Objective Assessment Application for Preschool Child Development

Guillaume Odendaal
Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Stellenbosch, South Africa
Guillaume.odendaal@gmail.com

Anna Marie Clay
Agricultural and Biological Engineering, Mississippi State University, MS, USA
Amd406@msstate.edu

David Vandenheever
Agricultural and Biological Engineering, Mississippi State University, MS, USA
Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Stellenbosch, South Africa
davidvdh@abe.msstate.edu

Abstract— Current developmental screening tests are typically subjectively evaluated making them susceptible to bias and are time- and resource-intensive. We present here the development of a tablet application for developmental screening incorporating fine motor and language tests. The tablet application was built with modularity in mind to ease the process of adaptation for cultural and age-appropriate conversions. An accompanying assessment pipeline was constructed to automatically process the data from the tablet assessment into several different metrics. The initial results indicate the usefulness and feasibility of the proposed application.

Keywords—early child development; language assessment; fine motor assessment; computerized assessment

I. Introduction

In order for a child to develop at the correct pace, he/she needs the appropriate stimulation and care; even from birth, they should be continuously stimulated at the appropriate level (Agyei, van der Weel, & van der Meer, 2016). The stimulation typically happens at home or in childcare facilities. For every child, the preschool years are the most important, because brain development and neural plasticity are at their peak (Chugani, 1998). Children that do not receive appropriate stimulation and care are at risk of not developing to their fullest potential. In 2016, the World Health Organization (WHO) estimated that 43% of children in low- and middle-income countries (LMICs) (250 million) were unable to reach their full development potential due to a lack of correct stimulation and care (WHO, 2018). The presence of neurodivergence in some children increases this risk. Neurodivergence is defined as a brain that functions in ways that diverge significantly from the dominant societal standards of “normal” and may include disorders such as Autistic Spectrum Disorder (ASD), Attention Deficit (Hyperactive) Disorder (ADD/ADHD), and dyslexia.

It has been shown that early intervention strategies can mitigate some of the effects of these disorders (Barnett, 1998), (Gorey, 2001). However, before interventions can be implemented, awareness of the problem needs to be obtained by the caregiver or guardians/parents. In LMICs, factors such as poverty, illiteracy of parents, and scarcity of resources can increase the lack of necessary stimulation and care children should receive (Engle & Black, 2008). These factors also contribute to the delayed discovery of neurodivergent disorders. There is therefore a need to assess several cognitive functions in young children to identify neural developmental issues and proceed with the necessary interventions.

There are several models of cognition defined by various bodies that each contain a set of cognitive domains (Harvey, 2019), (Baron & Leonberger, 2012), (Sabanathan, Wills, & Gladstone, 2015). Although there is no consensus on which model and accompanying domains are correct, five domains were present in all cognition models: attention, memory, executive functioning, language, and motor skills. Attention pertains to a person’s ability to focus on and differentiate between important information and non-relevant information, and how long this focus can be held. Memory touches on all aspects that we think of as memory, which includes working memory, prospective memory, explicit memory, procedural memory, and semantic memory (Harvey, 2019). Executive functioning is a person’s ability to execute a cognitive set (rules to follow when given a task, e.g., sorting cards according to color and not suit), mental flexibility, and inhibition. Language consists of two sub-domains, receptive and expressive language. The former is a person’s ability to understand language such as instructions, and the latter is the ability to express or convey meaning/ideas. Motor skills also consist of two sub-domains, fine and gross motor. The former pertains to small movements of the fingers and hand, manipulation of objects, and typing/writing, whereas the latter refers to larger movements such as walking, sitting, standing up, balance, and physical strength.

The developed assessment application presented here focuses on two domains, fine motor and language. Although it is important to assess all cognitive domains to acquire a full picture, this is not always possible. Fine motor assessment, as opposed to overall motor skills, was chosen because the nature of gross motor assessments makes it difficult to assess using a tablet. Furthermore, motor skills are the first to develop within the cognitive domain (Casey, Tottenham, Liston, & Durston, 2005), and is therefore well aligned to test early development. Language was chosen as the second domain as it also develops early on before other higher cognitive functions such as executive functioning (Shonkoff, Boyce, Cameron, & et al, 2008).

Motor function and control mostly reside in the motor cortex of the brain. Motor function impairment has been found to accompany ADHD (Dewey, Cantell, & Crawford, 2007), (Pitcher, Piek, & Hay, 2003), dyslexia (Fawcett & Nicolson, 1995), and anxiety disorders (Erez, Gordon, Sever, Sadeh, & Mintz, 2004). Motor function impairment, also classified as a disorder known as Developmental Coordination Disorder (DCD), has been found to affect certain facets of life, such as mathematical skills, reading, and writing (Alloway, 2007). Longitudinal studies have also linked motor ability to executive functioning later in life (Murray, et al., 2006) and academic performance in mathematics (Kurdek & Sinclair, 2001). Motor function has also been shown to predict levels of anxiety and depressive symptomatologies (Piek, Barrett, Smith, Rigoli, & Gasson, 2010).

