

ORIGINAL ARTICLE

Characterizing cognitive control abilities in children with 16p11.2 deletion using adaptive ‘video game’ technology: a pilot study

JA Anguera^{1,2}, AN Brandes-Aitken¹, CE Rolle¹, SN Skinner¹, SS Desai¹, JD Bower³, WE Martucci³, WK Chung⁴, EH Sherr^{1,5} and EJ Marco^{1,2,5}

Assessing cognitive abilities in children is challenging for two primary reasons: lack of testing engagement can lead to low testing sensitivity and inherent performance variability. Here we sought to explore whether an engaging, adaptive digital cognitive platform built to look and feel like a video game would reliably measure attention-based abilities in children with and without neurodevelopmental disabilities related to a known genetic condition, 16p11.2 deletion. We assessed 20 children with 16p11.2 deletion, a genetic variation implicated in attention deficit/hyperactivity disorder and autism, as well as 16 siblings without the deletion and 75 neurotypical age-matched children. Deletion carriers showed significantly slower response times and greater response variability when compared with all non-carriers; by comparison, traditional non-adaptive selective attention assessments were unable to discriminate group differences. This phenotypic characterization highlights the potential power of administering tools that integrate adaptive psychophysical mechanics into video-game-style mechanics to achieve robust, reliable measurements.

Translational Psychiatry (2016) 6, e893; doi:10.1038/tp.2016.178; published online 20 September 2016

INTRODUCTION

Cognition is typically associated with measures of intelligence (for example, intellectual quotient (IQ)¹), and is a reflection of one's ability to perform higher-level processes by engaging specific mechanisms associated with learning, memory and reasoning. Such acts require the engagement of a specific subset of cognitive resources called cognitive control abilities,^{2–5} which engage the underlying neural mechanisms associated with attention, working memory and goal-management faculties.⁶ These abilities are often assessed with validated pencil-and-paper approaches or, now more commonly with these same paradigms deployed on either desktop or laptop computers. These approaches are often less than ideal when assessing pediatric populations, as children have highly varied degree of testing engagement, leading to low test sensitivity.^{7–9} This is especially concerning when characterizing clinical populations, as increased performance variability in these groups often exceeds the range of testing sensitivity,^{7–9} limiting the ability to characterize cognitive deficits in certain populations. A proper assessment of cognitive control abilities in children is especially important, as these abilities allow children to interact with their complex environment in a goal-directed manner,¹⁰ are predictive of academic performance¹¹ and are correlated with overall quality of life.¹² For pediatric clinical populations, this characterization is especially critical as they are often assessed in an indirect fashion through intelligence quotients, parent report questionnaires¹³ and/or behavioral challenges,¹⁴ each of which fail to properly characterize these abilities in a direct manner.

One approach to make testing more robust and user-friendly is to present material in an optimally engaging manner, a strategy particularly beneficial when assessing children. The rise of digital health technologies facilitates the ability to administer these types of tests on tablet-based technologies (that is, iPad) in a game-like manner.¹⁵ For instance, Dundar and Akcayir¹⁶ assessed tablet-based reading compared with book reading in school-aged children, and discovered that students preferred tablet-based reading, reporting it to be more enjoyable. Another approach used to optimize the testing experience involves the integration of adaptive staircase algorithms, as the incorporation of such approaches lead to more reliable assessments that can be completed in a timely manner. This approach, rooted in psychophysical research,¹⁷ has been a powerful way to ensure that individuals perform at their ability level on a given task, mitigating the possibility of floor/ceiling effects. With respect to assessing individual abilities, the incorporation of adaptive mechanics acts as a normalizing agent for each individual in accordance with their underlying cognitive abilities,¹⁸ facilitating fair comparisons between groups (for example, neurotypical and study populations).

Adaptive mechanics in a consumer-style video game experience could potentially assist in the challenge of interrogating cognitive abilities in a pediatric patient population. This synergistic approach would seemingly raise one's level of engagement by making the testing experience more enjoyable and with greater sensitivity to individual differences, a key aspect typically missing in both clinical and research settings when testing these populations. Video game approaches have previously been utilized in clinical adult populations (for example, stroke,^{19,20}

¹Department of Neurology, University of California, San Francisco, San Francisco, CA, USA; ²Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA; ³Akili Interactive Labs, Boston, MA, USA; ⁴Department of Pediatrics, Columbia University Medical Center, New York, NY, USA and ⁵Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA. Correspondence: JA Anguera or EJ Marco, University of California, San Francisco, Mission Bay – Sandler Neurosciences Center, UCSF MC 0444, 675 Nelson Rising Lane, Room 502, San Francisco, CA 94158, USA.

E-mail: joaquin.anguera@ucsf.edu or elysa.marco@ucsf.edu

Received 6 March 2016; revised 13 July 2016; accepted 18 July 2016

Table 1. Task description

	<i>Project: EVO</i>	<i>Flanker task</i>	<i>Visual search task</i>
Task description		Indicate direction of central target, which is flanked by distractors that are either in the same or opposite direction as the target	Search for a target (green 'II'), among a number of red and non-vertically aligned green distracting 'IIs', indicating the side of the II opening being on the top or the bottom (as shown).
Primary Reference	Anguera et al. ¹⁸	Eriksen and Eriksen ³⁹	Treisman (1982)
Type of Attention Measured	Selective attention with distractors, alone and while multi-tasking	Selective attention with distractors	Selective attention with distractors, with search component
Primary measure to assess attention	Response time to targets under differing distraction loads	Response time to congruent versus incongruent targets	Response time to targets under high versus low loads of distractors
Respond to select stimuli	Yes	Yes	Yes
High/Low Difficulty Loads	Yes	No	Yes
Distracting or Irrelevant Stimuli	Yes	Yes	Yes
Trial-by-Trial Feedback	Yes	Yes	Yes
Incorporates high-level art and music to create immersive experience	Yes	No	No
Uses Adaptivity	Yes	No	No
Involves multi-tasking	Yes	No	No

schizophrenia²¹ and traumatic brain injury²²⁻²⁴); however, these are examples of using existing entertainment-based video games for assessment purposes rather than scientifically derived assessments that use video game mechanics for clinical assessments and/or training. This difference highlights the dissociation between two types of interactive digital media: those designed primarily for entertainment ('video games') and those created for the purpose of cognitive assessment or enhancement.⁶ There are few examples of scientifically derived serious games used for clinical assessments, and, to the best of our knowledge, no examples of an entertainment-quality video game developed or validated for profiling cognitive abilities in clinical populations. Indeed, our previous work demonstrated the utility of incorporating adaptive algorithms in a video game for enhancing cognitive control,¹⁸ suggesting that similar cognitively targeted technology could be especially powerful in characterizing cognitive abilities in both healthy and clinical populations.

