A novel framework to measure executive function in Down syndrome, with applications for early clinical diagnosis of dementia

Elizabeth Smith
University of Bristol
Cardiff University

Craig Hedge
Cardiff University

Christopher Jarrold
University of Bristol

Corresponding Author:

Liz Smith, PhD

Mail: liz.smith@bristol.ac.uk

University of Bristol

School of Experimental Psychology

12a Priory Road

Bristol, BS8 1TU

Craig Hedge, PhD,

Cardiff University, CF24 4HQ, UK

Chris Jarrold, Professor

University of Bristol, BS8 1TU, UK

Acknowledgements

This research was funded by the Foundation Jérôme Lejeune.

We would like to thank all participants and their families for supporting this research.

A novel framework to measure executive function in Down syndrome, with applications for early clinical diagnosis of dementia

Abstract

Executive function (EF) decline is a consistent early sign of Alzheimer's disease (AD) among adults with Down syndrome (DS); baseline measures of EF for individuals with DS are vital to allow detection of meaningful decline. We developed a framework to extract measures of three core components of EF (memory updating, inhibitory, and temporal components) within one task. Increases in memory load, inhibitory load, and temporal demands led to significant increases in reaction times and significant decreases in accuracy among eighteen adults with DS and 18 typically developing matched individuals; thus the expected effects of all three manipulations were detected. Good test-retest reliability indicated that this framework has the potential to provide a simple, baseline EF measure for individuals with DS.

Keywords: Down syndrome, Alzheimer's disease, Dementia, Executive function

Down syndrome (DS) is caused by an extra 21st chromosome (either partial or complete). This results in various complications; one of which is a greatly increased risk of developing dementia, specifically early onset Alzheimer's disease (AD) (Cork, 1990; Crayton et al., 1998; Hartley et al., 2015). Indeed in rare cases the onset of dementia can occur exceptionally early among individuals with Down syndrome, with signs of cognitive decline emerging in some individuals in their 30's (Ballard, Mobley, Hardy, Williams & Corbett, 2016; Holland et al., 1998; Prasher & Krishnan, 1993). Accumulation of amyloid-b in the brain appears to be related to early AD related cognitive decline among individuals with Down syndrome (Hartley et al., 2017). The characteristic neuropathological signs of AD, such as neurofibrillary tangles and amyloid plaques, are thought to be present in the brains of all individuals with Down syndrome beyond the age of 40 (Cork, 1990), however, the clinical signs of AD are present in only a proportion of these individuals (Krinsky-McHale et al., 2008; Tyrrell et al., 2001). We currently lack reliable measures of key cognitive processes in Down syndrome (Keeling et al., 2017) which could be used to detect the decline in cognitive/behavioral functioning over time that may signify the need for clinical assessment. Here, we present our initial findings from a novel task developed for this purpose.

Executive function as a potential window to dementia related cognitive decline in Down syndrome

The pre-existing cognitive difficulties associated with Down syndrome (e.g., Baddeley & Jarrold, 2007; Chapman 1997; Martin, Klusek, Estigarribia & Roberts, 2009) pose particular challenges to the diagnosis of AD in the population with Down syndrome (Stanton & Coetzee, 2004). At present, the diagnostic tools that clinicians use to assess AD symptomatology among individuals with intellectual difficulties are the same as those used among typically developing populations (e.g., DSM-5, APA, 2013; and the ICD-10, WHO, 2011). However, the appropriateness of these tools to assess individuals with learning difficulties is debateable (see

Strydom, Livingston, King & Hassiotis, 2007; Westphal, 2013), as self-report of decline may be difficult for individuals with learning difficulties to communicate (Roberts, Price, & Malkin, 2007).

In addition, AD presentation appears atypical in the early stages among individuals with Down syndrome; the earliest stage of AD in those with Down syndrome tends to be characterised by decline in frontal lobe functions (Ball et al., 2006; Ball et al., 2008; Fonseca, Yokomizo, Bottino & Fuentes, 2016), with Das et al. (1995) suggesting that there may be particularly strong links between planning and attention problems and the presence of dementia in the population with Down syndrome. A recent review carried out by Lautarescu, Holland and Zaman (2017) also strongly supports the notion that decline in executive function (i.e., frontal lobe functions) is a key early indicator of dementia in adults with Down syndrome. It is therefore important that deterioration in executive function is detected among adults with Down syndrome. However, due to large variability in cognitive function from one individual to the next in the population with Down syndrome (Nieuwenhuis-Mark, 2009), detecting deterioration in performance (e.g., in executive function tasks) among individuals with Down syndrome is problematic. For example, Carr (1988) gave the Leiter International Performance Scale (Leiter, 1980) to a sample of 21-year-olds with Down syndrome, and found IQs ranging from 8-67 (mean = 42); this translates to a mental age range of 1-8 years. Thus, it is difficult to compare an individual's cognitive or behavioral performance (e.g., executive functions) to that which one would expect for an individual with Down syndrome in the 'normal range' on a given task, as this expected level could be very different from one individual with Down syndrome to another. Given the consistent evidence that executive function is a key area of decline in the early stages of AD it is essential that reliable baseline performance measures of executive function are developed for individuals with Down syndrome to allow for accurate measurement of any cognitive decline.

Components of executive function

There is an array of components under the umbrella term of executive/frontal functions, and a number of different tasks are used to test these different functions in the literature. Existing research measuring executive function among individuals with Down syndrome tends to involve multiple tasks to tap various different functions (Costanzo et al., 2013; Lanfranchi et al., 2010; Pennington et al., 2003; Will, Fidler, Daunhaurer & Gerlach-McDonald, 2017), such as separate tasks to assess working memory (e.g., backward digit span), inhibition (e.g., Stroop-type tasks, Stroop, 1935; and go/no-go tasks), planning tasks (e.g., tower of London type tasks, Shallice, 1982), shifting (e.g., rule shifting such as card sorting tasks, Nelson, 1976), and verbal fluency (e.g., saying as many words as possible starting with a particular letter, Newcombe, 1969). However, measuring separate core functions with separate tasks can be very time consuming. Alternatively, parent and caregiver report measures have also been used to assess executive functions among individuals with Down syndrome, specifically the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy & Kenworthy, 2000), and its preschool version (BRIEF-P; Gioia, Espy & Isquith, 2003) (e.g., Daunhauer et al. 2014; Lee et al., 2011; Loveall, Conners, Tungate, Hahn & Osso, 2017).

The findings regarding executive function in the Down syndrome population have been mixed, with many studies suggesting impairment in executive functioning relative to individuals' other mental abilities (Borella, Carretti, & Lanfranchi, 2013; Lanfranchi et al., 2010; Reid et al., 2017; Rowe, Lavender & Turk, 2006), but some indicating that executive function is not impaired relative to mental age (Pennington et al., 2003). However, a profile of strengths and weaknesses in executive function has been suggested to be associated with Down syndrome, with areas of executive function such as inhibition often suggested to be less impaired, relative to other aspects such as shifting (Costanzo et al., 2013; Lee et al., 2015; Loveall et al., 2017), and working memory and planning (Daunhauer et al., 2014).

Existing executive function tasks used with adults with Down syndrome

Using separate tasks is not ideal when it comes to drawing comparisons across the different components of executive function. As noted by Miyake (2000): 'Any executive task strongly implicates other cognitive processes that are not directly relevant to the target executive function' (p. 52). The different types of cognitive demands imposed by different tasks may lead to better or worse performance among individuals with Down syndrome. For example, specific difficulties have been reported in verbally mediated shifting tasks (Landry et al., 2012) and some studies find impairment on verbal but not visual inhibition tasks (Costanzo et al., 2013; Cornish, Scerif & Karmiloff-Smith, 2007); these patterns could be related to the specific verbal difficulties associated with Down syndrome (e.g., Brock & Jarrold, 2005; Jarrold, Baddeley & Phillips, 1999) rather than reflecting fundamental executive problems.

Existing measures often lack reliability, can be insensitive to change among individuals with Down syndrome, and may not be suitable for repeated assessment (See Edgin et al., 2017; Esbensen et al., 2017, & Keeling et al., 2017 for overviews). Esbensen et al. (2017) suggested that the BRIEF (parent/caregiver report measure) has adequate test-retest reliability among individuals with Down syndrome (see also D'Ardhuy et al., 2015). However, given the potentially subjective nature of third party report, it is important that report-based measures are supplemented by adequate direct tests of behavior.

A novel measure of executive function for adults with Down syndrome

Perhaps for the reasons outlined above, many of the commonly used tasks that are each assumed to reflect a particular component of executive functioning do not tend to correlate well with one another (Hedge, Powell & Sumner, 2018; Rabbit, 1997). A first aim of the current study was to take a novel approach and to extract measures of core components of executive function within one simple framework; a subsidiary objective was to assess whether these three

measures were independent. A second main aim was to then assess the reliability of this framework among the Down syndrome population, as it is important that any task used as a regular, repeated baseline assessment is reliable.

The ability to extract multiple independent measures within the same single task allows for controlled comparisons across these different key functions, e.g., participants' level of attention and motivation are the same across the measures. The task format and demands are also controlled across the different executive function measures. In addition, the use of a single task to simultaneously extract multiple components is efficient; for example, extracting one level of each measure on every trial would substantially reduce the time required to measure multiple components of executive function. This is particularly important given the motivational difficulties reported among the population with Down syndrome (Fidler, Philofsky, Hepburn & Rogers, 2005; Ruskin, Mundy, Kasari & Sigman, 1994; Wishart, 2001).

In a theoretical context, the essence of executive control is the ability to maintain task goals in order to prevent inappropriate responses to external stimuli (e.g., Norman & Shallice, 1986); this statement provides a clear *a priori* model of the interplay between working memory and prepotent response inhibition. More recently, theorists have begun to distinguish typically developing children's reactive and proactive control: the former reflecting control functions that follow the onset of a change in the environment, the latter reflecting those that anticipate such a change (Bravier, Gray & Burgess, 2007; Chatham, Frank & Munakata, 2009; Chevalier, Huber, Wiebe & Espy, 2013). In a task switching paradigm, participants show faster reaction times when they are pre-cued to the stimulus dimension that will be relevant on the upcoming trial (see Chevalier et al., 2014; & Monsell, 2003). Note that this is not reducible to the influence of basic speed of processing, rather it depends on higher-level executive factors to engage in preparation for the upcoming task. We therefore propose a theoretically driven framework in which we have an *a priori* starting point, focussing on working memory and

inhibition, but in addition accommodate the distinction between reactive vs. proactive control by manipulating temporal parameters. We therefore explored the interplay between: i) working memory, and ii) response inhibition, which iii) unfolds over time due to the need to exert executive control in response preparation.