Language is an integral part of how people interact with the world and is a complex construct. The complex nature of language can be described as a system comprising of many dimensions, namely phonology (sound system), lexicon (vocabulary), semantics (meaning), grammar (structure), pragmatics (communicative functions and conventions for language use), and discourse (integration of utterances into longer stretches of conversation or narrative). The acquisition of language starts shortly after birth when infants start to discriminate between different sound contrasts (McMurray & Aslin, 2005). The prelinguistic period is characterized by speech sounds, babbling, and longer sequences of sounds trying to mimic adult speech (Saaristo-Helin, Kunnari, & Savinainen-Makkonen, 2011); this is followed by gestures, indicating wants and interactions (Behne, Liszkowski, Carpenter, & Tomasello, 2012), and finally basic language comprehension such as recognizing his/her name and associating words with objects (Tincoff & Jusczyk, 1999). Language acquisition speeds up once the first word is spoken, ending the pre-linguistic period. On average, children will acquire 10 words per month up to about 50 words, whereby this acquisition rate increases to about 30 words per month (Goldfield & Reznick, 1990). Two-word speech, indicating basic grammatical knowledge developing (Schipke & Kauschke, 2011), becomes more apparent and develops into three- and four-word utterances. Auxiliary verbs are next to develop, and only later in language development come questions and negative sentences (Tyack & Ingram, 1977). Finally, language development ends at the end of the child's preschool years (Hoff, 2009). Language deficiencies have been linked to various developmental and educational outcomes, such as the link with memory (Conti-Ramsden & Durkin, 2007), social behavior, quality of friendships (Durkin & Conti-Ramsden, 2007), emotional difficulties, and academic failure (St Clair, Pickles, Durkin, & Conti-Ramsden, 2011), (Conti-Ramsden, Durkin, Simkon, & Knox, 2009).

Developmental assessment is a wide field of study, having specific tests and procedures depending on the domain assessed. To simplify this, two categories of developmental assessment are considered: classical and computerized. Classical developmental assessment employs a test using pen and paper, usually administered by a trained medical professional. Alternatively, a computerized developmental assessment is an application with which the child interacts, which in turn calculates the necessary results. Computerized assessments are commonly given via tablets. Tablet technology is commonly preferred due to its lightweight and compact design (Kucirkove, 2014) and ability to quickly distribute newer versions of test development cycles. Furthermore, it has been shown that preschool children can successfully interact with tablet technologies (Nacher, Jaen, Navarro, Catala, & Gonzalez, 2015) as they are becoming increasingly popular in early education (Chiong & Shuler, 2010), (Geist, 2012).

Here we describe the development and validity of an in-house developed tablet assessment application to asses fine motor and language abilities.

II. Methods

A. Test items

A total of eighteen test items were implemented, ten language-related and eight fine motor-related. The process of selecting these tests started by collecting data from current standardized developmental assessments. Six computerized developmental assessments and fifteen classical developmental assessments (8 language-related, 4 fine motor-related, and 3 assessment batteries containing both language and fine motor assessments) were analyzed and filtered according to implementability, whether or not they would keep their construct validity when implemented on a tablet, and presence in the standardized tests. The list of developmental assessments that were used in the construction of this tablet application are listed for reference: Griffiths Mental Developmental Scales (GDMS), Clinical Evaluation of Language Fundamentals (CELF) (Wiig, Secord, & Semel, 2006), British Picture Vocabulary Scale - III (BPVS3) (Dunn & Dunn, 2009), Early Repetition Battery (ERB) (Seeff-Gabriel, Chiat, & Roy, 2008), Receptive One Word Picture Vocabulary Test (ROWPVT) (Martin & Brownell, 2010), Expressive One Word Picture Vocabulary Test (EOWPVT) (Martin & Brownell, 2010), Early Years Toolbox (EYT) (Howard & Melhuish, 2017), Denver Developmental Scales Test (Frankenburg, Dodds, Archer, Shapiro, & Bresnick, 1992), Zurich Neuromotor Assessment (ZNA), McCarron Assessment of Neuromuscular Development (MAND) (McCarron, 1976), Peabody Developmental Motor Scales 2 (PDMS-2) (Folio & Fewell, 2000), Peabody Picture Vocabulary Test - IV (PPVT-4) (Dunn & Dunn, PPVT-4 Peabody Picture Vocabulary Test, 2007), Bruininks-Oseretsky Test of Motor Proficiency 2nd edition (BOT-2) (Bruininks & Bruininks, 2005), and Movement Assessment Battery for Children (MABC-2) (Henderson, Sugden, & Barnett, 2007).

Using a Java-based Android Studio integrated development environment (IDE), images were created (using GIMP) and sourced (from Google, taking note of copyright and selecting only free-to-use images).

The first set of tests, known as option selection tests, focused on measurements of language skills. For this assessment, individuals are shown a stimulus, the stimulus needs to be understood, and the individual must respond by touching an option. Test items similar to these can be found in a variety of standardized developmental screening tests such as BAS3's, BPVS-3, ROWPVT-4, PPVT-4, EYT, DDST, and CELF. The following metrics are recorded for each test: time to first response, time to correct response, and the number of options selected.

a) *Object Recall*
This is a short-term memory test in which an individual is shown a single object on the application's screen for a short period of time. Once the object is no longer visible, a grid of various objects, including the originally shown object, is presented. The individual is then asked to choose the original object.

b) *Choose Associated Word*
Here the individual is shown an object and four words. The individual is asked to select the word that best matches the object. If the individual desires to audibly hear a word, upon selection of the word, the tablet will audibly read it aloud.

c) *Choose Associated Object*
Here the individual is shown a word and four objects and is asked to select the appropriate object described by the word. Again, if the individual desires to audibly hear the given word, upon selection of the word, the tablet will read it aloud.

d) *Follow Instructions*
This item tests the individual's ability to follow instructions. The test begins with the tablet reading aloud a set of instructions to the individual. The individual is then allowed the opportunity to perform the given instructions. Examples of instruction sets for this test include, "select the cat on top of the table" (as opposed to the cat below the table), "select the cat on the left", and "select the cat on the right."

e) *Choose Picture*
This test item is an analysis of both receptive language skills and phonological memory. Individuals are given a description of an object and asked to select the correct object matching the description from four options. Examples include "green triangle" and "blue circle." This test item measures the individual's ability to recall in memory a few selectors (such as 'green' and 'triangle') but also understand the prompt given and connect it to a visual image.