Here we administered a novel digital platform embedded with adaptive algorithms designed to assess cognitive control abilities associated with selective attention to children with and without a specific 16p11.2 BP4-BP5 deletion. This population was selected as children with the 16p11.2 BP4-BP5 deletion show high prevalence of inattention as well as language and social challenges when evaluated using clinical records and parent report tools.^{25,26} Nineteen percent of deletion carriers meet criteria for attention deficit disorder,²⁷ with 26% reaching DSM-IV-TR diagnostic criteria for autism spectrum disorders. Measurements of selective attention that involve distraction and inhibitory function like the Flanker task have shown attentional impairments in both attention deficit/hyperactivity disorder (ADHD)²⁸⁻³⁰ and autistic children.^{31,32} However, the visual search task (which measures selective attention with distraction under an increasing distractibility load) has shown similar effects in children with ADHD,³³⁻³⁵ but the opposite effects in those with autism.^{36,37} These findings epitomize recent meta-analysis findings by Karalunas et al.³⁸ that suggest these types of assessments typically show small to moderate effect sizes when attempting to dissociate children with autism or ADHD to neurotypical controls. Despite evidence of clinically significant attention challenges in the 16p11.2 population, there has not yet been a study utilizing direct assessment measures of attention abilities in the presence of distraction.

Our study was predominantly conducted over the course of two 16p11.2 family meetings sponsored by the Simons Foundation in 2013 and 2015 with the exception of a single local family. The assessments were conducted in a semi-private conference room that allowed for a 'real-world' testing environment but also inherently requires greater diagnostic sensitivity than testing in a controlled environment. These meetings facilitated our ability to recruit individuals with the 16p11.2 deletion who live throughout the world as well as their siblings who are age matched but do not carry the 16p11.2 deletion. Siblings represent an ideal comparison group due the sharing of 50% of their additional genetic material and the shared familial environment. We hypothesized that 16p11.2 deletion carriers would show slower and more variable reaction times, in accordance with historic literature from ADHD cohorts, compared with age-matched non-carrier siblings and unrelated healthy adolescents. We also hypothesized that this novel approach in assessing selective attention would reveal group-level deficits that traditional non-adaptive tests (Flanker and Visual Search) would not be able to uncover, given the modest sensitivity that non-adaptive platforms have shown in distinguishing between healthy and similarly affected populations.³⁸

MATERIALS AND METHODS

Participants

One hundred eleven children participated in this study: 91 children (mean age 10.7 years ± 2.2, 41 females) who were not carriers for 16p11.2 deletion, and 20 children (mean age 10.1 years ± 3.0, 6 females) who were carriers for 16p11.2 deletion (Table 1). 16p11.2 deletion carriers and their families were recruited from the broader Simons Variations in Individuals Project (VIP), where families across the world with this genetic disorder were invited to attend a family conference and participate in different research projects. In addition, 75 unrelated, unaffected children were recruited from ongoing work involving the characterization of attention abilities using a novel digital cognitive assessment platform (Project: EVO (EVO)), and tested in a traditional laboratory setting. This 75-person sample was included to provide a representative depiction of performance on this novel task in children (who are inherently variable from both an age and ability perspective), and determine whether any observed differences between affected and unaffected sibling could be replicated against a

Table 2. Demographic profiles of participants^a

	Carriers N = 13	Non-carrier siblings N = 11	Unrelated controls N = 24	F-value (P-value)
<i>IQ (subset)</i>				
Verbal IQ	84.3 ± 11.4	108.3 ± 13.3	124.9 ± 11.9	51.12 (< 0.0001)
Non-verbal IQ	88.5 ± 10.1	100.8 ± 10.6	113.6 ± 12	19.73 (< 0.0001)
Full scale IQ	85.5 ± 10.5	103.6 ± 12.5	119.1 ± 10.5	41.3 (< 0.0001)
ADHD diagnosis	6 (46%)	2 (22%)	0	
Autism spectrum disorder	1 (8%)	1 (9%)	0	
Learning disorder	2 (15%)	0	0	
Anxiety disorder	0	1 (14%)	0	
Mood disorder	1 (8%)	0	0	
Age (full sample)	N = 20 10.1 ± 3	N = 16 10.1 ± 2.8	N = 75 10.8 ± 1.4	2.41 (0.054)
Gender (full sample)	6 female	7 female	34 female	1.54 (0.46) ^b

Abbreviations: ADHD, attention deficit/hyperactivity disorder; IQ, intellectual quotient. ^aCognitive assessments and clinical evaluations were conducted as part of the broader Simons Foundation 16p11.2 and UCSF Sensory Neurodevelopment and Autism study. Verbal IQ was calculated from the standard verbal reasoning score of the DAS and the verbal comprehension index of the WISC-IV. Non-verbal IQ is calculated from the standard non-verbal reasoning score of the DAS and the perceptual reasoning index of the WISC-IV. Full scale IQ is calculated from general conceptual ability score of the DAS and the full scale IQ composite score of the WISC-IV. ^bChi-square with 2 df.

larger sample of unaffected children given the non-random convenient sample selection at the family meeting.

Forty-eight individuals of the total sample had IQ data with either the Differential Abilities Scales—2nd edition⁴⁰ or the Wechsler Intelligence Scale for Children-IV⁴¹ through the broader Simons VIP (<https://sfari.org/resources/autism-cohorts/simons-vip>) and UCSF Sensory Neurodevelopment and Autism Program, with verbal IQ (VIQ) and non-verbal IQ (NVIQ) specifically assessed (Table 2). Legal guardians provided written informed consent, and minors gave assent to participate. The study was approved by the University of California, San Francisco Committee on Human Research. All subjects who took part at the family meetings were asked to complete two non-adaptive traditional assessments (Flanker and Visual Search) and one novel, adaptive assessment. Test administration occurred in semi-private rooms at the conference, with participants seated at tables with one administrator present to both explain and monitor each assessment. Headphones were provided to participants to decrease any environmental distractions and create a standardized testing experience for each participant. All assessments were administered using an iPad 2 in a counter-balanced manner.