Summary

To summarize, the aims of this study are first to assess whether the expected effects of all three previously outlined measures of executive function are found when assessed simultaneously using a single task, with a subsidiary objective of assessing whether these effects are independent of one another. Second, the study aims to determine whether each of the three measures are reliable among individuals with Down syndrome. This proof of concept study is an important first step. If found to be valid and reliable, implying that practice effects do not confound performance, this novel task has the potential to provide a simple, efficient approach to measuring baseline executive function among individuals with Down syndrome, and could be incorporated into annual health checks.

Method

Participants:

Participants were 18 individuals with Down syndrome (mean age = 30.3 years, *SD* = 6.9 years, range = 20-49 years, 5 females), and 34 typically developing (TD) individuals (mean age = 7.6 years, *SD* = 2.0 years, range = 5-10 years, 21 females). The 18 individuals with Down syndrome were recruited in XXXX, UK, and the surrounding area. Four of the individuals with Down syndrome lived at home with parent/s, 9 individuals lived in shared supported accommodation, 3 individuals lived in their own home/flat, and for 2 individuals their living situation was not reported. None of the individuals reported a diagnosis of AD. The 34 TD children were recruited from schools in XXXX, UK, and participated in a quiet room at their

school. The 18 individuals with DS and the 34 TD individuals each completed the children's version of the Raven's Coloured Progressive Matrices (Raven, Raven & Court, 1998) to assess their nonverbal mental ability, and completed the main task within the same session for practicality. Prior to analyses of the data, 18 of the TD children were matched to the individuals with Down syndrome on an individual basis for nonverbal mental ability raw scores (DS mean raw score: 19.61, (*SD* = 5.54, raw score range = 9-30) TD mean raw score: 19.78 (*SD* = 6.08, raw score range = 7-30), $t = -.086$, $p = .93$)), (Nonverbal mental age matched TD sample: mean chronological age = 7.1 years, *SD* = 1.4 years, range = 5-10 years, 11 females). The analyses were focussed on the matched groups, however the data set for the entire sample of 34 TD children is provided online via the Open Science Framework (osf.io/z3gxp).

Design

Our study was designed to address two primary aims. First, we sought to verify whether our task could be used to successfully extract measures of the three components of executive functioning described above. We administered the task to individuals with Down syndrome and typically developing individuals and assessed accuracy and reaction times to determine whether the effects of each manipulation were found. Second, we sought to determine the reliability of this framework for use among the population with Down syndrome. All individuals with Down syndrome were re-tested approximately 4 weeks following their initial session.

Executive functioning task:

The executive function task was a new task designed and programmed by the authors. The task was presented on a tablet device (Microsoft Surface Pro; 10.6 inch screen). The general structure of the task was as follows. On each trial a single coloured alien appeared on screen. The participant's task was to assign the alien to one of two 'spaceships', presented on the left and right of the screen, according to a rule that they were given at the start of each

block. The participant was instructed to respond as quickly and accurately as possible by touching one of two response buttons (one button corresponding to the right spaceship, and one corresponding to the left spaceship).

The task was divided into 6 blocks, each containing 4 practice trials followed by 24 experimental trials. Participants were given 4000ms within which to respond on each trial; if the participant did not respond within this time frame then the trial was skipped. Upon a participant pressing their response they moved onto an inter-trial screen displaying a large 'ready' button and the two squares outlining where the response buttons were located (see Figure 1); the large size of the ready button was to aid accessibility. After pressing the ready button the next trial was presented after 2000ms; participants were instructed to place their index fingers over each of the placeholders (white squares) so that they were ready to respond to the next trial. At the start of each block the participant was introduced to the memory load rule for that block. The instructions were shown to the participant pictorially (as shown in Figure 1). The blocks were completed in a counterbalanced order.

On each trial participants received accuracy feedback. Feedback was also given at the end of each block. To motivate the participants, they received 'slime' to feed to the aliens at the end of the block. Completing the task took approximately 20-25 minutes.

To manipulate the different components of executive functioning, we manipulated the complexity of the response rule that had to be maintained in each block (memory), the location of the alien stimulus on each trial (inhibition), and the time window available for the response on each trial (reactive vs. proactive control), as described in detail below. One level of each of the three variables was extracted on each trial. See Figure 1 for an overview of the task progression, and Figure 2 for an illustration of each manipulation.

[*Insert Figure 1 here*]

Memory Load

We varied the memory load across blocks by manipulating the complexity of the rule by which participants assigned the alien stimuli to the correct spaceship across three levels (no memory load, low memory load, high memory load). Participants were informed of the rule at the start of each block. In the no memory load blocks, the participant was simply required to visually match the stimulus to the spaceship of the same colour (blue alien = blue spaceship, green alien = green spaceship). In the low memory load blocks, the participant was required to remember a one-to-one association between the stimuli and response options based on non-matching colours (pink alien = blue spaceship, orange alien = green spaceship). In the high memory load blocks, the participant was required to remember a two-to-one association (yellow OR grey alien = blue spaceship, red OR brown alien = green spaceship). Each stimulus (alien colour) was presented an equal number of times within each block (e.g., no memory load block: 50% blue stimuli, 50% green stimuli).

Inhibition

Inhibition was simultaneously but independently manipulated by varying the presentation location of the stimulus (left or right) on a trial by trial basis. This inhibition manipulation was based on the Simon task (e.g., Simon, 1969); the Simon task rests on the assumption that the spatial location of the stimulus automatically activates the corresponding response option. For example, a stimulus presented on the left side of the screen would automatically activate the left-hand response option. Participants are typically slower and less accurate on trials where the stimulus is presented on the opposite side to the correct response option (incongruent trials) compared to when the stimulus is presented on the same side as the correct response option (congruent trials) (see Simon, 1969) We implemented this by presenting the alien on either the left or right of the screen on each trial in a pseudo-random order. Across the block as a whole, half of the trials involved the stimuli appearing next to the

correct response button (congruent), and half involved the stimuli appearing next to the incorrect button (incongruent).

Reactive and proactive control.

We manipulated the extent to which participants were able to exert proactive control by varying the delay between stimulus onset and the appearance of the response buttons within each block. The response buttons were the two spaceship buttons. Two white squares on either side of the screen functioned as placeholders to indicate where the spaceship buttons would show up. These placeholders were on the screen constantly until replaced by the spaceship buttons (see Figure 1). For half of the trials in each block there was a delay of 1000ms between the onset of the stimuli and the appearance of the response buttons, and for the other half of trials the response buttons were available immediately; trials of each type were interleaved and presented in a pseudo-random order to prevent participants anticipating the format of the next trial. The logic of this manipulation followed from findings from task switching studies, where individuals show reduced performance costs when they receive a cue before each trial to alert them to the relevant stimulus feature (Monsell, 2003). Here, we assumed that when a participant was immediately presented with the response options, that they would respond more 'reactively'. By delaying the onset of the response options, participants have the opportunity to use the time to prepare their response set, which would lead to faster reaction times (see e.g., Chevalier, 2015).

[*Insert Figure 2 here*]

Practice trials

Prior to the main trials of the executive function task all participants were shown how to play the game on the tablet and completed practice trials; there were 24 initial practice trials (8 for each memory load set). In the practice trials all stimuli were presented in the centre of the screen to allow participants to get used to the task and the memory rules without having to

attend to the other manipulations. Participants were subsequently shown that the stimuli (aliens) would sometimes pop up closer to the green spaceship button or closer to the blue spaceship button; they were told to ignore the location of the stimuli and instead always follow the memory rules. Participants were also shown examples of proactive and reactive trials and were told that they were able to respond as soon as the spaceship buttons appeared. Any participants struggling on the practice trials were given the opportunity to repeat these practice trials. However, if participants were unable to grasp the task after more than three rounds of the practice trials, or if they were unable to attend to the task they were excluded from the main experiment (n = 3; 2 DS, 1 TD). In addition, the first four trials of each block (6 blocks) were also practice trials; on these trials stimuli were again presented centrally to ensure that the participant was maintaining the given memory rule, prior to the introduction of the congruency manipulation.

Data Analyses:

Accuracy cut off:

We first assessed accuracy at the level of each of the six blocks. If the average accuracy within a block was below 60% across all experimental trials, we discarded those trials from our analysis. The accuracy cut off was used to exclude blocks in which we assumed that participants were not attending to the task, and were guessing (i.e., responding randomly); guessing would be indicated by scores close to chance. It is common practice in cognitive control studies to apply an accuracy cut off for this reason (e.g., De Simoni & von Bastian, 2018; Kane, Poole, Tuholski & Engle, 2006; Rey-Mermet, Gade & Oberauer, 2018; Zwaan et al., 2017). The 60% accuracy cut off in the current study resulted in four participants in the group with DS and two participants in the TD group missing data from either the low or high memory load conditions. However, participants' data remained in the analysis for the blocks

in which they were at or above the accuracy cut off; this was to minimize the loss of data given the modest sample size.

Analysis:

Data were analysed with linear mixed effects models using the lme4 (Bates, Maechler, Bolker & Walker, 2015) package in R (R core team, 2016). To obtain p-values for main effects and interactions analogous to traditional repeated measures ANOVA, denominator degrees of freedom were calculated using Satterthwaite's approximation implemented in the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2016). Test-retest reliability was assessed separately for the effects of inhibition, memory load and delay in the group of individuals with Down syndrome. Reliability was calculated using intraclass correlation coefficients (ICC) using a two-way random effects model for absolute agreement.

Results

Descriptive statistics for the Down syndrome group (N = 18) and the subset of matched typically developing children (N = 18) are shown in Table 1. These descriptive data were calculated after excluding blocks in which accuracy was below 60%, in line with the main analyses.