The next set of test group assessments, labeled as placement accuracy, focused on measurements of fine motor skills. Present in many standard developmental screening tests such as DDST, BOT-2, PDMS, ZNA, MABC, and MAND, movements such as picking up, manipulating objects, and placing them in a desired position are generally used to measure fine motor ability. Although the actions of picking up and manipulating objects are not tested within this developmental screening application, the motor planning and on-screen manipulation aspects are still present. Furthermore, motor planning along with spatial intelligence is required to build structures or fit objects into place within a larger structure. The metrics recorded include the translation and rotation of all the puzzle pieces as well as the total time spent. The difference between the desired and completed coordinates and orientation of the puzzle pieces can be determined from this. We determined error in the X direction, error in the Y direction, Euclidian error, as well as rotation error. Distance errors were in pixels and rotation errors in degrees.

f) *Place Object Exactly*
In this test item, the individual is shown an incomplete object with the missing part shown to the side. The individual is tasked to move and rotate the missing part to fit and complete the shown object.

g) *Build Puzzle*
This test item is similar to building a puzzle. The individual is shown several pieces that fit together which must be rotated and moved to complete the image.

The next set of fine motor skill tests requires the individual to perform time-related tapping tasks. These are similar to standard developmental screening tests such as ZNA, MAND, and a tablet test [Pitchford and Outhwaite, 2016]. The test items involve tap-related tasks that are commonly used to measure manual processing speed and manual coordination. According to Avanzino et al., a person's ability to keep rhythm using motor movements is a predictor of fine motor ability (Avanzino, et al., 2016)

h) *Speed Tap*
Here the individual is tasked with tapping a dot displayed in the middle of the screen as quickly and accurately as possible in a set amount of time. This tests an individual's manual processing speed. The number of taps and the coordinates of the tap on the screen are recorded for accuracy.

i) *Rhythmic Tap*
This test assesses visuomotor coordination by evaluating how well the individual can match and continue a specific rhythmic beat. An auditory (1000 Hz beep for 300 ms) and visual (border of screen flashes between black and white) rhythmic beat is given for a predefined time before vanishing and the individual is asked to continue tapping the screen at the same rhythmic beat. The coordinates of the tap on the screen are recorded along with the timing which can be used to determine congruency with the given beat.

The next two motor skill assessment tests are based on two standardized line/path tracing-related tests known as MABC and BOT-2 as well as two standardized connecting dots-related tests, PDMS and BOT-2.

j) *Connect the Dots*
This is the classic connect the dots game where the individual is instructed to connect the numbered dots by tracing a straight line between two dots on the screen using their finger and starting at the dot numbered as one. For each segment (between two sequential dots) the perpendicular distance is determined between the drawn line and the correct line.

k) *Trace Path*
This test requires the individual to trace their finger along a predefined path. The correct path is divided into several points and the distance between each point and the nearest point on the drawn path is

determined. The time to complete the task is also recorded.

The next test is used to assess precision motor skills and is related to tests in PDMS and BOT-2.

l) *Color Between the Lines*
Here the individual is shown an image outline to color in, as well as a color palette to the side. The individual must color in between the lines, using a variety of stroke sizes. Measurements for post analysis includes the number of pixels that were required to be colored in, but were not, and the number of pixels that should not be colored, but were. Although not instructed to the individual, all pixels outside of the image's border and those that formed the image's outline were expected to remain unaltered. The overall score for the test is calculated as the percentage of error pixels.

Another image-based test requires individuals to copy and redraw a given picture. This test is based on similar tests in the DDST, PDMS and Griffiths.

m) *Draw Object Shown*
For this test the individual is presented with an image of an object on the left side of the screen and asked to redraw the object, using their finger, on the right side of the screen. Individuals are not constrained to draw the image in the exact orientation or scale. Since there is no accepted way of measuring how well the individual can redraw the image presented, six different metrics were evaluated: Sum of Squared Differences (SSD), Cosine Similarity (CS), Hausdorff Distance (HD) (Rucklidge, 1996), Scale Invariant Feature Transform (SIFT) (Lowe, 1999), a Convolutional Neural Network (CNN) based feature extraction method, and a machine learning based image similarity application protocol interface (API). The SIFT algorithm extracts keypoints from an image and descriptors of those keypoints. These keypoints and their accompanying descriptors are compared to the keypoints and descriptors of other images to find matches. The CNN feature extraction uses a modified pretrained ResCNN model, specifically ResNet-152. This model was specifically selected for its size and accuracy (He, Zhang, Ren, & Sun, 2016), (Anwar, 2019). The fully connected neural network at the end of the ResNet-152 network (that has the purpose of classifying images) was removed and the raw feature vectors were used to compare images. The final metric, DeepAI's Image, requires two images as the input and returns a similarity score as the output.

The last set of tests involve auditory processing. These tests involve the oral recall of numbers and sentences which is derived from related test items in BAS3, CELF, and Griffith; orally identifying the antonym of an image as based on similar tests in DDST and Griffiths; the oral description of a picture, which is an expressive language assessment, and based on tests in DDST, BAS3, and EOWPVT; and the inclusion of a novel computerized test item which assesses the pronunciation of a word. Medical professionals commonly look out for incorrect pronunciation when evaluating children undergoing an assessment battery. This is useful as language can be identified through this meta-analysis process (analysis of how the individual pronounces a word rather than what is heard). For all the tests in this set audio files were recorded and stored for post-processing. Two automatic speech recognition (ASR) techniques were used to make transcriptions of the audio recordings, namely DeepSpeech2 (Amodei, Ananthanaravanan, Anubhai, Bai, & et al, 2016), and Google's Speech-to-text API. Two metrics were used to assess the tests, word error rate (WER) and character error rate (CER). Word error rate allocated a point for each word in an individual's transcription that matched a word in the desired/correct transcription. Character error rate, a more lenient metric, calculated the number of changes needed in order for the two transcriptions to match (ie., the number of letters that need to be added, removed, or altered).