Project: EVO

EVO is a digital cognitive assessment and intervention system developed by Akili Interactive Labs to assess and train cognitive control abilities in clinical populations through immersion in an adaptive, high-interference environment that is built to look and feel like a consumer-grade action video game (Figure 1a). EVO game development is based upon the principles utilized and validated by Anguera *et al.*¹⁸ in their development of a video game (NeuroRacer). EVO is designed for playing on iOS mobile devices using a consumer game engine (UNITY) with high-level art, music, feedback and storylines to ensure engagement with children and adolescents, whereas NeuroRacer was not developed with such assets or distribution protocols in mind. Similar to NeuroRacer, EVO assesses perceptual discrimination, while single- and multi-tasking (that is, while performing a simultaneous visuomotor tracking task). The perceptual discrimination task requires selective attention in the presence of distraction (with distraction here consisting of the visuomotor tracking component) to correctly respond to specific colored stimuli (tapping anywhere on the screen when a target appears), while ignoring all others, much like a standard 'Go/NoGo' task. Visuomotor tracking involves navigating one's character through a dynamically moving environment with the goal of avoiding the walls and obstacles.

Unlike NeuroRacer, EVO utilizes adaptive algorithms to change game difficulty on a trial-by-trial basis (as opposed to the block-by-block approach used in NeuroRacer) for both the tracking task (adapting the speed of the forward path and sensitivity of the user's motions) and discrimination task (adapting the response window for a target), with real-time feedback making the participant aware of their performance. More specifically, the adaptive algorithm makes proportional changes in game

play difficulty depending upon participants' performance from an 80% accuracy median, an approach comparable to that used in NeuroRacer. Although most cognitive adaptive procedures include simple staircases, EVO employs rapidly adapting algorithms suited for measuring threshold accuracy in a fast-paced environment. The EVO assessments lasted a total of 7 min, during which time participants completed a specified number of both correct and incorrect trials (~100 trials, with an ITI of 1000 ms ± 500) allowing the adaptive algorithm to settle on a prescribed level of difficulty that would forecast discrimination accuracy. The outcome measures acquired from the EVO platform include: a calculated threshold level, mean response time and response time variability to perceptual targets during the single and multi-tasking conditions.

Non-adaptive traditional cognitive assessments

We used a battery of validated neuropsychological cognitive tests integrated into an iOS tablet-format app to assess attention-based cognitive control in a comparable manner to that of EVO: (1) visual search task: participants were presented with an array of either 4 or 12 Landolt squares with an opening on one side until participants located the target (a green box with a gap in the top or bottom) and indicated the location of the gap (top, bottom) by tapping on a box with either 'top' or 'bottom' (Figure 1b). There are two distinct modes: a feature search where red Landolt squares were present in addition to the single green target; and a conjunction search where distractor boxes were green and red, with all green distractor boxes having gaps on either side, whereas red distractor boxes had gaps on the top and bottom similar to the green target box. Each task consisted of 100 trials, with inter-trial intervals jittered between 1200 and 1800 ms in 100 ms increments. Task performance was assessed by examining the mean response time (to correct responses) for each trial type possibility (set size: 4-item trials, 12-item trials; set type: feature search trials, conjunction search trials), with response time cost between target identification for feature and conjunction trials across each set size also determined ((cf. ref. 42); Visual Search cost = set size 12 – set size 4 for feature and conjunction trials separately). (2) Flanker task: based on the original Eriksen and Eriksen task,³⁹ participants responded to the direction of a central arrow, with flanking arrows either having the same (congruent) or different (incongruent) directionality (Figure 1c). The visual search and flanker tasks lasted a total of 7 min each, during which time the player watched an instructional video, played a practice round and completed the assessment. Task performance was assessed by examining response time to correct responses to each trial type, and calculating a cost between these trial types (cf. Lee *et al.*, 2012; Flanker Cost = Incongruent response time (RT) – Congruent RT). See Table 1 for a comparison of tasks.

Statistical analyses

Analysis of covariances (ANCOVAs), covarying for age, were used to test for main effects and interactions between groups and measures, with planned follow-up t-tests (assessing equality of variance using Levene's Test for

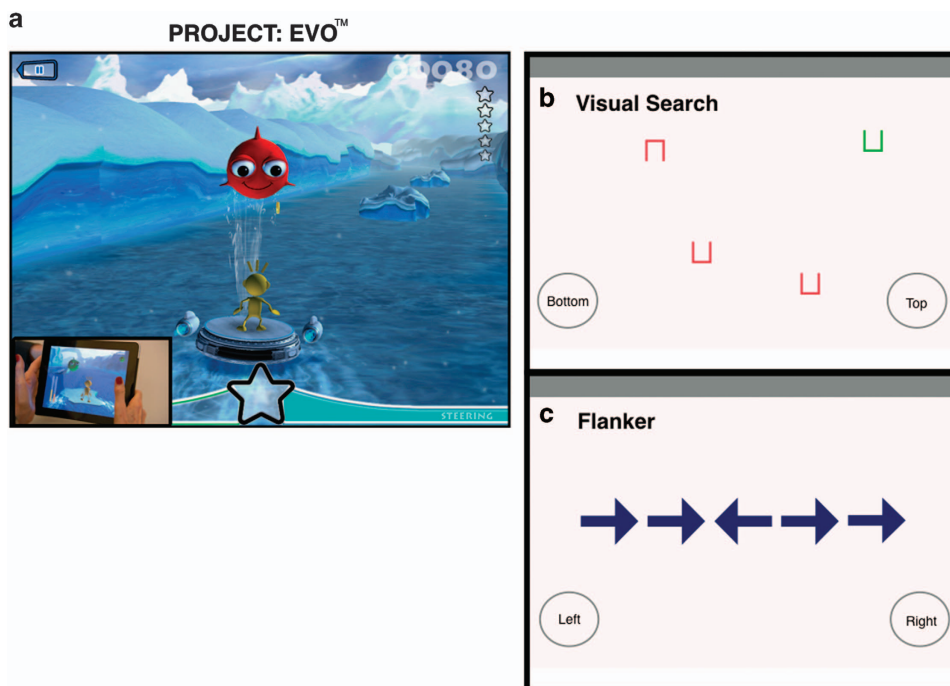


Figure 1. Screen shots of Project: EVO. (a) Image of participant playing Project: EVO showing the individual steering the character, while anticipating the appearance of target stimuli. (b) Image of the visual search task, (c) Image of the flanker task (incongruent trial type).