[Insert Table 1 here]

Individuals with Down syndrome were compared to the subset of mental age matched TD children to explore whether those with Down syndrome were affected by the three manipulations as expected, and to determine how their level of performance compared to that of this matched TD group. Analyses are presented first for reaction times, and second for accuracy, with the same analysis carried out for both dependent measures.

Reaction Time:

A linear mixed effects model was carried out with three within subjects factors (3 x 2 x 2): memory load (No memory load, low memory load, high memory load), temporal load (Immediate vs. proactive), and inhibitory load (Low load: congruent vs high load: Incongruent), and one between subjects factor: population (Down syndrome vs. typically developing).

The main effect of population was not significant, $F(1, 33) = 0.75$, $p = .39$. The main effect of memory load was significant, $F(1, 351) = 30.99$, $p < .001$, with post-hoc comparison showing that the difference between no load (M = 1114ms, SD = 303ms) and low load (M = 1233ms, SD = 272ms) conditions was significant ($t = -5.03$, $p < .001$), as was the difference between no load and high load conditions (M = 1341ms, SD = 335ms), ($t = -7.7$, $p < .001$), and low load and high load conditions ($t = -2.72$, $p = .01$). The main effect of temporal load was also significant, $F(1, 349) = 570.78$, $p < .001$, reflecting faster reaction times in the delay condition (M = 881ms, SD = 242ms) relative to the no delay condition (M = 1432ms, SD = 260ms). The final main effect of inhibitory load was marginally significant, $F(1, 349) = 3.02$, $p = .08$, reflecting faster reaction times on the congruent trials (M = 1136ms, SD = 247ms) relative to the incongruent trials (M = 1177ms, SD = 247ms). There was a significant interaction of memory load x temporal load, $F(2, 349) = 4.78$, $p = .01$, with post-hoc comparison showing that the effect of temporal load (the difference between RT's for immediate vs. delayed responses) increased under greater memory loads (temporal effect under no memory load: mean difference = 490ms, $t(35) = 11.88$, $p < .001$; low memory load: mean difference = 595ms, $t(35) = 13.99$, $p < .001$; high memory load: mean difference = 675ms, $t(35) = 15.42$, $p < .001$) as can be seen in figure 3; In other words the difference between no load, low load, and high load RT's was more pronounced in the immediate (no delay) condition ($F(2, 158) = 28.29$, $p < .001$) relative to the delay condition ($F(2, 158) = 5.90$, $p = .003$). All remaining interactions

were non-significant, with all F values < 2.48, and all p values > .10; this included all interactions with population.

[Insert Figure 3 here]

Accuracy:

The same linear mixed effects model as used in the reaction time analysis was also carried out with the accuracy data. To be consistent with the reaction time data this analysis also applied the 60% accuracy cut off. Thus, the overall accuracy levels in this analysis are slightly overestimated.

The main effect of population was not significant, $F(1, 33) < .01$, $p = .96$. The main effect of memory load was significant, $F(1, 354) = 60.31$, $p < .001$, with post-hoc comparisons showing that the difference between no load ($M = 96.0$, $SD = 7.0$) and low load conditions ($M = 87.4$, $SD = 11.2$) was significant ($t = 6.11$, $p < .001$), as was the difference between no load and high load conditions ($M = 80.1$, $SD = 13.3$), ($t = 10.92$, $p < .001$), and low load and high load conditions ($t = 4.83$, $p < .001$). The main effect of temporal load was also significant, $F(1, 349) = 5.06$, $p = .03$, reflecting higher accuracy in the delay condition ($M = 84.5$, $SD = 14.5$) relative to the no delay condition ($M = 82.0$, $SD = 15.6$). The final main effect of inhibitory load was marginally significant, $F(1, 349) = 3.45$, $p = .06$, reflecting higher accuracy on congruent trials ($M = 84.1$, $SD = 13.6$) relative to the incongruent trials ($M = 82.3$, $SD = 16.6$). There were no significant interactions among these factors for accuracy; all F values < 2.06, all p values > .13, again including all interactions involving population.

Test-retest reliability:

Typical interpretations of ICC values are: excellent (.8), good/substantial (.6), and moderate (.4) levels of reliability (Cicchetti & Sparrow, 1981; Fleiss, 1981; Landis & Koch,

1977). Negative ICCs are reported as zero (Giraudeau, 1996). In our analysis the participants' 'costs' of each manipulation at time 1 were correlated with their costs of that manipulation at time 2. The memory load cost referred to the difference between participants' overall performance in the high memory load condition relative to the low memory load condition, as we were primarily interested in the cost of high vs. low memory load, rather than the presence vs. absence of memory load. The temporal cost was the difference between participants' overall performance in the no delay condition relative to overall performance in the delay condition. Finally, the inhibition cost was the difference between participants' overall performance in the incongruent condition relative to their overall performance in the congruent condition. Costs were calculated for each of these three factors, for participants' reaction times and accuracy.

Reaction times:

The overall RT inhibitory cost had excellent test-retest reliability from time 1 to time 2 (single measures ICC = .688, average measures ICC = .815, p = .001). For the overall RT memory load cost there were five participants excluded from the correlation due to missing data at either time 1 or time 2 for either the low or high memory load conditions. However, the memory cost also had good test-retest reliability from time 1 to time 2 (single measures ICC = .438, average measures ICC = .609, p = .059). The RT temporal cost had slightly poorer test-retest reliability from time 1 to time 2 (single measures ICC = .240, average measures ICC = .387, p = .149). The correlations between time 1 and time 2 for each factor are shown diagrammatically in Figure 4.

[Insert Figure 4 here]

Accuracy:

The overall accuracy inhibition cost was associated with moderate test-retest reliability from time 1 to time 2 (single measures ICC = .306, average measures ICC = .469, p = .090). In contrast, the overall accuracy memory load cost had poor test-retest reliability from time 1 to time 2 (single measures ICC= .025, average measures ICC = .049, p = .469). Finally, the accuracy temporal cost had poor test-retest reliability from time 1 to time 2 (single measures ICC = 0, average measures ICC = 0, p = .664). The correlations between time 1 and time 2 are shown in Figure 5.

[Insert Figure 5 here]

Reliability of Raven's Coloured Progressive Matrices:

The Raven's Coloured Progressive Matrices was found to have excellent test-retest reliability among the adults with Down syndrome (single measures ICC = .811, average measures ICC = .895, p < .001).

Discussion

The first aim of this study was to test the feasibility of using a single task to extract measures of three components of executive function, with adults who have Down syndrome and a TD comparison group, with a subsidiary objective to assess whether the three components were independent. A second aim was to assess the reliability of this task for the group of adults with Down syndrome, to test whether such a task would have potential to be further developed as a baseline measure for this population in future. Both of these aims were met; the single task was effective in extracting each of the three components in both populations, and the reasonable reliability of the task among adults with Down syndrome indicates that the task shows promise as a potential novel baseline measure of executive function for this population.

Suitability of the novel executive function task for individuals with Down syndrome

For the above aims to be realised it was essential that individuals with Down syndrome in the target age range were able to complete this task. The developmental level of the group with Down syndrome was between age 5-10 years (based on their nonverbal mental age scores), and these participants had relatively few missing cells of data (i.e., below the accuracy cut off). There were four individuals who were below the 60% accuracy cut off for either the low or high memory load conditions, and we interpreted these individuals as not attending to the task rules and simply guessing/responding randomly in these blocks. Therefore, some individuals with Down syndrome may have difficulty with some aspects of the task. However, as there are three levels of memory load it would be possible to use the no memory load performance in the memory load cost calculation (rather than high memory load minus low memory load) for these individuals. Indeed the overall levels of accuracy were very high. There were two participants that were excluded as they were unable to do the task after the practice trials. We do not have mental age information for these individuals. However, for individuals with a mental age equivalent of 5-10 years this task tends to be appropriate. It is common for adults with Down syndrome to have a mental age within this range (see e.g., Carr, 2000; Vakil & Lifshitz-Zehavi, 2012).

Executive functions in Down syndrome and mental age matched controls

Overall, the difference in performance, either in terms of task accuracy or reaction times, between those with Down syndrome and those without Down syndrome matched for nonverbal mental ability did not reach significance. Thus, the executive functions measured here (working memory, inhibition and reactive vs. proactive control) were not significantly impaired relative to nonverbal mental ability in individuals with Down syndrome in this age range; this is in line with the findings of Pennington et al. (2003). However, this finding is in contrast to various other studies that have reported impairment on tasks of executive function among individuals with Down syndrome relative to their mental age (Costanzo et al., 2013;

Lanfranchi et al., 2004; 2009; 2010; Rowe et al., 2006). Note that our aim was to develop a task that can be used to track executive functioning over time, not to detect group differences. For this, it was critical that individuals with Down syndrome could perform the task adequately at baseline. However, it is worth considering whether the absence of significant group differences is surprising.

One possible reason for conflicting evidence in previous research is the employment of different tasks and matching procedures across studies. For example, in the study by Rowe et al. (2006), participants were matched for age and vocabulary, and showed differences in a number of tests including Raven's progressive matrices (the matching measure used in the current study). Thus, in the current study, particularly given the potential association between measures of fluid intelligence such as the Ravens matrices and executive functioning (Duncan, Burgess & Emslie, 1995; Friedman et al., 2006), it could be argued that the matching measure (Ravens matrices) plays a role in the close similarities in performance in executive function in the Down syndrome group and the TD group. However, Friedman et al. (2006) found that inhibition and shifting, in contrast to memory updating, were very weakly related to traditional measures of fluid intelligence (here including Raven's matrices and block design). It is important to keep in mind that performance on measures such as vocabulary may be more influenced by continued experience into adulthood, compared to fluid abilities.

There were also no significant interactions between any of the within-subjects manipulation factors and population in the current study, for either reaction times or accuracy data, indicating that both groups were affected by these manipulations in a similar way. The group with Down syndrome did tend to have slightly slower overall reaction times relative to the matched TD group, as well as slightly lower accuracy overall, but these differences did not approach significance. Task modality may have also played a role in the performance of the individuals with Down syndrome in the current study. It is perhaps worth noting that

impairment in domains such as visuo-spatial working memory relative to mental age has been found in previous research but only when the task was highly demanding (e.g., Lanfranchi, Carretti, Spanò & Cornoldi, 2009; Lanfranchi, Cornoldi & Vianello, 2004), for example when items were presented simultaneously (more demanding) rather than one at a time. This is in contrast to verbal working memory where the same group of individuals with Down syndrome were impaired regardless of the task demands (Lanfranchi et al., 2009; 2004). We designed our task to limit verbal demands; and our findings indicate that this was effective, making the task well suited for use with individuals with Down syndrome.