n) *Number Recall*
In this test a sequence of numbers is read out load. Following this, the individual is asked to repeat the sequence verbally. With each increasing trial, the sequence length increases by one number.

o) *Sentence Recall*
Similar to the previous test, in this test a sentence is read out load and the individual is asked to verbally repeat the test. Each subsequent trial will increase the sentence length by one word.

p) *Give Opposite*
In this test the individual is shown a word, the word is also read out loud, and the individual is then asked to verbally provide an antonym for the word.

q) *Describe Picture*
This test item shows a picture of a scene, and the individual is asked to verbally describe the scene. For this test item there is no original transcription with which to compare for correctness, and, therefore, three keywords were assigned to each picture. During post-processing analysis, each recording was checked for the number of keywords. If a keyword was present, a point was awarded for that picture. For example, if the individual was shown an image of a boy sitting on a chair reading a book. The three keywords associated with that image would be boy/person, sitting/sit, and read/reading.

r) *Word Pronunciation*
For the final test the individual is shown a word which is also read out loud. The individual is then asked to simply repeat the word.

B. *Testing.*

We tested each of the test items to see if we can clearly distinguish between good versus bad attempts. This serves as an initial validation test for each test item and does not constitute a methodologically sound experiment that will form part of future work. For the testing, one of the authors proceeded to perform the tests by intentionally simulating good and bad attempts. In addition, we also performed these initial tests to determine which metric assessment techniques worked best for the *Draw Objects Shown* test item as well as for the auditory items.

III. Results

For test items a) through e) the application correctly detected whether the individual made the correct selection, how many selections were made, and the total time to selection. For test items f) and g), using different objects (and different numbers of puzzle pieces for g), we clearly showed measurable differences for all the error metrics for the simulated good versus bad attempts ($p<0.01$).

Figure 1 shows the results for test item h) with attempt 1 simulating the good attempt and attempt 2 simulating a bad attempt. The scenarios refer to the different times given to the individual to tap the screen starting with 10s and increasing by 5s to 30s. Figure 2 shows the results for test item i) again with attempt 1 simulating the good attempt and attempt 2 simulating a bad attempt. The scenarios refer to different rhythmic beats of 2 Hz, 1 Hz, 0.67 Hz, 0.5 Hz, and 0.4 Hz.

Figure 3 shows the error distance results for test item j) consisting of 12 segments. Similarly, figure 4 shows the mean error results for test item k) for three different line paths and for the good and bad attempts.

We were also able to distinguish between the two attempts for test item l) by calculating an error pixel percentage. Figure 5 shows an example with the good attempt giving a score of 97.7% and the bad attempt resulting in a score of 69.6%.

For test item m) the ResNet model was the only one that could reliably score the correctly drawn image higher than the two incorrectly drawn images.

For the various auditory test items, it was found that Google's Speech-to-text API worked best to transcribe the recordings.

IV. Discussion

Our initial results indicate the feasibility of the developed assessment application for preschool children. The application was developed to be modular (it is easy to add or remove tests) and easily modifiable (one can quickly and effortlessly change the specific pictures, words, or sentences used in the different test items). Our results indicate that it is possible to distinguish between good and bad attempts at the numerous different language and fine motor assessment items.

The first set of tests, collectively called the object selection tests, are used to assess language skills. A quick correct answer when making a selection gives an indication of the individual's receptive language skills. Faster processing speed with regard to language would allow the individual to perceive the stimulus and form an answer quicker (Leonard, et al., 2007). Furthermore, the selection of multiple answers can indicate hesitation or possible confusion which could warrant further investigation.

The next couple of tests (f and g), requiring the individual to move and rotate items, assess fine motor skills. Lower distance errors (where objects were placed closer to the desired locations) would correlate with better fine motor skills as the individual was better able to manipulate and translate the object(s) on screen. More specifically, test item g) gives insight into the individual's visuospatial intelligence as it is used to determine how pieces moved around and placed should fit together. More puzzle pieces require more spatial organization, therefore increasing the number of pieces indicates varying levels of visuospatial intelligence (Cameron, et al., 2012).

Both of the tapping tests (h and i) required the individual to tap the dot as accurately as possible. Here, the metric distance error indicates how far off the individual was from the dot's center. Lower scores would indicate better fine motor skills, such as visuomotor integration (using one's visual perception to guide where to tap on the screen). Furthermore, motor timing has a strong link to good overall motor performance (Falter & Noreika, 2011). A person's rhythmic capability indicates the ability to estimate time (sub- and supra-second) and uphold motor timing rhythm (the rhythm of movement and timing of movements). We also measured variance in the tapping frequency with higher variance indicating less consistency, which in turn could indicate motor timing ability (Noreika, Falter, & Rubia, 2013).

The two tracing accuracy test items (j and k) used distance error as a metric, with mean and variance calculated. Higher variance, average error, and overall distance error would indicate that the individual had traced the line (or between two dots) with less accuracy, thus indicating less fine motor precision and control. Therefore, less distance error would correlate with better fine motor ability (Cohen, Bravi, Bagni, & Minciacchi, 2018). Coloring an image (item l) requires fine motor control (Wehrmann, Chiu, Reid, & Sinclair, 2006), therefore better fine motor ability and control over fine movements would allow the individual to color the image more accurately and would result in fewer error pixels present. Furthermore, better fine motor control would allow the individual to redraw a picture more accurately (item m). Thus, the similarity score generated between the stock image (presented as a stimulus to the individual) and the drawn image can be used to indicate fine motor ability (Vimercati, et al., 2015). The higher the similarity score, the better the fine motor ability. The ResNet model was the only measure that consistently indicated that the intended drawn image was more similar than that of the intentionally incorrectly drawn images. This metric, however, still needs refinement.