Equality of Variances) and the Greenhouse–Geisser correction used when appropriate. For the subsample of the participants with IQ measurements (see Table 2 for details by participant cohort), non-verbal IQ was also used as a covariate in a separate analysis (in addition to age) to test for main effects and interactions as above. All effect size values were calculated using Cohen's *d* (ref. 43) and corrected for small sample bias using the Hedges and Olkin approach.⁴⁴

RESULTS

Twenty children with the same 16p11.2 BP4-BP5 deletion, 16 non-carrier siblings, and 75 unaffected unrelated children overlapping in age completed the EVO assessment. Eighteen deletion carrier and 19 non-carrier children (a mix of non-carrier siblings (14) and unrelated children (5)) completed the non-adaptive traditional assessments (the Flanker and Visual Search tasks). Comparison of IQ data from those 48 participants that were part of the larger Simons VIP project revealed both VIQ and NVIQ being significantly lower in deletion carriers than controls (Table 2). The mean carrier results for VIQ (84.3 ± 11.4) and NVIQ (88.5 ± 10.1) were in agreement with other Simons VIP Consortium study groups²⁷ who have access to larger samples ($N = 85$) and documented the average VIQ and NVIQ of deletion carriers to be 79 ± 18 and 86.8 ± 15.1 , respectively.

Diagnostic assessments: EVO Levels

Comparison of children with 16p11.2 deletion with their non-carrier siblings. Game play level for visuomotor tracking and perceptual discrimination during single-task and multi-tasking conditions reflects performance that approaches 80% accuracy. With respect to visuomotor tracking (for example, navigating without hitting walls or obstacles), a 2 (between factor: group) \times 2 (within factor: condition) repeated measures ANCOVA examining the participant level for optimal tracking performance during single- and multi-tasking conditions revealed a group main effect ($F(1,34) = 17.19$, $P < 0.001$, Cohen's $d = 1.36$) but neither an effect of condition ($F(1,34) = 1.26$, $P = 0.27$) nor an interaction ($F(1,34) = 0.74$, $P = 0.40$).

This result suggests that children with 16p11.2 deletion played EVO at a lower visuomotor tracking level than their non-carrier siblings in each condition. Thus, multi-tasking did not differentially impair play for children with the 16p11.2 deletion. A similar analysis assessing the perceptual discrimination level (for example, responding to targets and ignoring distractors) during single- and multi-tasking conditions revealed a group main effect ($F(1,34) = 16.45$, $P < 0.001$, Cohen's $d = 1.03$) but again no effect of condition ($F(1,34) = 0.19$, $P = 0.66$) or interaction ($F(1,34) = 0.91$, $P = 0.35$), suggesting carriers thresholded to a lower discrimination level than their non-carrier siblings. It should be noted that each of the primary effects of interest remain significant when Non-verbal IQ was used as an additional covariate in the ANCOVA analyses ($F(1,18) \geq 4.6$, $P \leq 0.045$), although this analyses should be considered exploratory given the sample size involved (Table 2).

Comparison of children with 16p11.2 deletion with unrelated neurotypical controls. Comparisons between children with 16p11.2 deletion and the unrelated neurotypical children also revealed a group main effect for both visuomotor tracking and perceptual discrimination level ($F(1,92) \geq 33.02$, $P \leq 0.001$, $d \geq 1.44$ in each case), as well as a main effect of condition (single- vs multi-tasking; $F(1,92) \geq 10.10$, $P = 0.002$), but no condition by group interaction ($F(1,92) \leq 3.25$, $P \geq 0.074$ in each case). As above, the primary effects of interest remain significant when Non-verbal IQ was used as an additional covariate in the ANCOVA analyses ($F(1,33) \geq 9.5$, $P \leq 0.004$).

Selective attention: EVO response time and response time variability

Comparison of children with 16p11.2 deletion with their unaffected siblings. In our next set of analyses, we assessed response time and response time variability on the perceptual discrimination task (that is, responding to specific targets and ignoring non-targets). These analyses echoed the level-based findings: response time to targets revealed a main effect of group ($F(1,33) = 15.50$,

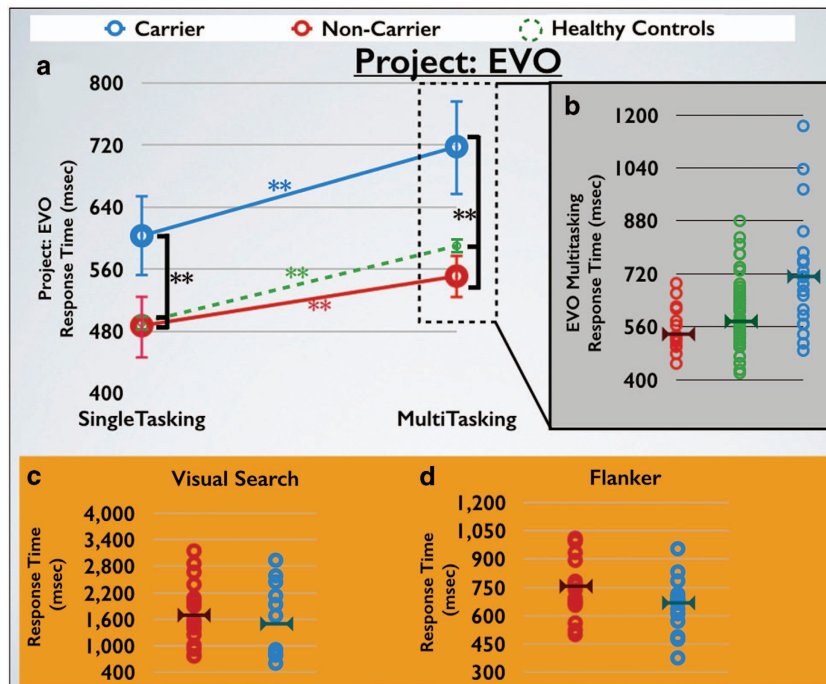


Figure 2. Project: EVO selective attention performance. (a) EVO single- and multi-tasking response time performance for each group (carriers, non-affected siblings and non-affected control groups). (b) EVO multi-tasking RT. (c) Visual search task performance for the conjunction 12 conditions (most difficult). (d) Flanker task performance for the incongruent trial type. Error bars represent s.e., horizontal bars on each plot represent the mean. $**P < 0.01$. RT, response time.

$P < 0.001$, $d = 1.29$), but neither a main effect of condition ($F(1,33) = 0.12$, $P = 0.74$) nor a condition by group interaction ($F(1,33) = 0.98$, $P = 0.33$; Figures 2a and b). The same approach examining response time variability revealed a group main effect ($F(1,33) = 9.08$, $P = 0.005$, $d = 1.01$), but neither a condition ($F(1,33) = 2.50$, $P = 0.12$) nor an interaction ($F(1,33) = 2.21$, $P = 0.15$), suggesting that children with the 16p11.2 deletion have greater performance variability than siblings regardless of condition. However, as expected, each group showed an increase in RT from the single-task to the multi-tasking condition ($P < 0.004$ in each case). As a whole, these results suggest that the deletion carrier group showed slower response times that were magnified and more variable regardless of task complexity relative to their own siblings (There were no group differences with respect to false-positive rate for either condition ($P > 0.30$ in each case)). The inclusion of Non-verbal IQ as an additional covariate in the ANCOVA analyses revealed a very modest trending effect for response time ($F(1,20) = 2.8$, $P = 0.11$), with a significant effect for response time variability ($F(1,20) = 5.5$, $P = 0.029$).