Measurement and independence of executive functions

Increases in memory load, inhibitory load, and temporal demands, did lead to significant increases in RT and significant decreases in accuracy for both populations. Thus, the expected effects of each of these manipulations can be detected with this simple experimental framework, both in TD individuals, as well as among individuals with Down syndrome. It should be acknowledged that the inhibitory load effect was marginal, however the effect of inhibition was in the expected direction for both populations. The reliability analysis highlighted that there were large individual differences in the inhibition effect for the individuals with Down syndrome, that may have reduced the overall group level effect. Nonetheless, the size of this inhibition effect is a limitation of the current framework, and it may be appropriate to adapt this aspect of the design in future research to tap into larger congruency effects across individuals.

The temporal delay led to a very large decrease in RT's and large increase in accuracy, indicating that both populations were able to prepare for their subsequent responses when given a one second delay. The one second delay does not force participants to use proactive control, in that the participant does not have to use the available time to be proactive and anticipate their response. For example, in studies using the cued task switching paradigm, when children

are given a cue that a task switch is about to happen they have the opportunity to be proactive and prepare for the switch, however, young children (e.g., age 3 or 4) do not tend to be proactive in these circumstances. Instead they rely on reactive control, while older children (age 7+) show the expected tendency to engage proactive control (Chatham, Frank, & Munakata, 2009; Chevalier et al., 2014; Chevalier, Martis, Curran & Munakata, 2015; Killikelly & Szűcs, 2013). In the current study participants' preparation to respond prior to the availability of the response buttons appears to indicate the engagement of proactive control, as would broadly be expected based on the developmental level (5-10 years) of the individuals in this study.

The effect of memory load was also large, with both populations responding more slowly and less accurately with increasing memory load. In turn the RT benefit in the delay condition relative to the immediate condition was greater with increasing memory load, as highlighted by the significant memory load x temporal load interaction. Our interpretation of this interaction is that participants' use of the delay period (which we attribute to proactive control) absorbs the extra processing time required for higher memory loads. Given that the memory rule and the location of the response buttons remain consistent in a block in our framework, the preparation time can be used to select the appropriate response, and therefore respond faster when the response buttons subsequently become available. When the rules place greater demands on memory, participants may require more response preparation, i.e., thinking through the more complicated response mappings. It is therefore possible that in the delay condition, participants' reliance on the additional time to prepare their response in the low or high memory load conditions makes the difference in no, low and high memory load performance less pronounced (though note the effect of memory load was still highly significant in both the immediate and the delay condition). We should note that processing speed is a key element in the extent to which participants might benefit from the delay period in the different load conditions, in that we assume that at least some participants are able to

process the stimulus and prepare a response in less than 1 second (the duration of the delay) on some trials. However, it is the spontaneous choice to engage in processing during this delay period that we attribute to the engagement of proactive control.

Thus, the temporal manipulation was not independent from the memory load manipulation. In contrast, there were no significant interactions involving inhibitory load, hence this appears to be an independent component of executive function. The notion of independence of different executive function components fits with the findings of Miyake et al. (2000) and may reflect specific genetic influences unique to the separate components of executive function (Friedman et al., 2008).

Importantly, a crucial finding in the current study, was that the tested framework appeared to capture relatively reliable individual differences among individuals with Down syndrome for each measure extracted in terms of reaction times and accuracy (as shown in Figures 4 & 5). The reliability for each component in the current study was greater for reaction times relative to accuracy. One possibility is that RT's allow greater sensitivity because they are continuous, however RT's may also be subject to greater noise. Another possibility is that the greater reliability for RT's relative to accuracy may in part be due to the accuracy cut off, where scores were excluded if they fell below 60%. With two buttons to choose from there is a floor of 50% to all accuracy data. Additionally, scores cannot be above 100% for accuracy, whereas RT's are not constrained in the same way. Improved reliability for participants' accuracy should be an aim for future development of such a framework if it is to be used as a diagnostic indicator in combination with or instead of reaction time.

The current work therefore represents a successful proof-of-concept study, indicating that a framework such as this could be further adapted in order to provide a simple, fast, and effective baseline screening measure against which decline in executive functions in those with Down syndrome can be benchmarked. Task impurity issues have traditionally made measuring

executive functions problematic (see Miyake, 2000; Karr et al., 2018). Although our current approach does not eliminate the impurity issue, by extracting executive subcomponents from a single task we are able to keep some of these sources of variance constant across the subcomponents of interest. If different tasks were used for each subcomponent, we would not know whether a decline in performance on one of those tasks was due to a change in the executive subcomponent of interest, or whether it was due to other processes unique to that task.

We should note that a degree of normal age-related decline in executive function over long periods of times would also be expected among individuals with Down syndrome, as it would among individuals without Down syndrome (Devenny, Silverman, Hill, Jenkins, Sersen & Wisniewski, 1996). However, decline in executive function related to AD can potentially occur at a much earlier age and be much more significant than what would be expected in normal ageing among adults with Down syndrome. Establishing baseline performance would therefore be important to flag such instances and prompt individuals to seek further clinical assessment to determine if the changes reflect early onset AD.

Future directions

One way to improve the task in the future could be to incorporate both RT and accuracy into a single measure. RT and accuracy measures do not always correlate with each other, and can be confounded by, for example, speed-accuracy trade-offs (Hedge et al. 2018). Incorporating measures that account for both speed and accuracy could enhance the validity of executive function measures in the future (for an overview, see Verhaeghen, 2013). Indeed we reanalysed our data with a recently proposed 'balanced integration score'; a composite of both RT and accuracy (Liesefeld & Janczyk, 2018), and the results of this were consistent with our main analyses (see Appendix 3). In future research it would also be informative to assess the

convergent validity of each of the measures extracted within this single task relative to other measures of executive functioning.

Given the consistency of findings reporting changes in executive function as the earliest indicator of dementia related decline among individuals with Down syndrome (e.g., Ball et al., 2006; Das et al., 1995; Holland, Hon, Huppert & Stevens, 2000; Lautarescu, Holland & Zaman, 2017), the potential to adapt the framework tested in the current study has important implications. Having a reliable baseline measure against which to track potential decline of executive functions in the Down syndrome population is an extremely valuable goal. Burt and Aylward (2000) have previously suggested that baseline measures should be obtained annually in the adult Down syndrome population. Individuals could complete a task such as that used in the current study at home to track progress, with the supervision of parents or carers if required. We emphasise that we do not expect performance on such a task to be the basis of a diagnosis of cognitive status (e.g., AD) in itself. Rather, our aim is to develop a reliable and valid tool that is easily administered without clinical supervision, but can be used to identify individuals for appropriate clinical assessment. We should note that the sample size in the current study was relatively small, nonetheless the expected effects of each measure were found, and reasonable reliability was observed.

Conclusion

In conclusion, the current proof-of-concept study shows that a task such as the one used in the current study can be used to obtain multiple components of executive function within a single, fast computerized assessment; such a task could be suitable for many individuals with Down syndrome given the overall high levels of accuracy observed in this group in the current study. In addition, these measures appear to be relatively well suited to tracking abilities over time. The framework used here therefore offers a means of capturing individual changes over time in executive function in individuals with Down syndrome. Future research using an adapted

version of this framework with large samples is warranted, and would have considerable potential for improving clinical practice.

References

American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Publishing.

Baddeley, A., & Jarrold, C. (2007). Working memory and Down syndrome. Journal of Intellectual Disability Research, 51(12), 925-931.

Ball, S. L., Holland, A. J., Hon, J., Huppert, F. A., Treppner, P., & Watson, P. C. (2006). Personality and behaviour changes mark the early stages of Alzheimer's disease in adults with Down's syndrome: findings from a prospective population-based study. *International Journal of Geriatric Psychiatry, 21(7),* 661-673.

Ball, S. L., Holland, A. J., Treppner, P., Watson, P. C., & Huppert, F. A. (2008). Executive dysfunction and its association with personality and behaviour changes in the development of Alzheimer's disease in adults with Down syndrome and mild to moderate learning disabilities. *British Journal of Clinical Psychology, 47(1),* 1-29.

Ballard, C., Mobley, W., Hardy, J., Williams, G., & Corbett, A. (2016). Dementia in Down's syndrome. *The Lancet Neurology, 15(6),* 622-636.

Bates, D., Maechler, M., Bolke, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67 (1),* 1-48, doi:10.18637/jss.v067.i01.

Beveridge, M., Jarrold, C., & Pettit, E. (2002). An experimental approach to executive fingerprinting in young children. *Infant and Child Development, 11,* 107-123.

Borella, E., Carretti, B., & Lanfranchi, S. (2013). Inhibitory mechanisms in Down syndrome: Is there a specific or general deficit? *Research in Developmental Disabilities, 34(1),* 65-71.

Braver T. S., Gray J. R., Burgess G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In Conway A. R. A., Jarrold C., Kane M. J., Miyake A., Towse J. N., editors. (Eds.), Variation in working memory (pp. 76–106). New York City, NY: Oxford University Press.

Brock, J., & Jarrold, C. (2005). Serial order reconstruction in Down syndrome: evidence for a selective deficit in verbal short-term memory. *Journal of Child Psychology and Psychiatry, 46(3),* 304-316.

Burt, D. B., & Aylward, E. H., D. (2000). On behalf of the members of the Working Group for the Establishment of Criteria for the Diagnosis of Dementia in Individuals with intellectual. Test battery for the diagnosis of dementia in individuals with intellectual disability, *Journal of Intellectual Disability Research, 44(2),* 175-180.

Carr, J. (1988). Six weeks to twenty-one years old: A longitudinal study of children with Down's syndrome and their families: Third Jack Tizard Memorial Lecture. *Journal of child Psychology and Psychiatry, 29(4),* 407-431.

Chapman, R. S. (1997). Language development in children and adolescents with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews, 3(4),* 307-312.

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 5529-5533.