Finally, the audio analysis test items are to be viewed with as much confidence as the confidence in the automatic speech recognition system. These systems can influence the results if they have not been specifically trained to avoid bias. The WER and CER metrics measureed how similar the spoken transcription is to the true transcription. Less WER and CER indicate a closer match between the two indicating that the individual understood the stimulus and objective. Therefore, lower WER and CER indicates better receptive language (Viding, et al., 2004).

There is a word of caution here. All stimuli (objects and words) used in this work were merely used as placeholders to demonstrate the capabilities of the application and accompanying processing pipeline. These stimuli might not suit all environments. The specific stimuli used can affect the results in several ways and are dependent on the individual's familiarity with the stimulus, the individual's home language, whether the stimulus has an ambiguous meaning, or if the stimulus is too difficult for the individual's age. When choosing suitable stimuli for the test one should be guarded against biasing results through the use of specific stimuli and only cultural and age-appropriate stimuli should be used. Furthermore, the results derived by the processing pipeline are only valid if the individual understood what was expected of them. If the individual did not understand what is needed of them and performed the test item, the results would not yield an accurate representation of the construct being measured. The degree to which the individual understands the stimuli needs to be noted and taken into account when administering the test.

V. CONCLUSION

This work reviews the development and characterization of an automated tablet assessment application and its accompanying processing pipeline. This tablet application focused on two main domains, fine motor and language, specifically for preschool children. Eighteen test items were sourced, created, and implemented on a tablet application, and processing pipelines created for each of the test items. The results indicate that the application works as intended and is able to record multiple measures simultaneously. In publishing the characterization and results of the automated tablet assessment application design, it is hoped this application will serve as a framework for the development of objective assessment tests in the future.

REFERENCES

Agyei, S. B., van der Weel, F. R., & van der Meer, A. L. (2016). Longitudinal study of preterm and full-term infants: High-density EEG analysis of cortical activity in response to visual motion. *Neuropsychologia, 84*, 89-104.

Alloway, T. P. (2007). Working memory, reading, and mathematical skills in children with developmental coordination disorder. *Journal of Experimental Child Psychology, 96*, 20-36.

Amodei, D., Ananthanaravanan, A., Anubhai, R., Bai, J., & et al. (2016). Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. *Proceedings of the 33rd International Conference on Machine Learning.* New York.

Anwar, A. (2019). *Difference between AlexNet, VGGNet, ResNet, and Inception*. Retrieved from Towards Data Science: https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaaecccc96

Avanzino, L., Pelosin, E., Vicario, C. M., Lagravinese, G., Abbruzzesse, G., & Martino, D. (2016). Time Processing and Motor Control in Movement Disorders. *Frontiers in Human Neuroscience, 10*, 631.

Barnett, S. (1998). Long-term cognitive and academic effects of early childhood education on children in poverty. *Preventitive Medicine, 27*, 204-207.

Baron, I. S., & Leonberger, K. A. (2012). Assessment of intelligence in the preschool period. *Neuropsychology Reviews, 22*, 334-344.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology, 30*, 359-375.

Bruininks, R. H., & Bruininks, B. D. (2005). *BOT-2 Bruininks-Oseretsky Test of Motor Proficiency* (2nd ed.).

Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine Motor Skills and Executive Function Both Contribute to Kindergarten Achievement. *Child Development, 83*, 1229-1244.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in COgnitive Sciences, 9*, 104-110.

Chiong, C., & Shuler, C. (2010). *Learning: Is there an app for that?* Joan Ganz Cooney Center.

Chugani, H. T. (1998). A critical period of brain development: Studies of cerebral glucose utilization with PET. *Preventitive Medicine, 27*, 184-188.

Cohen, E. J., Bravi, R., Bagni, M. A., & Minciacchi, D. (2018). Precision in drawing and tracing tasks: Different measures for different aspects of fine motor control. *Human Movement Science, 61*, 177-188.

Conti-Ramsden, G., & Durkin, K. (2007). Phonological short-term memory, language, and literacy: Developmental relationships in early adolescence in young people with SLI. *Journal of Child Psychology and Psychiatry, 48*, 147-156.

Conti-Ramsden, G., Durkin, K., Simkon, Z., & Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language and COmmunication Disorders, 44*, 15-35.

Dewey, D., Cantell, M., & Crawford, S. G. (2007). Motor and gestural performance in children with autism spectrum disorders, developmental coordination disorder, and/or attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society, 13*, 246-256.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Peabody Picture Vocabulary Test* (4th ed.).

Dunn, L. M., & Dunn, D. M. (2009). *The British picture vocabulary scale* (3rd ed.). GL Assessment Limited.

Durkin, K., & Conti-Ramsden, G. (2007). Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. *Child Development, 78*, 1441-1457.

Engle, P. L., & Black, M. M. (2008). The effect of poverty on child development and educational outcomes. *Annals of the New York Academy of Sciences, 1136*, 143-356.

Erez, O., Gordon, C. R., Sever, J., Sadeh, A., & Mintz, M. (2004). Balance dysfunction in childhood anxiety: Findings and theoretical approach. *Journal of Anxiety Disorders, 18*, 341-356.

Falter, C. M., & Noreika, V. (2011). Interval Timing Deficits and Abnormal Cognitive Development. *Frontiers in Integrative Neuroscience, 5*, 1-2.

Fawcett, A. J., & Nicolson, R. I. (1995). Persistent deficits in motor skill of children with dyslexia. *Journal of Motor Behavior, 27*, 235-240.

Folio, M. R., & Fewell, R. R. (2000). *PDMS-2 Peabody Developmental Motor Scales* (2nd ed.).