Comparison of children with 16p11.2 deletion with unrelated neurotypical controls. A similar pattern of effects was observed when comparing the deletion carrier children with the unrelated control group for both the response time and response time variability analysis. There was a main effect of group ($F(1,92) \geq 23.86$, $P \leq 0.001$, $d \geq 1.2$ in each case), but no condition by group interaction ($F(1,92) \leq 0.43$, $P \geq 0.51$ in each case). However, a main effect of condition was present in this contrast ($F(1,92) \geq 6.11$, $P \leq 0.015$, in each case), indicative of slower and more variable RT when multi-tasking for all children. This suggests that children with 16p11.2 deletion show differences in RT and RT variability relative to both their siblings and the unrelated neurotypical controls. As above, the primary effects of interest remain significant when Non-verbal IQ was used as an additional covariate in the ANCOVA analyses ($F(1,33) \geq 5.4$, $P \leq 0.027$).

Non-adaptive traditional cognitive assessments: children with 16p11.2 deletion and neurotypical children

We evaluated response time and response time variability performance for the Flanker and Visual Search selective attention tasks on all trial types and task measures between children with 16p11.2 deletion and unaffected children. Across all of these tests and measures, no significant group difference was observed ($F \leq 3.7$, $P \geq 0.064$, $d \leq 0.28$) in each case; see Figures 2 and 3, Table 3 for depiction of these results). Given that there was a trend toward significance for the Flanker response time cost (incongruent RT – congruent RT), further examination of this result revealed that the carrier group actually exhibited an inverse cost (for example, they performed better on the harder condition of the Flanker task), a result that is inconsistent with design and utility of this task across a wealth of literature.³⁰ The inclusion of Non-verbal IQ as an additional covariate in the ANCOVA analyses resulted in the same pattern of effects observed above.

DISCUSSION

The present findings demonstrated that children with 16p11.2 deletion show visuomotor and cognitive control deficits relative to age-matched non-carrier siblings and neurotypical unrelated children when using an adaptive, scientifically inspired digital platform for cognitive assessment. Furthermore, cognitive control deficits were not observed when using traditional non-adaptive assessments (Flanker and Visual Search). These results have two important implications: first, children with 16p11.2 deletion have selective attention deficits that likely affect their learning and real-world function, and second, these deficiencies may be overlooked if using traditional assessments.

The two traditional selective attention assessments used here have consistently shown differences between healthy and attentionally deficient groups,^{45–49} with similar effects seen when

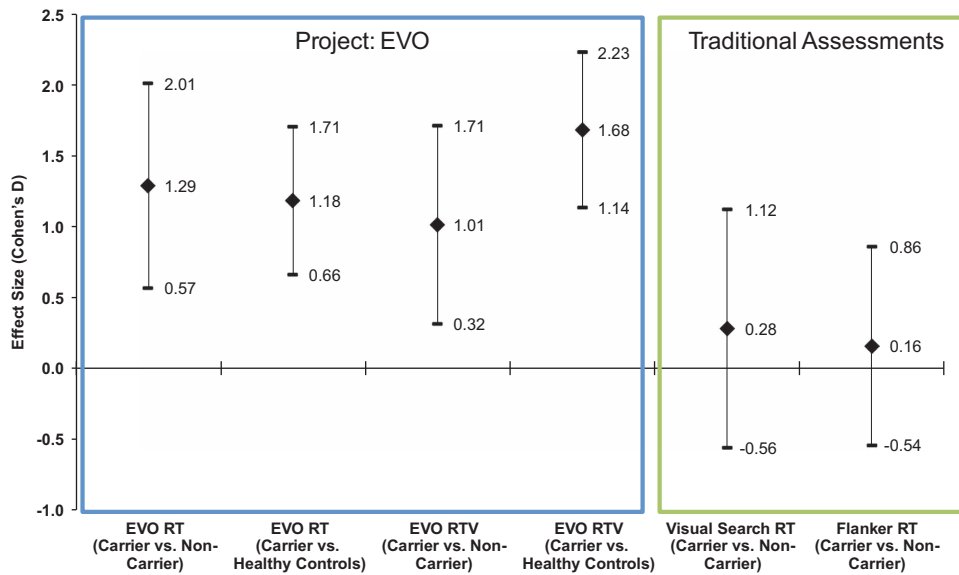


Figure 3. Illustration of effect sizes. Effect sizes (Cohen's *D*) for EVO and iPad assessments are displayed, with these values calculated from repeated measures estimated marginal means for group main effects. RT, response time.

Table 3. Between-group task descriptions

Task	Trial type	Carrier		Non-carrier ^a	
		Group mean RT (s.d.) N = 18		Group mean RT (s.d.) N = 19	
Visual search (ms)	Conjunction: 4-item	1175	(419)	1163	(400)
	Conjunction: 12-item	1517	(863)	1616	(833)
	Feature: 4-item	983	(327)	906.33	(315)
	Feature: 12-item	971	(225)	926.4	(243)
Flanker (ms)	Incongruent	657	(144)	754	(163)
	Congruent	676	(125)	712	(140)
		N = 20		Non-carrier unaffected siblings (N = 16)	Non-related healthy controls (N = 75)
EVO (ms)	Single-tasking RT	603	(144)	487 (126) ^b	491 (80) ^b
	Multi-tasking RT	718	(178)	551 (95) ^b	591 (67) ^b

Abbreviation: RT, response time. ^a5 of these individuals were non-related healthy controls. ^b*P* < 0.05 between-group difference from the carrier group.

deployed on an iPad⁴² or internet browser⁵⁰ (however, see Bauer *et al.*⁵¹ as a point of caution regarding assumptions of validity and reliability when digitally converting testing tools). However, these assessments failed to reveal group differences between the children with 16p11.2 deletion and their siblings. One may question whether the differential results between the adaptive versus non-adaptive assessments reflect variability in the cognitive challenge presented. Indeed, increased cognitive load has been shown to negatively affect performance in children,^{52,53} young adults^{54–57} and older adults.^{56,58,59} However, this does not appear to be the case here: no group differences emerged for the flanker task, nor any for the visual search task which has both a low and high selective attention load. These findings suggest that enjoyable technology that engages a participant with adaptive mechanics can reveal phenotypic differences in highly variable populations.