Chevalier, N., Huber, K. L., Wiebe, S. A., & Espy, K. A. (2013). Qualitative change in executive control during childhood and adulthood. *Cognition, 128,* 1-12.

Chevalier, N., James, T. D., Wiebe, S. A., Nelson, J. M., & Espy, K. A. (2014). Contribution of reactive and proactive control to children's working memory performance: Insight from item recall durations in response sequence planning. *Developmental Psychology*, *50*(7), 1999.

Chevalier, N., Martis, S. B., Curran, T., & Munakata, Y. (2015). Metacognitive processes in executive control development: The case of reactive and proactive control. *Journal of Cognitive Neuroscience*, *27*(6), 1125-1136.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing Criteria for Establishing Interrater Reliability of Specific Items–Applications to Assessment of Adaptive-Behavior. *American Journal of Mental Deficiency, 86(2),* 127–137.

Costanzo, F., Varuzza, C., Menghini, D., Addona, F., Gianesini, T., & Vicari, S. (2013). Executive functions in intellectual disabilities: a comparison between Williams syndrome and Down syndrome. *Research in Developmental Disabilities, 34(5),* 1770-1780.

Cork, L. C. (1990). Neuropathology of Down syndrome and Alzheimer disease. American *Journal of Medical Genetics, 37(S7),* 282-286.

Cornish, K., Scerif, G., & Karmiloff-Smith, A. (2007). Tracing syndrome-specific trajectories of attention across the lifespan. *Cortex, 43(6),* 672-685.

Crayton, L., Oliver, C., Holland, A., Bradbury, J., & Hall, S. (1998). The Neuropsychological Assessment of Age Related Cognitive Deficits in Adults with Down's Syndrome. *Journal of Applied Research in Intellectual Disabilities, 11(3),* 255-272.

d'Ardhuy, X. L., Edgin, J. O., Bouis, C., de Sola, S., Goeldner, C., Kishnani, P., ... & Spiridigliozzi, G. (2015). Assessment of cognitive scales to examine memory, executive function and language in individuals with Down syndrome: Implications of a 6-month observational study. *Frontiers in Behavioral Neuroscience*, *9*, 300.

Das, J. P., Divis, B., Alexander, J., Parrila, R. K., & Naglieri, J. A. (1995). Cognitive decline due to aging among persons with Down syndrome. *Research in Developmental Disabilities, 16(6),* 461-478.

Daunhauer, L. A., Fidler, D. J., Hahn, L., Will, E., Lee, N. R., & Hepburn, S. (2014). Profiles of everyday executive functioning in young children with Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *119*(4), 303-318.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44,* 2037-2078.

De Simoni C & von Bastian CC (2018) Working memory updating and binding training: Bayesian evidence supporting the absence of transfer. *Journal of Experimental Psychology: General, 147(6),* 829-858

Devenny, D. A., Silverman, W. P., Hill, A. L., Jenkins, E., Sersen, E. A., & Wisniewski, K. E. (1996). Normal ageing in adults with Down's syndrome: A longitudinal study. *Journal of Intellectual Disability Research*, *40*(3), 208-221.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33(3),* 261-268.

Edgin, J. O., Anand, P., Rosser, T., Pierpont, E. I., Figueroa, C., Hamilton, D., ... & Nguyen-Driver, M. (2017). The Arizona Cognitive Test Battery for Down Syndrome: Test-Retest Reliability and Practice Effects. *American Journal on Intellectual and Developmental Disabilities*, *122*(3), 215-234.

Esbensen, A. J., Hooper, S. R., Fidler, D., Hartley, S. L., Edgin, J., d'Ardhuy, X. L., ... & Rafii, M. (2017). Outcome measures for clinical trials in Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *122*(3), 247-281.

Fidler, D. J., Philofsky, A., Hepburn, S. L., & Rogers, S. J. (2005). Nonverbal requesting and problem-solving by toddlers with Down syndrome. *American Journal on Mental Retardation*, *110*(4), 312-322.

Fleiss, J. L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley.

Fonseca, L. M., Yokomizo, J. E., Bottino, C. M., & Fuentes, D. (2016). Frontal lobe degeneration in adults with Down syndrome and Alzheimer's disease: a review. *Dementia and Geriatric Cognitive Disorders, 41(3-4),* 123-136.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17(2),* 172-179.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137,* 201-225.

Gioia, G. A., Isquith, P. K., & Espy, K. (2003). Construct validity of the behavior rating inventory of executive function. Preschool version. *Journal of the International Neuropsychological Society*, *9*, 297-306.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Behavior rating inventory of executive function. *Lutz, Florida: Psychological Assessment Resources*.

Giraudeau, B. (1996). Negative values of the intraclass correlation coefficient are not theoretically possible. *Journal of Clinical Epidemiology, 49(10),* 1205.

Hartley, D., Blumenthal, T., Carrillo, M., DiPaolo, G., Esralew, L., Gardiner, K., ... & Lott, I. (2015). Down syndrome and Alzheimer's disease: Common pathways, common goals. *Alzheimer's & Dementia*, *11*(6), 700-709.

Hartley, S. L., Handen, B. L., Devenny, D., Mihaila, I., Hardison, R., Lao, P. J., ... & Christian, B. T. (2017). Cognitive decline and brain amyloid-β accumulation across 3 years in adults with Down syndrome. *Neurobiology of Aging*, *58*, 68-76.

Hedge, C., Powell, G., Bompas, A., Vivian-Griffiths, S., & Sumner P. (2018). Low and variable correlation between reaction time costs and accuracy costs explained by accumulation models: meta-analysis and simulations. *Psychological Bulletin, 144(11),* 1200-1227.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50(3),* 1166-1186.

Holland, A. J., Hon, J., Huppert, F. A., & Stevens, F. (2000). Incidence and course of dementia in people with Down's syndrome: findings from a population-based study. *Journal of Intellectual Disability Research, 44(2),* 138-146.

Holland, A. J., Hon, J., Huppert, F. A., Stevens, F., & Watson, P. (1998). Population-based study of the prevalence and presentation of dementia in adults with Down's syndrome. *The British Journal of Psychiatry, 172(6),* 493-498.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44(11),* 2017-2036.

Jarrold, C., Baddeley, A., & Phillips, C. (1999). Down syndrome and the phonological loop: The evidence for, and importance of, a specific verbal short-term memory deficit. *Down Syndrome Research and Practice, 6(2),* 61-75.

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of" executive attention". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 749.

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological bulletin.*

Keeling, L. A., Spiridigliozzi, G. A., Hart, S. J., Baker, J. A., Jones, H. N., & Kishnani, P. S. (2017). Challenges in measuring the effects of pharmacological interventions on cognitive and adaptive functioning in individuals with Down syndrome: A systematic review. *American Journal of Medical Genetics Part A*, *173*(11), 3058-3066.

Killikelly, C., & Szűcs, D. (2013). Delayed development of proactive response preparation in adolescents: ERP and EMG evidence. *Developmental cognitive neuroscience*, *3*, 33-43.

Krinsky-McHale, S. J., Devenny, D. A., Gu, H., Jenkins, E. C., Kittler, P., Murty, V. V., ... & Ye, L. (2008). Successful aging in a 70-year-old man with down syndrome: a case study. *Intellectual and Developmental Disabilities*, *46*(3), 215-228.

Kuznetsova, A., Brockhof, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-33. https://CRAN.R-project.org/package=lmerTest.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33(1),* 159–174.

Landry, O., Russo, N., Dawkins, T., Zelazo, P. D., & Burack, J. A. (2012). The Impact of Verbal and Nonverbal Development on Executive Function in Down Syndrome and Williams Syndrome. *Journal on Developmental Disabilities, 18(2).*

Lanfranchi, S., Carretti, B., Spanò, G., & Cornoldi, C. (2009). A specific deficit in visuospatial simultaneous working memory in Down syndrome. *Journal of Intellectual Disability Research, 53(5),* 474-483.

Lanfranchi, S., Cornoldi, C., & Vianello, R. (2004). Verbal and visuospatial working memory deficits in children with Down syndrome. *American Journal on Mental Retardation, 109(6),* 456-466.

Lanfranchi, S., Jerman, O., Dal Pont, E., Alberti, A., & Vianello, R. (2010). Executive function in adolescents with Down Syndrome. *Journal of Intellectual Disability Research, 54(4),* 308-319.

Lautarescu, B. A., Holland, A. J., & Zaman, S. H. (2017). The Early Presentation of Dementia in People with Down Syndrome: a Systematic Review of Longitudinal Studies. *Neuropsychology Review, 1-15.*

Lee, N. R., Anand, P., Will, E., Adeyemi, E. I., Clasen, L. S., Blumenthal, J. D., ... & Edgin, J. O. (2015). Everyday executive functions in Down syndrome from early childhood to young adulthood: evidence for both unique and shared characteristics compared to youth with sex chromosome trisomy (XXX and XXY). *Frontiers in Behavioral Neuroscience, 9.*

Lee, N. R., Fidler, D. J., Blakeley-Smith, A., Daunhauer, L., Robinson, C., & Hepburn, S. L. (2011). Caregiver report of executive functioning in a population-based sample of young children with Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *116*(4), 290-304.

Liesefeld, H. R., & Janczyk, M. (2018). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1076-x

Loveall, S. J., Conners, F. A., Tungate, A. S., Hahn, L. J., & Osso, T. D. (2017). A cross-sectional analysis of executive function in Down syndrome from 2 to 35 years. *Journal of Intellectual Disability Research*, *61*(9), 877-887.