Frankenburg, W. K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. (1992). The Denver II: A major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics, 89*, 91-97.

Geist, E. (2012). A qualitative examination of two-year-olds interaction with tablet based interactive technology. *Journal of Instructional Psychology, 39*, 26-35.

Goldfield, B. A., & Reznick, J. S. (1990). Early Lexical Acquisition: Rate, Content, And The Vocabulary Spurt. *Journal of Child Language, 17*, 171-183.

Gorey, K. M. (2001). Early Childhood Education: A Meta-Analytic Affirmation of the Short- and Long-Term Benefits of Educational Opportunity. *School Psychology Quarterly, 16*, 9-30.

Harvey, P. D. (2019). Domains of cognition and their assessment. *Dialogues in Clinical Neuroscience, 21*, 227-237.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp. 770-778). Las Vegas, USA.

Henderson, S. E., Sugden, D. A., & Barnett, A. (2007). *Movement Assessment Battery for Children* (2nd ed.).

Hoff, E. (2009). *Language Development* (4th ed.). Belmont, CA: Wadsworth.

Howard, S. J., & Melhuish, E. (2017). An Early Years Toolbox for Assessing Early Executive Function, Language, Self-Regulation, and Social Development: Validity, Reliability, and Preliminary Norms. *Journal of Psychoeducational Assessment, 35*, 255-275.

Kucirkove, N. (2014).). iPads in early education: Separating assumptions and evidence. *Frontiers in Psychology, 5*, 715.

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*, 451-455.

Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research, 50*, 408-428.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the IEEE International Conference on Computer Vision*, (pp. 1150-1157). Kerkyra, Greece.

Martin, N. A., & Brownell, R. (2010). *EOWPVT-4: Expressive One-Word Picture Vocabulary Test* (4th ed.).

Martin, N. A., & Brownell, R. (2010). *ROWPVT-4: Receptive One-Word Picture Vocabulary Test* (4th ed.).

McCarron, L. T. (1976). *MAND : McCarron assessment of neuromuscular development, fine and gross motor abilities.* Common Market Press.

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95*, B15-26.

Murray, G. K., Veijola, J., Moilanen, K., Miettunen, J., Glahn, D. C., Cannon, T. D., . . . Isohanni, M. (2006). Infant motor development is associated with adult cognitive categorisation in a longitudinal birth cohort study. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 47*, 25-29.

Nacher, V., Jaen, J., Navarro, E., Catala, A., & Gonzalez, P. (2015). Multi-touch gestures for prekindergarten children. *International Journal of Human Computer Studies, 73*, 37-51.

Noreika, V., Falter, C. M., & Rubia, K. (2013).). Timing deficits in attention-deficit/hyperactivity disorder (ADHD): Evidence from neurocognitive and neuroimaging studies. *Neuropsychologia, 51*, 235-266.

Piek, J. P., Barrett, N. C., Smith, L. M., Rigoli, D., & Gasson, N. (2010). Do motor skills in infancy and early childhood predict anxious and depressive symptomatology at school age? *Human Movement Science, 29*, 777-786.

Pitcher, T. M., Piek, J. P., & Hay, D. A. (2003). Fine and gross motor ability in males with ADHD. *Developmental Medicine and Child Neurology, 45*, 525-535.

Rucklidge, W. (1996). *Efficient visual recognition using the Hausdorff distance.* Berlin: Springer.

Saaristo-Helin, K., Kunnari, S., & Savinainen-Makkonen, T. (2011). Phonological development in children learning Finnish: A review. *First Language, 31*, 342-363.

Sabanathan, S., Wills, B., & Gladstone, M. (2015). Child development assessment tools in low-income and middle-income countries: How can we use them more appropriately? *Archives of Disease in Childhood, 100*, 482-488.

Schipke, C. S., & Kauschke, C. (2011). Early word formation in German language acquisition: A study on word formation growth during the second and third years. *First Language, 31*, 67-82.

Seeff-Gabriel, B., Chiat, S., & Roy, P. (2008). *Early Repetition Battery.* London: Pearson Assessment.

Shonkoff, J. P., Boyce, T. W., Cameron, J., & et al. (2008). *The Timing and Quality of Early Experiences Combine to Shape Brain Architecture.* Retrieved from The National Scientific Council on the Developing Child: http://www.developingchild.net

St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders, 44*, 186-199.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*, 172-175.

Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language, 4*, 211-224.

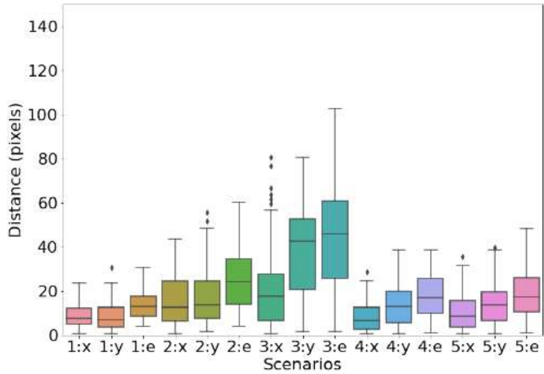
Viding, E., Spinath, F. M., Price, T. S., Bishop, D. V., Dale, P. S., & Plomin, R. (2004). Genetic and environmental influence on language impairment in 4-year-old same-sex and opposite-sex twins. *Journal of Child Psychology and Psychiatry, 45*, 315-325.

Vimercati, S. L., Galli, M., Stella, G., Caiazzo, G., Ancillao, A., & Albertini, G. (2015). Clumsiness in fine motor tasks: Evidence from the quantitative drawing evaluation of children with Down Syndrome. *Journal of Intellectual Disability Research, 59*, 248-256.