There is a general consensus that computerized response time measurements can act as a valuable indicator of cognitive ability.^{60–63} Several studies have shown that response time variability is increased in children and older adults compared with younger adults.^{61,64,65} Furthermore, response time variability has been found to distinguish groups of individuals with and without ADHD,^{66–69} as well as individuals on the autism spectrum.³⁸ However, a recent meta-analysis by Karalunas *et al.*³⁸ examining response time variability based on non-adaptive measures suggested these assessments only have small to moderate effect sizes (Hedges' *g* = 0.37, interpreted similar to Cohen's *d*) when attempting to dissociate children with autism or ADHD from neurotypical controls. Here the observed between-group effect sizes for response time performance and variability using the EVO assessment (including the correction for sample size and covarying for age) were quite high (*d* ≥ 0.83). These

findings suggest that the ability to detect group differences between populations that are inherently variable requires tools that demonstrate greater sensitivity. Even in cases for which sensitivity and specificity were found to be comparable between adaptive and non-adaptive platforms,^{70,71} adaptive platforms have the added benefit of requiring less total testing time than 'traditional' computerized testing, and are able to mitigate potential ceiling and floor effects,¹⁷ which is a concern for populations in which inter-individual variability is high.^{72–75} Here the use of adaptive algorithms in concert with entertainment-based video game factors likely contributed to participants being truly engaged in the testing experience. Thus, the observed null between-group differences observed in some studies may be due to a lack of sensitivity in the measurement tools being used as a function of participant engagement.

Parallels can be made between the mechanisms underlying the response time/response time variability effects observed here and similar effects reported in distinct populations. The field of cognitive aging has associated increased response times and variability with neural dedifferentiation (for example, where both structure and function becomes less focal with age^{76,77}) which in turn leads to increased neural noise⁷⁸ (a result also common to children with autism⁷⁹). Recent neuroimaging findings involving children with the 16p11.2 deletion have demonstrated that these individuals have irregularities in their white matter tracts connecting brain regions⁸⁰ that are indicative of improper differentiation. These findings are particularly intriguing given that deficient attentional resources have been associated with reduced frontal–posterior connectivity,¹⁸ with this measure also associated with increased response time variability in both children and older adults.⁸¹ These findings hint at the possibility that the impaired performance observed here stems from the children with the 16p11.2 deletion having less functional neural differentiation than their healthy counterparts at prefrontal regions, negatively affecting the generation of midline frontal theta activity for tasks requiring attentional resources. Future work examining these types of underlying neural correlates within this population would better elucidate the mechanisms associated with the impacted performance observed here.

Given the approach taken here, there are some clear limitations of this study. Of particular concern would be the sample size of the patient population is relatively small, the unrelated control group did not complete the non-adaptive assessments as they were involved in another research study, testing environments were inherently different between these groups, and we do not have a robust characterization of our sample that includes IQ due to the nature of how the data were collected (for example, at a family meeting). These factors must be taken into account when assessing the present findings, as those with higher IQs would be expected to have stronger selective attention abilities, and the observed group differences are subsumed in part by possible general cognitive delays. However, the carrier group is reasonably sized for a rare genetic event and genetically comparable since all the 16p11.2 deletion carriers have the same breakpoints and no other pathogenic copy-number variants. Furthermore, the differential effects observed between the adaptive and non-adaptive tests underscore the idea of using more sophisticated approaches to enhance testing sensitivity. One of the primary functions of the adaptive algorithms being incorporated is to ensure fair comparisons between groups with disparate abilities (in this case those with IQ differences). Although this approach would not remove the need and utility of collecting IQ data, it does provide the means for standardizing performance in situations where IQ would not be available. Another concern involves the identification of group differences in motor function or processing speed, or possible group differences in video game experience leading to the observed effects. While having these measures on each participant would be ideal, adaptivity facilitates having each

participant in a personalized testing state during testing, allowing for participants to have differences in these abilities without contaminating subsequent between-group comparisons (for example, see Anguera *et al.*¹⁸ for a similar approach involving older adults).

Finally, one may question whether EVO is actually testing a distinct type of attention from that measured by the flanker and visual search tasks, or is inadvertently reflecting group differences associated with general familiarity with video games. The common thread across each task is the engagement of selective attention resources, with each task assessing this construct in a distinct manner. Although the EVO platform has a video game presence, the underlying task being performed is at its core a selective attention task, with the adaptivity mechanics directly ensuring that differences in video game experiences are mediated for subsequent group comparisons. Although there was no quantitative evaluation demonstrating that participants found EVO to be more engaging than the other assessments, it is critical to keep in mind that engagement is not the sole driving factor in creating a more sensitive testing environment. Indeed, a more enjoyable experience can lead to enhanced engagement through increased participant motivation; however, heightened testing sensitivity requires the proper incorporation of these factors in addition to the proper titration of adaptive mechanics.¹ Furthermore, there may be situations where an individual with atypical attention would benefit from distinctly different testing environments that are less visually stimulating, providing further evidence that even the approach used here is not a panacea for all future assessment work. It should be noted that these adaptive mechanics are not directly related with making such an experience more enjoyable *per se*; however, 'fun' can make a testing experience more sensitive by encouraging greater participant engagement. In summary, the present findings suggest that children with cognitive control impairments when optimally engaged reveal measurable deficits in response time and response time variability relative to neurotypical controls.

CONFLICT OF INTEREST

JDB and WEM are employees of Akili Interactive Labs, which manufactures the Project: EVO device and supplied the device for use in this study. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank all of our participants and their families whose time and efforts made this work possible. We also thank the Simons VIP team for their help with the data collection. This work was supported by the Simons Foundation, the Wallace Research Foundation, the James Gates Family Foundation and the Kawaja-Holcombe Family Foundation (EJM). We are grateful to all of the families at the participating Simons Variation in Individuals Project (Simons VIP) sites, as well as the Simons VIP Consortium. We appreciate obtaining access to the phenotypic data on SFARI Base. Approved researchers can obtain the Simons VIP population data set described in this study by applying at <https://base.sfari.org>.

REFERENCES

- Goharpey N, Crewther DP, Crewther SG, Eklund LC, Nyman AS. *Intellectual Disability: Beyond Scores IQ*. Nova Science Publishers: Hauppauge, NY, 2009.
- Martinez K, Colom R. Working memory capacity and processing efficiency predict fluid but not crystallized and spatial intelligence: Evidence supporting the neural noise hypothesis. *Pers Indiv Differ* 2009; **46**: 281–286.
- Rose SA, Feldman JF. Prediction of IQ and specific cognitive-abilities at 11 years from infancy measures. *Dev Psychol* 1995; **31**: 685–696.
- Alloway TP. Working memory, but not IQ predicts subsequent learning in children with learning difficulties. *Eur J Psychol Assess* 2009; **25**: 92–98.
- Borella E, Carretti B, Mammarella IC. Do working memory and susceptibility to interference predict individual differences in fluid intelligence? *Eur J Cogn Psychol* 2006; **18**: 51–69.

