Speech Lang. Hear. Res. **36**(5), 948–958 (1993)
8. Dandache, S., Wouters, J., Ghesquière, P.: Development of reading and phonological skills of children at family risk for dyslexia: a longitudinal analysis from kindergarten to sixth grade. Dyslexia **20**(4), 305–329 (2014)
9. Eden, G.F., VanMeter, J.W., Rumsey, J.M., Maisog, J.M., Woods, R.P., Zeffiro, T.A.: Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. Nature **382**(6586), 66 (1996)
10. Ehri, L.C., Nunes, S.R., Willows, D.M., Schuster, B.V., Yaghoub-Zadeh, Z., Shanahan, T.: Phonemic awareness instruction helps children learn to read: evidence from the national reading panel's meta-analysis. Read. Res. Q. **36**(3), 250–287 (2001)
11. Eissa, M.: Behavioral and emotional problems associated with Dyslexia in adolescence. Curr. Psych. **17**(1), 39–47 (2010)
12. Elbro, C., Petersen, D.K.: Long-term effects of phoneme awareness and letter sound training: an intervention study with children at risk for dyslexia. J. Educ. Psychol. **96**(4), 660 (2004)
13. Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., Mascetti, G.G.: Visual-spatial attention in developmental dyslexia. Cortex **36**(1), 109–123 (2000)
14. Felton, R.H., Pepper, P.P.: Early identification and intervention of phonological deficits in kindergarten and early elementary children at risk for reading disability. School Psychol. Rev. **24**(3), 405–414 (1995)
15. Furnes, B., Samuelsson, S.: Phonological awareness and rapid automatized naming predicting early development in reading and spelling: results from a cross-linguistic longitudinal study. Learn. Individ. Differ. **21**(1), 85–95 (2011)
16. Gabrieli, J.D., Norton, E.S.: Reading abilities: importance of visual-spatial attention. Curr. Biol. **22**(9), R298–R299 (2012)
17. Gaggi, O., Galiazzo, G., Palazzi, C., Facoetti, A., Franceschini, S.: A serious game for predicting the risk of developmental dyslexia in pre-readers children. In: 2012 21st International Conference on Computer Communications and Networks (ICCCN), pp. 1–5. IEEE (2012)
18. Geurts, L., et al.: DIESEL-X: a game-based tool for early risk detection of dyslexia in preschoolers. In: Torbeyns, J., Lehtinen, E., Elen, J. (eds.) Describing and Studying Domain-Specific Serious Games. AGL, pp. 93–114. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20276-1_7
19. Hammill, D.D., Larsen, S.C.: Test of Written Language: TOWL4. PRO-ED, Austin (2009)
20. Hansen, P.C., Stein, J.F., Orde, S.R., Winter, J.L., Talcott, J.B.: Are dyslexics' visual deficits limited to measures of dorsal stream function? NeuroReport **12**(7), 1527 (2001)
21. Hutchins, E.L., Hollan, J.D., Norman, D.A.: Direct manipulation interfaces. Hum. Comput. Interact. **1**(4), 311–338 (1985)
22. Kevan, A., Pammer, K.: Predicting early reading skills from pre-reading measures of dorsal stream functioning. Neuropsychologia **47**(14), 3174–3181 (2009)
23. Kim, Y., Hong, G., Kim, K., Jang, H., Lee, J.: Receptive & Expressive Vocabulary Test (REVT). Seoul Community Rehabilitation Center, Seoul (2009)
24. Kirby, J.R., Parrila, R.K., Pfeiffer, S.L.: Naming speed and phonological awareness as predictors of reading development. J. Educ. Psychol. **95**(3), 453 (2003)

25. Kuosmanen, E., Kan, V., Visuri, A., Hosio, S., Ferreira, D.: Let's draw: detecting and measuring Parkinson's disease on smartphones. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–9, CHI 2020. ACM, New York (2020). https://doi.org/10.1145/3313831.3376864
26. Landerl, K., et al.: Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. Sci. Stud. Read. **23**(3), 220–234 (2019)
27. Landerl, K., Wimmer, H.: Development of word reading fluency and spelling in a consistent orthography: an 8-year follow-up. J. Educ. Psychol. **100**(1), 150 (2008)
28. Lombardino, L.J., Morris, D., Mercado, L., DeFillipo, F., Sarisky, C., Montgomery, A.: The early reading screening instrument: a method for identifying kindergarteners at risk for learning to read. Int. J. Lang. Commun. Disord. **34**(2), 135–150 (1999)
29. Lyytinen, H., Ronimus, M., Alanko, A., Poikkeus, A.M., Taanila, M.: Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. Nordic Psychol. **59**(2), 109–126 (2007)
30. MacKenzie, I.S.: A note on the information-theoretic basis for Fitts' law. J. Motor Behav. **21**(3), 323–330 (1989)
31. Moon, S., Byun, C.: Korean Kaufman assessment battery for children (K-ABC) (2003)
32. Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., Gerhardstein, P.: They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. J. Exper. Child Psychol. **137**, 137–155 (2015)