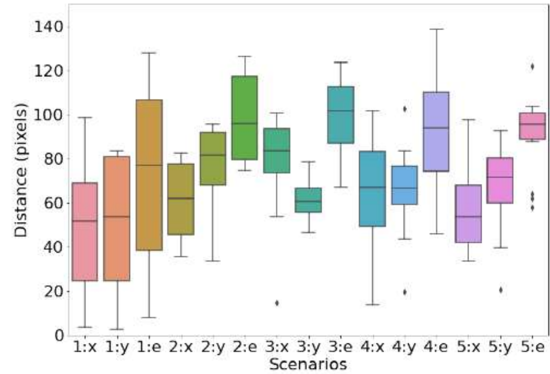
Wehrmann, S., Chiu, T., Reid, D., & Sinclair, G. (2006).). Evaluation of occupational therapy schoolbased consultation service for students with fine motor difficulties. *Canadian Journal of Occupational Therapy, 73*, 225-235.

WHO. (2018). *Improving early childhood development: WHO guideline.* WHO.

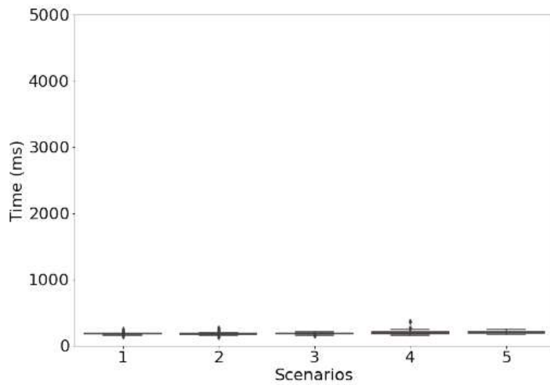
Wiig, E. H., Secord, W., & Semel, E. M. (2006). *CELF Preschool 2 UK: Clinical Evaluation of Language Fundamentals, Preschool.* Harcourt Assessment.



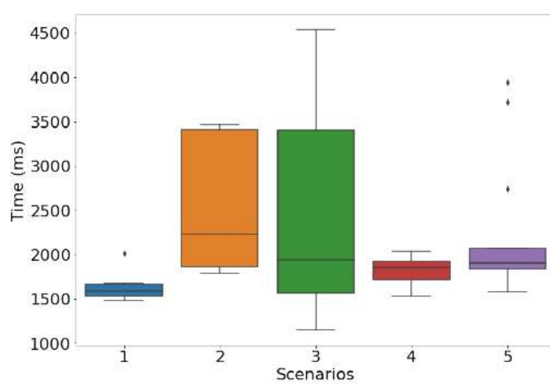
(a) Attempt 1 distance error variance.



(b) Attempt 2 distance error variance.

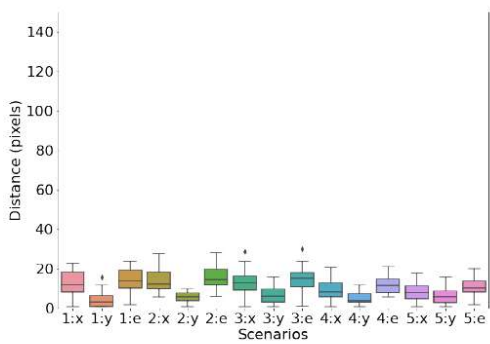


(c) Attempt 1 inter tap time variance.

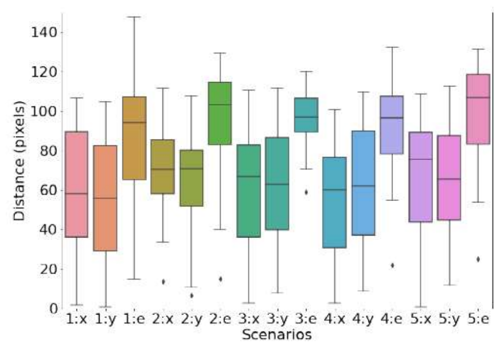


(d) Attempt 2 inter tap time variance.

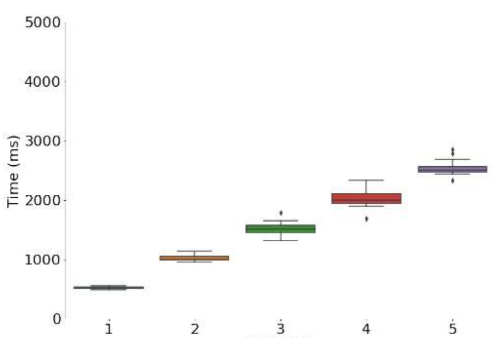
Fig. 1. Results of test item h) where a) and b) shows the mean x, y, and Euclidian distance errors from the middle of the target for attempt 1 and 2, respectively, and c) and d) show the inter-tap variability for attempt 1 and 2, respectively.



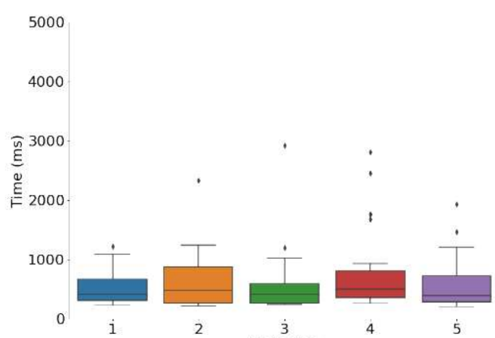
(a) Attempt 1 distance error variance.



(b) Attempt 2 distance error variance.