- 6 Anguera JA, Gazzaley A. Video games, cognitive exercises, and the enhancement of cognitive abilities. *Curr Opin Behav Sci* 2015; **4**: 160–165.
- 7 Walkley R. A compendium of neuropsychological tests - administration, norms and commentary. *Aust Psychol* 1992; **27**: 205.
- 8 Wolf S. A compendium of neuropsychological tests: Administration, norms, and commentary. *Integr Phys Behav Sci* 2000; **35**: 70–71.
- 9 Strauss E, Sherman E, Spreen O. *Compendium of Neuropsychological Tests: Administration Norms, and Commentary*. 3rd edn. Oxford University Press: New York, NY, 2006.
- 10 Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. *Psychol Rev* 2001; **108**: 624–652.
- 11 Coldren JT. Cognitive control predicts academic achievement in kindergarten children. *Mind Brain Educ* 2013; **7**: 40–48.
- 12 Chey J, Kim H. Neural basis for cognitive control enhancement in elderly people. *Alzheimers Dement* 2012; **8**: P1–167.
- 13 Hanson E, Nasir RH, Fong A, Lian A, Hundley R, Shen Y et al. Cognitive and behavioral characterization of 16p11.2 deletion syndrome. *J Dev Behav Pediatr* 2010; **31**: 649–657.
- 14 Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG et al. The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. *Biol Psychiatry* 2014; **77**: 785–793.
- 15 Naismith L, Lonsdale P, Vavoula GN, Sharples M. Mobile technologies and learning. *Futurelab Literature Review Series* 2004; **11**: 1–48.
- 16 Dundar H, Akcayir M. Tablet vs. paper: the effect on learners' reading performance. *Int Electron J Element Educ* 2012; **4**: 441–450.
- 17 Leek MR. Adaptive procedures in psychophysical research. *Percept Psychophys* 2001; **63**: 1279–1292.
- 18 Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowitz J et al. Video game training enhances cognitive control in older adults. *Nature* 2013; **501**: 97–101.
- 19 Lee G. Effects of training using video games on the muscle strength, muscle tone, and activities of daily living of chronic stroke patients. *J Phys Ther Sci* 2013; **25**: 595–597.
- 20 Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *J Neuroeng Rehabil* 2014; **11**: 108.
- 21 Han DH, Renshaw PF, Sim ME, Kim JI, Arenella LS, Lyoo IK. The effect of internet video game play on clinical and extrapyramidal symptoms in patients with schizophrenia. *Schizophr Res* 2008; **103**: 338–340.
- 22 Caglio M, Latini-Corazzini L, D'Agata F, Cauda F, Sacco K, Monteverdi S et al. Video game play changes spatial and verbal memory: rehabilitation of a single case with traumatic brain injury. *Cogn Process* 2009; **10**: S195–S197.
- 23 Ceranoglu TA. Star wars in psychotherapy: video games in the office. *Acad Psychiatry* 2010; **34**: 233–236.
- 24 Holmes EA, James EL, Coode-Bate T, Deerprouse C. Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? A proposal from cognitive science. *PLoS ONE* 2009; **4**: e4153.
- 25 Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 2008; **17**: 628–638.
- 26 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008; **358**: 667–675.
- 27 Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG et al. The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. *Biol Psychiatry* 2015; **77**: 785–793.
- 28 van Meel CS, Heslenfeld DJ, Oosterlaan J, Sergeant JA. Adaptive control deficits in attention-deficit/hyperactivity disorder (ADHD): the role of error processing. *Psychiatry Res* 2007; **151**: 211–220.
- 29 Cao J, Wang SH, Ren YL, Zhang YL, Cai J, Tu WJ et al. Interference control in 6-11 year-old children with and without ADHD: behavioral and ERP study. *Int J Dev Neurosci* 2013; **31**: 342–349.
- 30 Mullane JC, Corkum PV, Klein RM, McLaughlin E. Interference control in children with and without ADHD: a systematic review of flanker and Simon task performance. *Child Neuropsychol* 2009; **15**: 321–342.
- 31 Christ SE, Holt DD, White DA, Green L. Inhibitory control in children with autism spectrum disorder. *J Autism Dev Disord* 2007; **37**: 1155–1165.
- 32 Henderson H, Schwartz C, Mundy P, Burnette C, Sutton S, Zahka N et al. Response monitoring, the error-related negativity, and differences in social behavior in autism. *Brain Cogn* 2006; **61**: 96–109.
- 33 Mason DJ, Humphreys GW, Kent LS. Exploring selective attention in ADHD: visual search through space and time. *J Child Psychol Psych* 2003; **44**: 1158–1176.
- 34 Mullane JC, Klein RM. Visual search by children with and without ADHD. *J Attention Disord* 2008; **12**: 44–53.
- 35 Mason DJ, Humphreys GW, Kent L. Visual search, singleton capture, and the control of attentional set in ADHD. *Cogn Neuropsychol* 2004; **21**: 661–687.
- 36 O'Riordan MA. Superior visual search in adults with autism. *Autism* 2004; **8**: 229–248.
- 37 Plaisted K, O'Riordan M, Baron-Cohen S. Enhanced visual search for a conjunctive target in autism: a research note. *J Child Psychol Psychiatry* 1998; **39**: 777–783.
- 38 Karalunas SL, Geurts HM, Konrad K, Bender S, Nigg JT. Annual research review: Reaction time variability in ADHD and autism spectrum disorders: measurement and mechanisms of a proposed trans-diagnostic phenotype. *J Child Psychol Psychiatry* 2014; **55**: 685–710.
- 39 Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept Psychophys* 1974; **16**: 143–149.
- 40 Manual for the Differential Ability Scales. Harcourt Assessment: San Antonio, TX, 2007.
- 41 Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV). Pearson Assessments: San Antonio, TX, 2003.
- 42 Lee H, Baniqued PL, Cosman J, Mullen S, McAuley E, Severson J et al. Examining cognitive function across the lifespan using a mobile application. *Comput Hum Behav* 2012; **28**: 1934–1946.
- 43 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn, Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
- 44 Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. Academic Press: Orlando, FL, USA, 1985.
- 45 Salthouse TA. Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain Cogn* 2010; **73**: 51–61.
- 46 Fisk AD, Rogers WA. Toward an understanding of age-related memory and visual search effects. *J Exp Psychol Gen* 1991; **120**: 131–149.
- 47 Rogers WA, Fisk AD, Hertzog C. Do ability performance relationships differentiate age and practice effects in visual-search. *J Exp Psychol Learn* 1994; **20**: 710–738.
- 48 Madden DJ. Aging and distraction by highly familiar stimuli during visual-search. *Developmental Psychology* 1983; **19**: 499–507.
- 49 Ball KK, Beard BL, Roenker DL, Miller RL, Griggs DS. Age and visual-search - expanding the useful field of view. *J Opt Soc Am A Opt Image Sci Vis* 1988; **5**: 2210–2219.
- 50 Morrison GE, Simone CM, Ng NF, Hardy JL. Reliability and validity of the Neuro-Cognitive Performance Test, a web-based neuropsychological assessment. *Front Psychol* 2015; **6**: 1652.
- 51 Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Clin Neuropsychol* 2012; **26**: 177–196.
- 52 Gomaruk HK, Wijers AA, Minderaa RB, Althaus M. Do children with ADHD and/or PDD-NOS differ in reactivity of alpha/theta ERD/ERS to manipulations of cognitive load and stimulus relevance? *Clin Neurophysiol* 2009; **120**: 73–79.
- 53 Gomaruk HK, Wijers AA, Minderaa RB, Althaus M. ERP correlates of selective attention and working memory capacities in children with ADHD and/or PDD-NOS. *Clin Neurophysiol* 2009; **120**: 60–72.
- 54 Lavie N. Perceptual load as a necessary condition for selective attention. *J Exp Psychol Hum Percept Perform* 1995; **21**: 451–468.
- 55 Lavie N. Distracted and confused?: selective attention under load. *Trends Cogn Sci* 2005; **9**: 75–82.
- 56 Maylor EA, Lavie N. The influence of perceptual load on age differences in selective attention. *Psychol Aging* 1998; **13**: 563–573.
- 57 Rissman J, Gazzaley A, D'Esposito M. The effect of non-visual working memory load on top-down modulation of visual processing. *Neuropsychologia* 2009; **47**: 1637–1646.
- 58 Clapp WC, Gazzaley A. Distinct mechanisms for the impact of distraction and interruption on working memory in aging. *Neurobiol Aging* 2012; **33**: 134–148.
- 59 Clapp WC, Rubens MT, Gazzaley A. Mechanisms of working memory disruption by external interference. *Cereb Cortex* 2010; **20**: 859–872.
- 60 Barrett P, Eysenck HJ, Lucking S. Reaction-Time and Intelligence - a Replicated Study. *Intelligence* 1986; **10**: 9–40.
- 61 Hulstsch DF, MacDonald SW, Dixon RA. Variability in reaction time performance of younger and older adults. *J Gerontol Ser B Psychol Sci Soc Sci* 2002; **57**: P101–P115.
- 62 Jakobsen LH, Sorensen JM, Rask IK, Jensen BS, Kondrup J. Validation of reaction time as a measure of cognitive function and quality of life in healthy subjects and patients. *Nutrition* 2011; **27**: 561–570.
- 63 Schendel KL, Robertson LC. Using reaction time to assess patients with unilateral neglect and extinction. *J Clin Exp Neuropsychol* 2002; **24**: 941–950.
- 64 Garrett DD, Samanez-Larkin GR, MacDonald SW, Lindenberger U, McIntosh AR, Grady CL. Moment-to-moment brain signal variability: a next frontier in human brain mapping? *Neurosci Biobehav Rev* 2013; **37**: 610–624.
- 65 MacDonald SW, Nyberg L, Backman L. Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends Neurosci* 2006; **29**: 474–480.
- 66 Leth-Steensen C, Elbaz ZK, Douglas VI. Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychol* 2000; **104**: 167–190.