33. Nation, K., Cocksey, J., Taylor, J.S., Bishop, D.V.: A longitudinal investigation of early reading and language skills in children with poor reading comprehension. J. Child Psychol. Psych. **51**(9), 1031–1039 (2010)
34. Nicolson, R., Fawcett, A.: The Dyslexia Early Screening Test: DEST-2. Psychological Corporation (2004)
35. Nicolson, R.I., Fawcett, A.J.: Dyslexia, dysgraphia, procedural learning and the cerebellum. Cortex J. Devoted Study Nervous Syst. Behav. **47**, 117–127. (2011). https://doi.org/10.1016/j.cortex.2009.08.016
36. Norton, E.S., Wolf, M.: Rapid automatized naming (ran) and reading fluency: Implications for understanding and treatment of reading disabilities. Ann. Rev. Psychol. **63**, 427–452 (2012)
37. Piotrowska, B.: Investigating the performance and underlying mechanisms of a novel screening measure for developmental dyslexia: implications for early identification. Edinburgh Napier University, Ph.D. thesis, June 2018
38. Prange, A., Barz, M., Heimann-Steinert, A., Sonntag, D.: Explainable automatic evaluation of the trail making test for dementia screening. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–9 (2021)
39. Prange, A., Niemann, M., Latendorf, A., Steinert, A., Sonntag, D.: Multimodal speech-based dialogue for the mini-mental state examination. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA 2019, pp. 1–8. ACM, New York (2019). https://doi.org/10.1145/3290607.3299040
40. Prange, A., Sonntag, D.: Modeling cognitive status through automatic scoring of a digital version of the clock drawing test. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2019, pp. 70–77. ACM, New York (2019). https://doi.org/10.1145/3320435.3320452
41. Ramus, F., Pidgeon, E., Frith, U.: The relationship between motor control and phonology in dyslexic children. J. Child Psychol. Psych. **44**(5), 712–722 (2003)

42. Rauschenberger, M., Rello, L., Baeza-Yates, R.: A tablet game to target dyslexia screening in pre-readers. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 306–312 (2018)
43. Rauschenberger, M., Rello, L., Baeza-Yates, R., Bigham, J.P.: Towards language independent detection of dyslexia with a web-based game. In: Proceedings of the Internet of Accessible Things, pp. 1–10 (2018)
44. Rello, L., Ali, A., Bigham, J.P.: Dytective: toward a game to detect dyslexia. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 307–308. ACM (2015)
45. Saffer, D.: Designing Gestural Interfaces: Touchscreens and Interactive Devices. O'Reilly Media, Inc., Reading (2008)
46. Semel, E., Wiig, E., Secord, W.: Clinical Evaluation of Language Fundamentals, (CELF-4). The Psychological Corporation. San Antonio (2003)
47. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Techn. J. **27**(3), 379–423 (1948)
48. Shaywitz, B.A., et al.: Development of left occipitotemporal systems for skilled reading in children after a phonologically- based intervention. Biol. Psychiatry **55**(9), 926–933 (2004)
49. Singleton, C.: Intervention for dyslexia: a review of published evidence on the impact of specialist teaching. University of Hull (2009)
50. Singleton, C., Horne, J., Simmons, F.: Computerised screening for dyslexia in adults. J. Res. Reading **32**(1), 137–152 (2009)
51. Smith, K., Wallin, J.: Dibels Next Assessment Manual. Dynamic Measurement Group, Eugene (2011)
52. Snowling, M.J.: Early identification and interventions for dyslexia: a contemporary view. J. Res. Spec. Educ. Needs **13**(1), 7–14 (2013)
53. Stein, J.: Dyslexia: the role of vision and visual attention. Current Dev. Disorders Rep. **1**(4), 267–280 (2014)
54. Stein, J., Walsh, V.: To see but not to read; the magnocellular theory of dyslexia. Trends Neurosci. **20**(4), 147–152 (1997)
55. Swanson, H.L., Trainin, G., Necoechea, D.M., Hammill, D.D.: Rapid naming, phonological awareness, and reading: a meta-analysis of the correlation evidence. Rev. Educ. Res. **73**(4), 407–440 (2003)
56. Waber, D.P., et al.: Motor sequence learning and reading ability: is poor reading associated with sequencing deficits? J. Exper. Child Psychol. **84**(4), 338–354 (2003)
57. Wagner, R., Torgesen, J., Rashotte, C., Pearson, N.: Comprehensive Test of Phonological Processing-Second Edition (2013)
58. White, S., et al.: The role of sensorimotor impairments in dyslexia: a multiple case study of dyslexic children. Dev. Sci. **9**(3), 237–255 (2006)
59. Whitehurst, G., Lonigan, C.: Child development and emergent literacy. Child Dev. **69**, 848–872 (1998). Focus on Exceptional Children 44(3) (2011)
60. Wolff, P.H., Michel, G.F., Ovrut, M., Drake, C.: Rate and timing precision of motor coordination in developmental Dyslexia. Dev. Psychol. **26**(3), 349 (1990)
61. Ziegler, J.C., et al.: Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. Psychol. Sci. **21**(4), 551–559 (2010)