Martin, G. E., Klusek, J., Estigarribia, B., & Roberts, J. E. (2009). Language characteristics of individuals with Down syndrome. *Topics in Language Disorders, 29(2),* 112.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7(3),* 134-140.
Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*(4), 313-324.A
Newcombe, F. (1969). Missile wounds of the brain: A study of psychological deficits.
Nieuwenhuis-Mark, R. E. (2009). Diagnosing Alzheimer's dementia in Down syndrome: Problems and possible solutions. *Research in Developmental Disabilities, 30(5),* 827-838.
Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), Consciousness and self-regulation (pp. 1-18). New York: Plenum.
Pennington, B. F., Moon, J., Edgin, J., Stedron, J., & Nadel, L. (2003). The Neuropsychology of Down Syndrome: Evidence for Hippocampal Dysfunction. *Child Development, 74(1),* 75-93.
Prasher, V. P., & Krishnan, V. H. R. (1993). Age of onset and duration of dementia in people with Down syndrome: Integration of 98 reported cases in the literature. *International Journal of Geriatric Psychiatry, 8(11),* 915-922.
R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rabbit, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), Methodology of frontal executive function (pp. 1-38). East Sussex, UK: Psychology Press.
Raven, J., Raven, J. C., & Court, J. H. (1998). Coloured Progressive Matrices. Oxford, England: Oxford Psychologists Press.
Reid, D., Moss, J., Nelson, L., Groves, L., & Oliver, C. (2017). Executive functioning in Cornelia de Lange syndrome: domain asynchrony and age-related performance. *Journal of Neurodevelopmental Disorders, 9(1),* 29.
Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 44(4),* 501-526.
Roberts, J. E., Price, J., & Malkin, C. (2007). Language and communication development in down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews, 13(1),* 26-35.
Rowe, J., Lavender, A., & Turk, V. (2006). Cognitive executive function in Down's syndrome. *British Journal of Clinical Psychology, 45(1),* 5-17.
Ruskin, E. M., Mundy, P., Kasari, C., & Sigman, M. (1994). Object mastery motivation of children with Down syndrome. *American Journal on Mental Retardation*.
Shallice, T. (1982). Specific impairments of planning. *Phil. Trans. R. Soc. Lond. B*, *298*(1089), 199-209.
Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81,* 174-176.
Stanton, L. R., & Coetzee, R. H. (2004). Down's syndrome and dementia. *Advances in Psychiatric Treatment, 10(1),* 50-58.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.
Strydom, A., Livingston, G., King, M., & Hassiotis, A. (2007). Prevalence of dementia in intellectual disability using different diagnostic criteria. *The British Journal of Psychiatry, 191(2),* 150-157.

Tyrrell, J., Cosgrave, M., McCarron, M., McPherson, J., Calvert, J., Kelly, A., ... & Lawlor, B. A. (2001). Dementia in people with Down's syndrome. *International Journal of Geriatric Psychiatry*, *16*(12), 1168-1174.

Vakil, E., & Lifshitz-Zehavi, H. (2012). Solving the Raven Progressive Matrices by adults with intellectual disability with/without Down syndrome: different cognitive patterns as indicated by eye-movements. *Research in Developmental Disabilities*, *33*(2), 645-654.

Verhaeghen, P. (2013). *The elements of cognitive aging. Meta-analysis of age-related differences in processing speed and their consequences*. Oxford: Oxford University Press.

Westphal, A. (2013). Cognitive Dementia and Memory Service (CDAMS), Literature Review. Prepared for the Victorian Department of Health as part of the review of CDAMS practice guidelines, 2013.

Will, E., Fidler, D. J., Daunhauer, L., & Gerlach-McDonald, B. (2017). Executive function and academic achievement in primary–grade students with Down syndrome. *Journal of Intellectual Disability Research, 61(2),* 181-195.

Wishart, J. (2001). Motivation and learning styles in young children with Down syndrome. *Down syndrome Research and practice*, *7*(2), 47-51.

World Health Organization (2011). International Statistical Classification of Diseases and Related Health Problems. 10th Revision. Volume 2 Instruction Manual. Geneva, Switzerland: World Health Organization.

Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2017). Participant nonnaiveté and the reproducibility of cognitive psychology. *Psychonomic bulletin & review*, 1-5.

Table 1. *Means and standard deviations for participants.*

	Reaction times (milliseconds)			Accuracy (percentages)		
	DS (N = 18) M (SD)	TD matched (N = 18) M (SD)	Grand Mean	DS (N = 18) M (SD)	TD matched (N = 18) M (SD)	Grand Mean
No memory load						
No delay						
Congruent	1330 (256)	1363 (284)	1347	99.5 (2.0)	90.3 (11.2)	94.9
Incongruent	1364 (338)	1379 (290)	1372	96.3 (13.8)	93.5 (11.6)	94.9
Delay						
Congruent	925 (418)	791 (273)	858	98.2 (3.6)	96.3 (8.7)	97.2
Incongruent	976 (494)	782 (243)	879	96.3 (11.9)	97.7 (3.8)	97.0
Low memory load						
No delay						
Congruent	1536 (421)	1462 (273)	1499	89.1 (14.8)	86.1 (16.4)	87.6
Incongruent	1520 (217)	1603 (324)	1562	84.9 (15.6)	87.0 (16.7)	86.0
Delay						
Congruent	909 (224)	884 (302)	897	89.1 (11.7)	87.5 (10.4)	88.3
Incongruent	961 (249)	986 (472)	974	87.5 (17.5)	88.4 (11.8)	88.0
High memory load						
No delay						
Congruent	1784 (0494)	1568 (0271)	1676	77.6 (20.1)	82.8 (17.6)	80.2

Incongruent	1757 (0497)	1605 (0317)	1681	75.0 (19.5)	75.5 (23.5)	75.3
Delay						
Congruent	1117 (0355)	0823 (0312)	970	84.9 (15.6)	86.5 (13.6)	85.7
Incongruent	1047 (0346)	1025 (0469)	1036	75.0 (21.1)	83.3 (14.9)	79.2

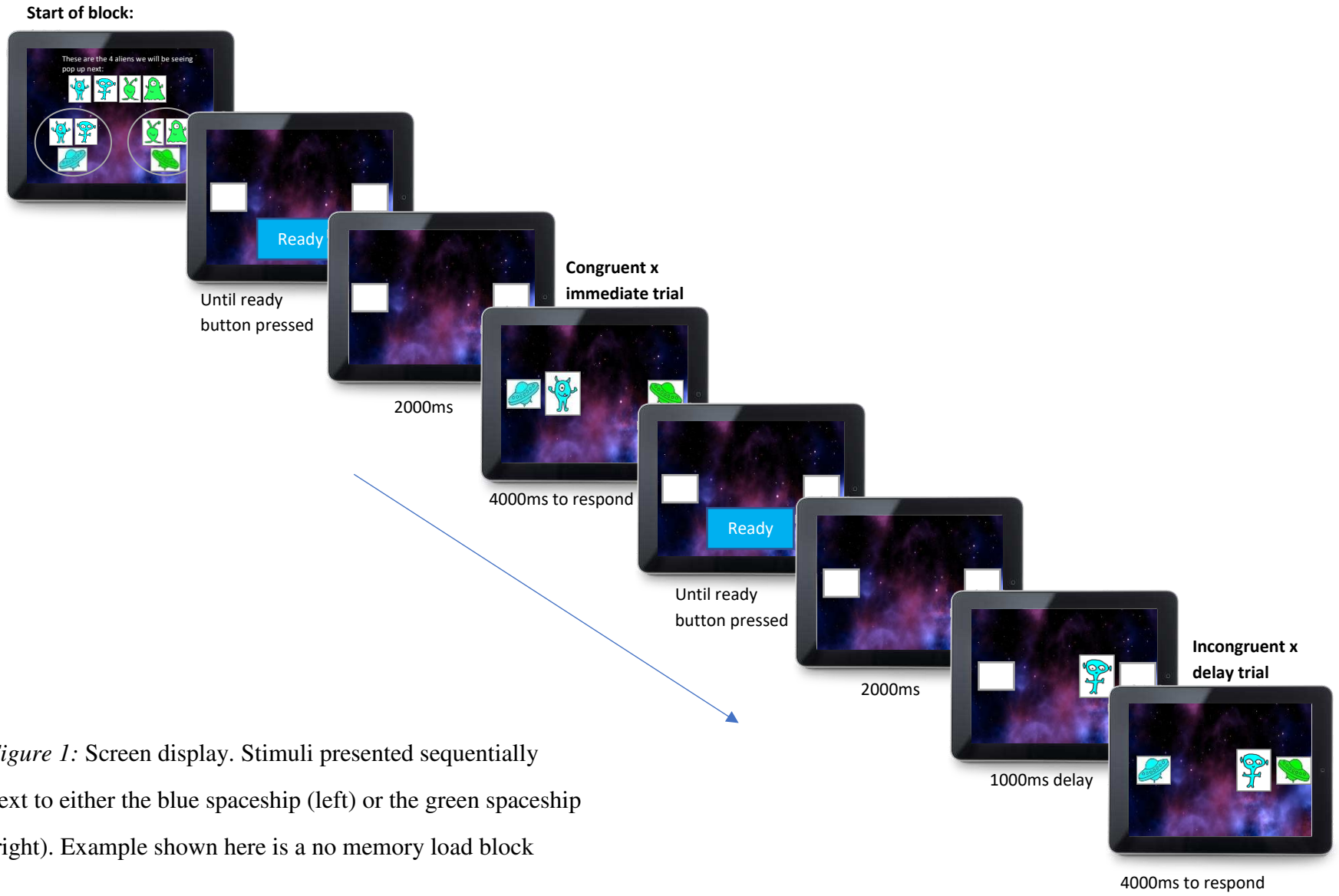


Figure 1: Screen display. Stimuli presented sequentially next to either the blue spaceship (left) or the green spaceship (right). Example shown here is a no memory load block