- 67 Bellgrove MA, Hawi Z, Kirley A, Gill M, Robertson IH. Dissecting the attention deficit hyperactivity disorder (ADHD) phenotype: sustained attention, response variability and spatial attentional asymmetries in relation to dopamine transporter (DAT1) genotype. *Neuropsychologia* 2005; **43**: 1847–1857.
- 68 Karalunas SL, Huang-Pollock CL, Nigg JT. Decomposing attention-deficit/hyperactivity disorder (ADHD)-related effects in response speed and variability. *Neuropsychology* 2012; **26**: 684–694.
- 69 Karalunas SL, Huang-Pollock CL, Nigg JT. Is reaction time variability in ADHD mainly at low frequencies? *J Child Psychol Psychiatry* 2013; **54**: 536–544.
- 70 Chien TW, Wu HM, Wang WC, Castillo RV, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. *Health Qual Life Out* 2009; **7**: 39.
- 71 Phipps JA, Zele AJ, Dang T, Vingrys AJ. Fast psychophysical procedures for clinical testing. *Clin Exp Optometry* 2001; **84**: 264–269.
- 72 Berkson G, Baumeister A. Reaction time variability of mental defectives and normals. *Am J Ment Defic* 1967; **72**: 262–266.
- 73 Castellanos FX, Sonuga-Barke EJ, Scheres A, Di Martino A, Hyde C, Walters JR. Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biol Psychiatry* 2005; **57**: 1416–1423.
- 74 Hultsch DF, MacDonald SW, Hunter MA, Levy-Bencheton J, Strauss E. Intraindividual variability in cognitive performance in older adults: comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology* 2000; **14**: 588–598.
- 75 Hetherington CR, Stuss DT, Finlayson MA. Reaction time and variability 5 and 10 years after traumatic brain injury. *Brain Inj* 1996; **10**: 473–486.
- 76 Goh JO. Functional dedifferentiation and altered connectivity in older adults: neural accounts of cognitive aging. *Aging Dis* 2011; **2**: 30–48.
- 77 West R, Murphy KJ, Armilio ML, Craik FI, Stuss DT. Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain Cogn* 2002; **49**: 402–419.
- 78 Voytek B, Kramer MA, Case J, Lepage KQ, Tempesta ZR, Knight RT et al. Age-related changes in 1/f neural electrophysiological noise. *J Neurosci* 2015; **35**: 13257–13265.
- 79 Milne E. Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Front Psychol* 2011; **2**: 51.
- 80 Owen JP, Chang YS, Pojman NJ, Bukshpun P, Wakahiro ML, Marco EJ et al. Aberrant white matter microstructure in children with 16p11.2 deletions. *J Neurosci* 2014; **34**: 6214–6223.
- 81 Papenberg G, Hammerer D, Muller V, Lindenberger U, Li SC. Lower theta inter-trial phase coherence during performance monitoring is related to higher reaction time variability: a lifespan study. *Neuroimage* 2013; **83**: 912–920.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016