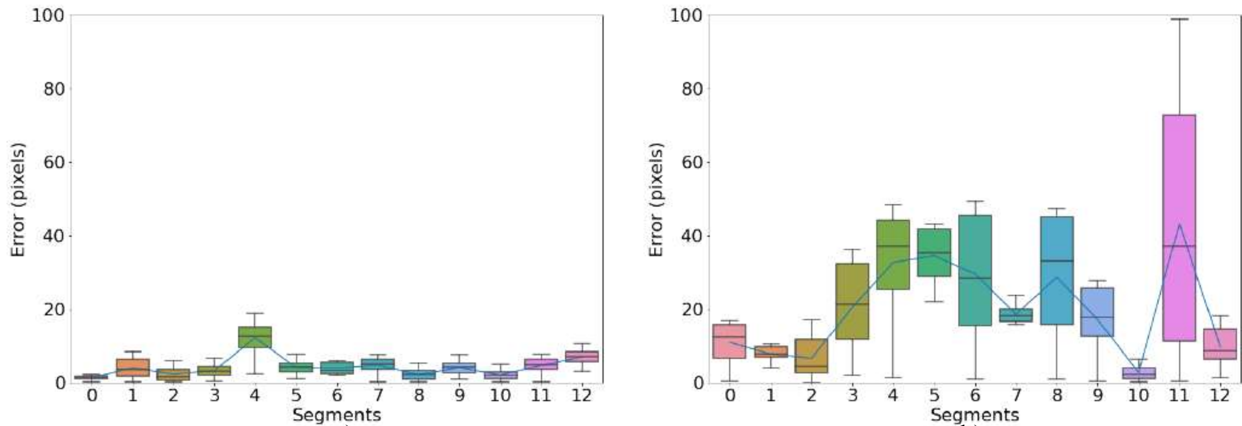


(c) Attempt 1 inter tap time variance.



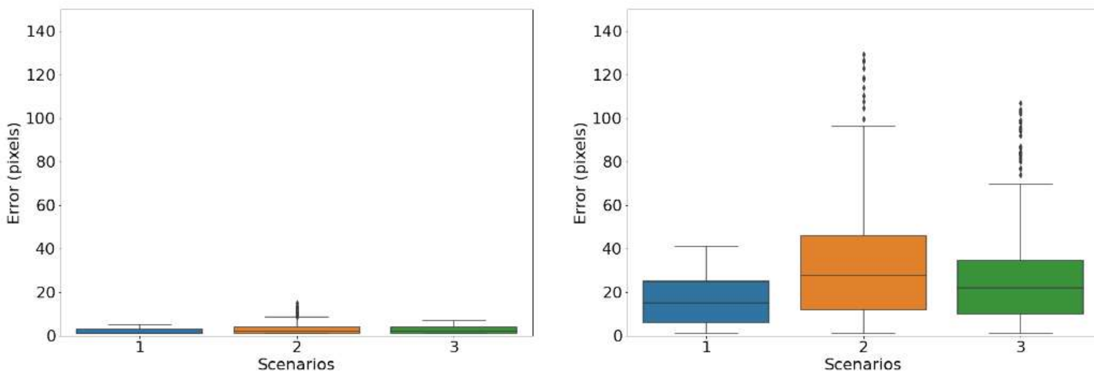
(d) Attempt 2 inter tap time variance.

Fig. 2. Results of test item i) where a) and b) shows the mean x, y, and Euclidian distance errors from the middle of the target for attempt 1 and 2, respectively, and c) and d) show the inter-tap variability for attempt 1 and 2, respectively.



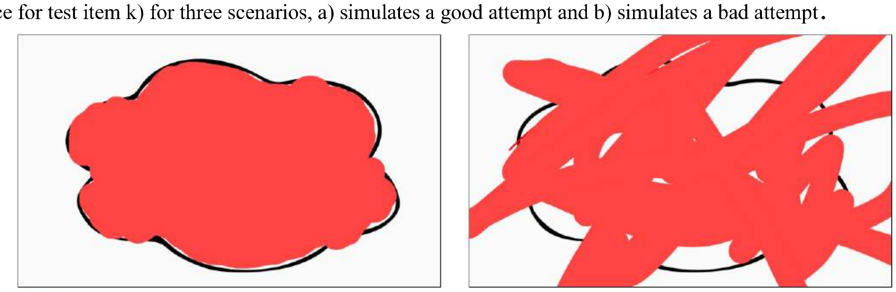
a) b)

Fig. 3. Perpendicular distance (in pixels) of the line segment drawn from the optimal line for a) good attempt and b) bad attempt.



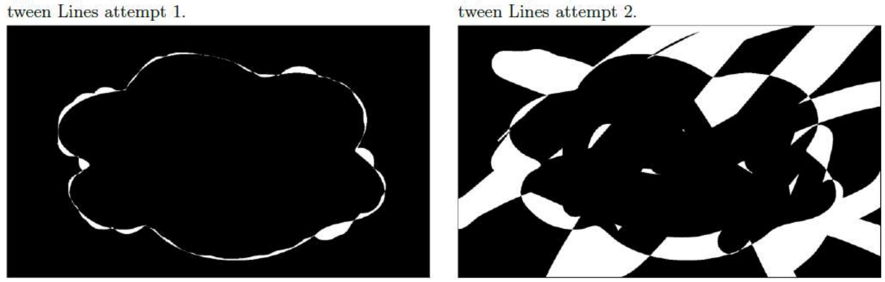
a) b)

Fig. 4. Mean error distance for test item k) for three scenarios, a) simulates a good attempt and b) simulates a bad attempt.



(a) Image captured from the Colour Between Lines attempt 1.

(b) Image captured from the Colour Between Lines attempt 2.



(c) Error pixel map for the Colour Between Lines attempt 1.

(d) Error pixel map for the Colour Between Lines attempt 2.

Fig. 5. Results for test item l) where a) attempt 1 simulates the good attempt and b) attempt 2 simulates the bad attempt. The error map is shown in c) for attempt 1 and d) for attempt 2.