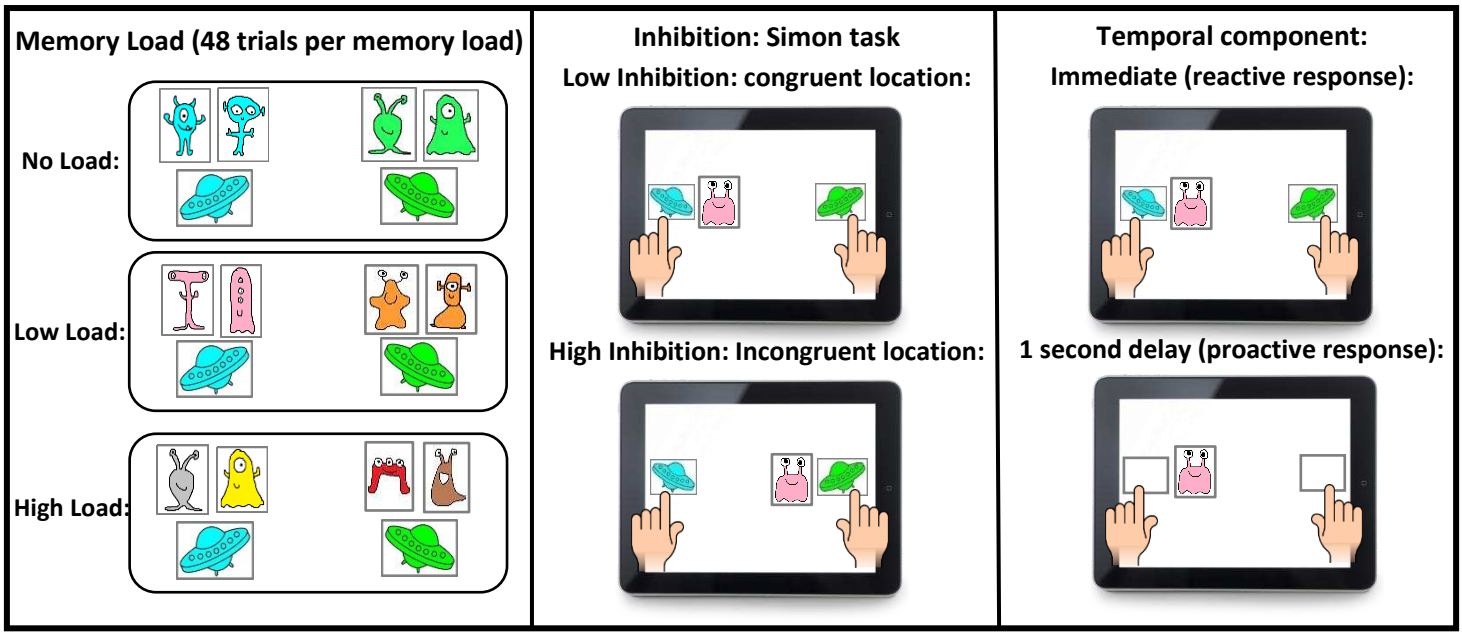


Figure 2. The three levels of memory load, two levels of inhibition and two levels of temporal load. The memory load rule was presented visually at the beginning of a block. No verbal labels were used to describe the memory rules, and, no verbal instructions for any of the three manipulations were given during the main task.

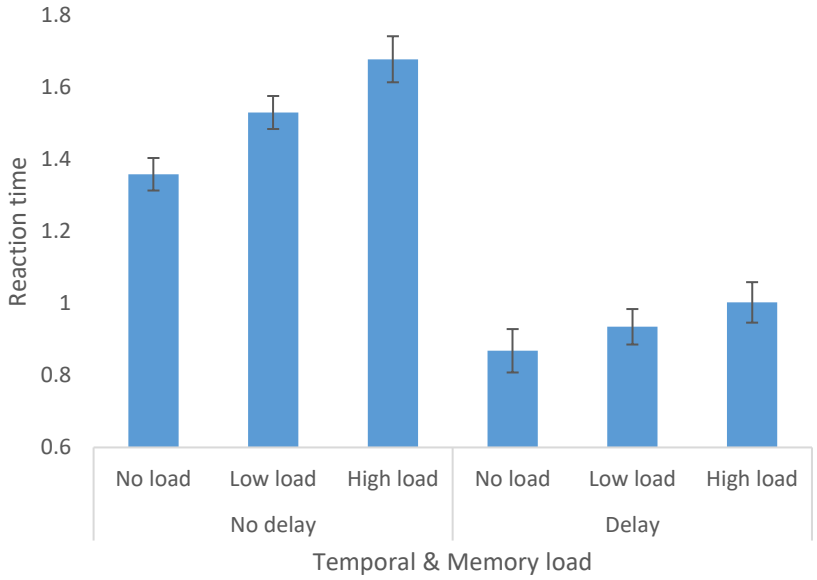
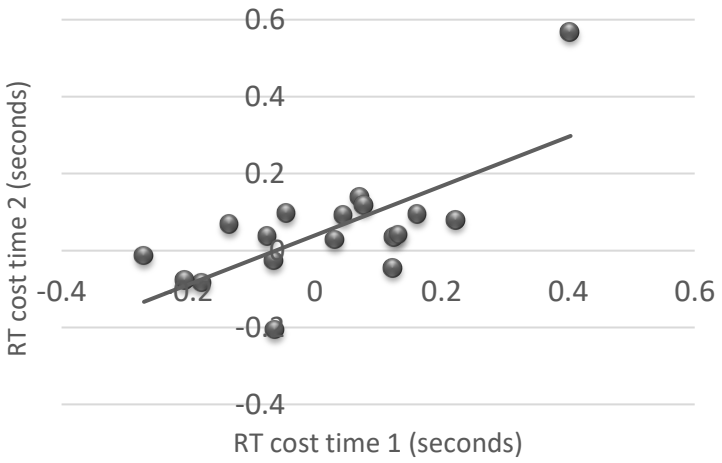
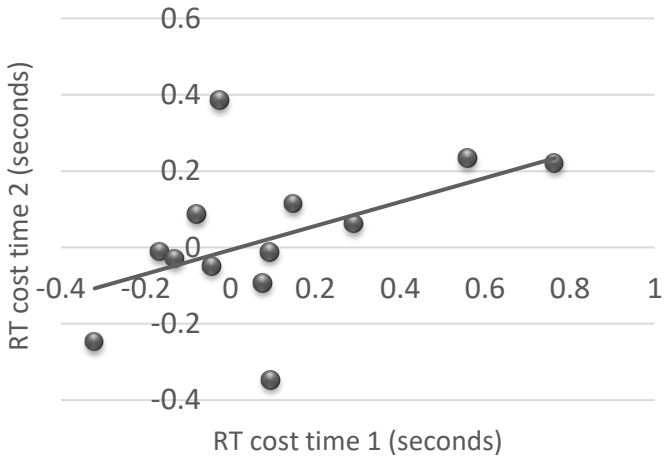


Figure 3. Graph showing memory load RT x temporal load RT across both populations. Error bars represent +/- 1 standard error. In the immediate (no delay) condition the significant effect of memory load was more pronounced, relative to the significant effect of memory load in the delay condition.

Inhibitory load cost:



Memory load cost:



Temporal load cost:

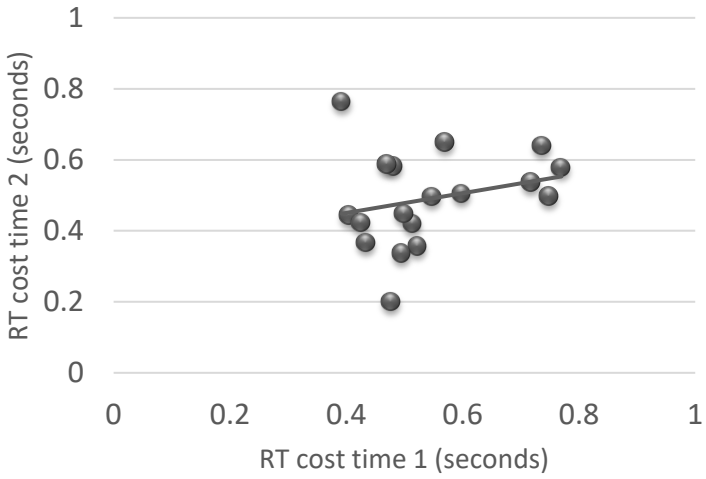
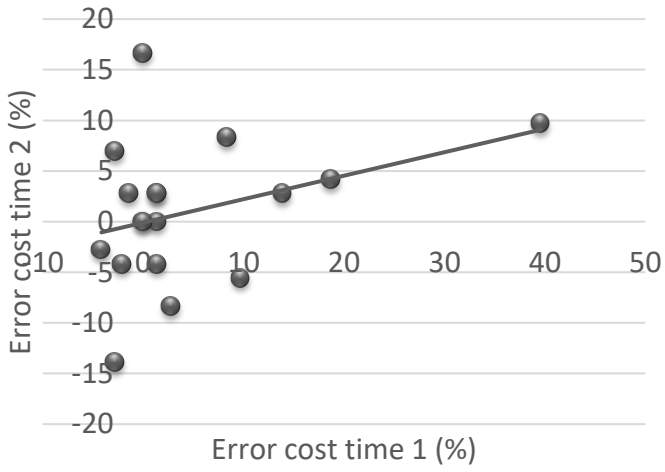
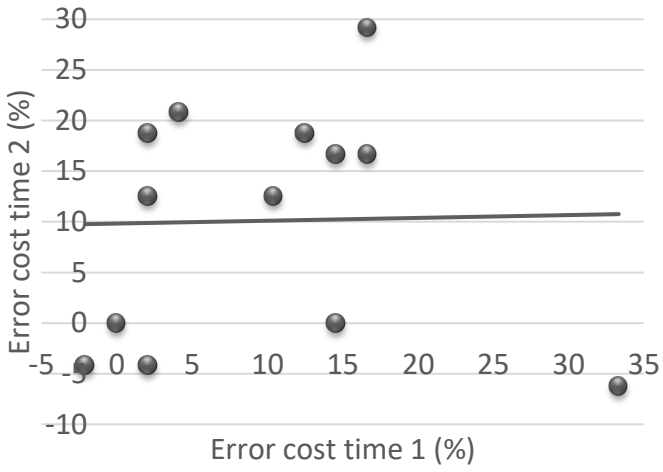


Figure 4. Scatterplots showing correlation of RT cost at time 1 x time 2 for each factor. Note. The range is different in each graph due to the different range of costs for each factor.

Inhibitory load cost:



Memory load cost:



Temporal load cost:

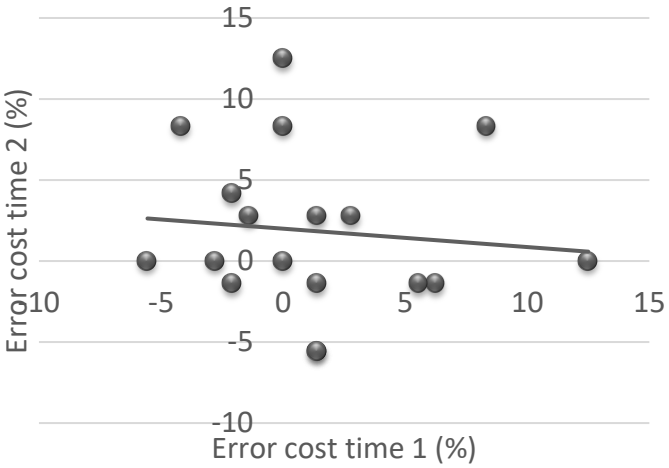


Figure 5. Scatterplots showing correlation of accuracy cost at time 1 x time 2 for each factor.

Note. The range is different in each graph due to the different range of costs for each factor.

Appendix 1. Results with full data set (not applying accuracy cut off).

Reaction times

Descriptive data:

Mean RTs

	No load	Low load	High load
DS	1.148696	1.340448	1.470766
TD	1.074703	1.223341	1.265424

	No delay	Delay
DS	1.586335	1.053605
TD	1.495693	0.879953

	Congruent	Incongruent
DS	1.295958	1.343982
TD	1.150806	1.22484

Linear mixed effects model output:

	df	*F*	*p*
Group	1, 34	1.85	.182
Memory load	2, 374	35.21	.000
Delay	1, 374	511.47	.000
Congruency	1, 374	5.78	.017
Memory load x delay	2, 374	3.03	.0497

Note. All remaining interactions were non-significant, with all F values < 2.67, and all p values > .10.

Accuracy

Descriptive data:

Mean Accuracy

	No load	Low load	High load
DS	97.56945	81.71297	72.56945
TD	93.86574	85.64815	74.18982

	No delay	Delay
DS	82.79321	85.10803
TD	82.79321	86.34259

	Congruent	Incongruent
DS	85.64815	82.25309
TD	84.5679	84.5679

Linear mixed effects model output:

	df	*F*	*p*
Group	1, 34	0.026	.872
Memory load	2, 374	100.723	.000
Delay	1, 374	5.196	.023
Congruency	1, 374	1.742	.188
Group x memory load	2, 374	3.091	.047

Note. All remaining interactions were non-significant, with all F values < 1.74, and all p values > .188.

Appendix 2. Descriptive data for time 2

Means and standard deviations for the Down syndrome group at time 2 (Accuracy cut off not performed):

	Reaction times (milliseconds)	**Accuracy (percentages)**
	DS Time 2 M (SD)	DS Time 2 M (SD)
No memory load		
No delay		
Congruent	1227 (214)	99.1 (2.7)
Incongruent	1289 (226)	98.6 (3.2)
Delay		
Congruent	822 (159)	99.1 (2.7)
Incongruent	880 (223)	100 (0)
Low memory load		
No delay		
Congruent	1469 (319)	86.6 (13.7)
Incongruent	1608 (494)	82.9 (18.0)
Delay		
Congruent	982 (358)	91.7 (9.5)
Incongruent	1015 (542)	86.6 (17.4)
High memory load		
No delay		
Congruent	1584 (523)	73.6 (20.7)
Incongruent	1606 (379)	73.1 (25.2)
Delay		

Congruent	1080 (384)	75.9 (17.6)
Incongruent	1075 (524)	74.1 (18.7)

Appendix 3. Composite scores

Here, we report the results of an analysis using the balanced integration score (BIS; Liesefeld & Janczyk, 2018). The BIS incorporates both reaction time and accuracy in to a single measure, and was shown to control well for speed-accuracy trade-offs compared to other measures.

The BIS is applied to participants' mean percentage correct (PC) and mean RT measures. First, for each condition and participant, PC and RT are separately z-scored by subtracting the overall mean (across all participants and conditions) and dividing by the overall standard deviation, e.g. for RT:

$$zRT_{ij} = \frac{RT_{ij} - \overline{RT}}{S_{RT}}$$

The i and j subscripts refer to participant and condition respectively. To calculate the BIS for each condition, the standardised RT score is subtracted from the standardised accuracy

$$BIS_{ij} = zPC_{ij} - zRT_{ij}$$

Note that, because the BIS consists of standardising the scores using the sample mean and standard deviation, it gives participants a value *relative* to other individuals in the sample, rather than an absolute performance score. This is not appropriate for all purposes, though is sufficient for assessing whether speed-accuracy trade-offs contaminate the sample effects reported in the main text. Supplementary Table 1 below displays the results from a linear mixed effects analysis (as conducted in the main text on RT and accuracy separately) performed on the BIS. We applied these to the data after removing blocks in which participants were below 60% accuracy

The results were generally consistent with our main analyses. We observed main effects of Memory Load, Delay and Congruency. The effect of group was not significant, nor were any of the interactions.

Supplementary Table 1: Results of linear mixed effects model conducted on balanced integration scores.

Effect	Sum of squares	Mean square	Num. DF	Denom. DF	F	p
Group	0.33	0.33	1.00	32.99	0.33	0.57
Memory Load	163.21	81.61	2.00	351.78	80.96	<.001***
Delay	209.64	209.64	1.00	348.98	207.98	<.001***
Congruency	5.68	5.68	1.00	348.98	5.63	0.02*
Group: x Memory Load	3.65	1.82	2.00	351.78	1.81	0.17
Group x Delay	2.10	2.10	1.00	348.98	2.08	0.15
Memory Load x Delay	5.40	2.70	2.00	348.98	2.68	0.07
Group x Congruency	0.03	0.03	1.00	348.98	0.03	0.85
Memory Load x Congruency	2.51	1.26	2.00	348.98	1.25	0.29
Delay x Congruency	0.09	0.09	1.00	348.98	0.09	0.76
Group x Memory Load x Delay	3.11	1.56	2.00	348.98	1.54	0.22
Group x Memory Load x Congruency	2.02	1.01	2.00	348.98	1.00	0.37
Group x Delay x Congruency	0.03	0.03	1.00	348.98	0.03	0.85
Memory Load x Delay x Congruency	0.37	0.18	2.00	348.98	0.18	0.83
Group x Memory Load x Delay x Congruency	0.16	0.08	2.00	348.98	0.08	0.92

Note: *p<.05, ***p<.001

Appendix 4. Parameter estimates from linear mixed effects models.

Supplementary Table 2. Estimates for linear mixed effects model of reaction time performance. Group (Down syndrome vs. typically developing) is a between subjects factor, all other factors are within-subjects.

Parameter	Estimate	Std. Error	DF	p
Fixed effects				
Intercept (DS group, no memory load, no delay, congruent)	1.33	0.09	94.25	<0.001
Group (typically developing)	0.03	0.12	94.25	0.782
Low memory load	0.27	0.09	349.53	0.002
High memory load	0.43	0.09	349.53	<0.001
Delay	-0.4	0.08	349.02	<0.001
Incongruent	0.03	0.08	349.02	0.675
Group * low memory load	-0.17	0.12	349.28	0.155
Group * high memory load	-0.19	0.12	349.52	0.117
Group * delay	-0.17	0.12	349.02	0.152
Low memory load * delay	-0.22	0.12	349.02	0.065
High memory load * delay	-0.26	0.12	349.02	0.030
Group * incongruent	-0.02	0.12	349.02	0.872
Low memory load *incongruent	-0.05	0.12	349.02	0.675
High memory load * incongruent	-0.06	0.12	349.02	0.607
Delay * incongruent	0.02	0.12	349.02	0.892
Group * low memory load * delay	0.22	0.17	349.02	0.197
Group * high memory load * delay	0.09	0.17	349.02	0.601
Group * low memory load * incongruent	0.18	0.17	349.02	0.296
Group * high memory load * incongruent	0.08	0.17	349.02	0.627
Group * delay * incongruent	-0.04	0.17	349.02	0.805
Low memory load * delay * incongruent	0.05	0.17	349.02	0.757
High memory load * delay * incongruent	-0.06	0.17	349.02	0.730
Group * low memory load * delay * incongruent	-0.07	0.24	349.02	0.780
Group * high memory load * delay * incongruent	0.25	0.24	349.02	3.013
Random effects (variance)				
Participant (intercept)	0.071			
Residual	0.061			

Note. Model estimated using restricted maximum likelihood. Degrees of freedom and p-values are calculated using Satterthwaite's approximation. Effects are relative to the reference categories (Down syndrome group, no memory load condition, no delay, congruent).

Supplementary Table 3. Estimates for linear mixed effects model of accuracy performance. Group (Down syndrome vs. typically developing) is a between subjects factor, all other factors are within-subjects.

Parameter	Estimate	Std. Error	DF	p
Fixed effects				
Intercept (DS group, no memory load, no delay, congruent)	99.54	3.42	210.96	<0.001
Group (typically developing)	-9.27	4.83	210.96	0.057
Low memory load	-11.49	4.24	350.62	0.007
High memory load	-22.03	4.24	350.62	<0.001
Delay	-1.38	4.11	349.42	0.736
Incongruent	-3.24	4.11	349.42	0.431
Group * low memory load	7.33	5.91	350.04	0.216
Group * high memory load	13.46	6	350.61	0.026
Group * delay	7.41	5.81	349.42	0.203
Low memory load * delay	1.38	5.99	349.42	0.817
High memory load * delay	8.68	5.99	349.42	0.148
Group * incongruent	6.48	5.81	349.42	0.265
Low memory load *incongruent	-0.93	5.99	349.42	0.877
High memory load * incongruent	0.64	5.99	349.42	0.915
Delay * incongruent	1.38	5.81	349.42	0.812
Group * low memory load * delay	-6.02	8.34	349.42	0.471
Group * high memory load * delay	-11.05	8.46	349.42	0.192
Group * low memory load * incongruent	-1.39	8.34	349.42	0.867
Group * high memory load * incongruent	-11.17	8.46	349.42	0.188
Group * delay * incongruent	-3.23	8.21	349.42	0.694
Low memory load * delay * incongruent	1.22	8.46	349.42	0.885
High memory load * delay * incongruent	-8.68	8.46	349.42	0.306
Group * low memory load * delay * incongruent	0.63	11.79	349.42	0.957
Group * high memory load * delay * incongruent	14.69	11.97	349.42	0.221
Random effects				
Participant (intercept)	58.53			
Residual	151.74			

Note. Model estimated using restricted maximum likelihood. Degrees of freedom and p-values are calculated using Satterthwaite's approximation. Effects are relative to the reference categories (Down syndrome group, no memory load condition, no delay, congruent).