
Title Developmental changes in executive functioning
Author(s) Kerry Lee, Rebecca Bull and Ringo M. H. Ho
Source *Child Development, 84*(6), 1933–1953
Published by Wiley-Blackwell

This is the pre-peer reviewed version of the following article: Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development, 84*(6), 1933–1953. doi: 10.1111/cdev.12096, which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1111/cdev.12096/abstract

Developmental Changes in Executive Functioning

Kerry Lee and Rebecca Bull

National Institute of Education, Singapore

Ringo M. H. Ho

Nanyang Technological University

Author Note

Kerry Lee and Rebecca Bull, Applied Cognitive Development Lab & Centre for Research in Pedagogy and Practice, National Institute of Education, Singapore; Ringo M. H. Ho, Department of Psychology, Nanyang Technological University.

This study was supported by grants from the Office of Educational Research and the Centre for Research in Pedagogy and Practice. Views expressed in this article do not necessarily reflect those of the National Institute of Education. We thank the children who participated in the study and school administrators who provided access and assistance. Special thanks to Siran Zhan and EeLynn Ng for assistance with data analyses. We also thank Su Yin Ang, Fannie Khng, Jeremy Ng, and the other research assistants who helped with data collection and project management.

Correspondence concerning this article should be addressed to Kerry Lee, National Institute of Education, 1 Nanyang Walk, Singapore 637616. Email: Kerry.Lee@nie.edu.sg. Telephone: +65 6219 6251, Fax: +65 6896 9845

Abstract

Although early studies of executive functioning in children supported Miyake et al.'s (2000) three factor model, more recent findings supported a variety of undifferentiated or two factor structures. Using a cohort-sequential design, this study examined whether there were age-related differences in the structure of executive functioning amongst 6 to 15-year-olds (*N* = 688). Children were tested annually on tasks designed to measure updating and working memory, inhibition, and switch efficiency. There was substantial task based variation in developmental patterns on the various tasks. Confirmatory factor analyses and tests for longitudinal factorial invariance showed that data from the 5 to 13-year-olds conformed to a two factor structure. For the 15-year-olds, a well separated three factor structure was found.

Keywords: Executive function, longitudinal study, factor structure, inhibition, switching,
working memory

Developmental Changes in Executive Functioning

During the last decade, a large number of studies have examined executive functioning (EF) and its relation to academic achievement and school readiness (e.g., Bull, Espy, & Wiebe, 2008; Lee, Ng, Bull, Pe, & Ho, 2011; Michel & Roebers, 2008; St Clair-Thompson & Gathercole, 2006; Swanson, Jerman, & Zheng, 2008), social competence and theory of mind (e.g., Carlson, Mandell, & Williams, 2004; Hughes & Ensor, 2007), and behaviours associated with developmental disorders such as ADHD and autism (e.g., Happe, Booth, Charlton, & Hughes, 2006; Pellicano, 2010). Although EF is commonly defined as processes that control, direct, or coordinate other cognitive processes, both its conceptualisation and measurement vary across studies. In its most general form, EF encompasses a large range of top-down control and monitoring processes, such as attentional control, planning, and the regulation of action.

A dominant model in the adult literature, proposed by Miyake et al. (2000), focused on three aspects of EF: switching, inhibition, and updating. Switching or shifting refers to the ability to move between alternative sets of mental operations. Inhibition refers to the ability to resist interference from competing or prepotent responses or processes. Updating refers to the ability or capacity to refresh and maintain information in working memory in the presence of new information. Miyake and colleagues found the three abilities to be separately identifiable, but yet moderately to strongly correlated. To explain the correlations between factors, Miyake et al. proposed that updating requires discarding irrelevant incoming information and suppressing obsolete information. Likewise, switching requires the deactivation or suppression of an obsolete mental set in favour of a new one. Accordingly, poor behavioural inhibition would lead to secondary deficiencies in EF.

Many studies have adopted the multi-factorial framework of Miyake et al, but few provide an empirical or a theoretical justification for the suitability of the model for a child population. Despite early evidence suggesting that EFs can be divided into the same subtypes in children (Lehto, Juujaervi, Kooistra, & Pulkkinen, 2003), findings from more recent studies are equivocal, with a number of studies failing to find evidence for differentiation into the three subtypes (e.g., Espy, Sheffield, Wiebe, Clark, & Moehr, 2011; Hughes, Ensor, Wilson, & Graham, 2010; Wiebe, Espy, & Charak, 2008; Willoughby, Blair, Wirth, & Greenberg, 2010). Evidence from the neurosciences also suggests that the prefrontal and parietal structures, which support EFs, have long maturation schedules that continue into adolescence and early adulthood (e.g., Blakemore & Choudhury, 2006; but see Tamnes et al., 2010). Thus, it is perhaps precipitous to assume that a model of EF, based on adults' data, will also hold true across childhood.

Issues parallel to these questions were raised more than sixty years ago in a closely related area. A review of the literature led Garrett (1946) to propose a differentiation hypothesis: "abstract or symbolic intelligence changes in its organization as age increases from a fairly unified and general ability to a loosely organized group of abilities or factors" (p. 373). It was argued that the general intelligence factor, *g,* varies in its role in subtests of intelligence, such that it has a smaller role amongst older children and adults than children, and amongst the more able than the less able. With some noted exceptions (e.g., Molenaar, Dolan, Wicherts, & van der Maas, 2010), there is general support for differentiation, including findings from a recent study involving 6,273 4 to 101 year olds (Tucker-Drob, 2009). Studies conducted with children and adolescents suggest that differentiation occurs by the early adolescent years, but it is not clear whether there are significant changes either before or after the early teenage years. In an earlier study, Fitzgerald, Nesselroade, and Baltes (1973) found adult-like factor structures amongst a sample of 1,891 12 to 17 year olds, but Stankov (1978) found only partial differentiation amongst 11 – 12 year olds. Focusing on younger children, Atkin et al. (1977) found evidence of differentiation amongst a sample of 5 to 11 year olds. However, clear evidence of differentiation was found only amongst the oldest children and only amongst white males. In a more recent study, using a large sample of 4 to 9 year olds, Facon (2007) found that the same crystallised-fluid factor structure provided a good fit to the children's data. In contrast, Hulur, Wilhelm, and Robitzsch (2011) found evidence for differentiation, but the magnitude of change was small amongst 2.5 to 7 year olds.

In the context of executive functioning, the differentiation hypothesis was addressed in Shing, Lindenberger, Diamond, Li, and Davidson (2010), in which memory maintenance and inhibition were found to be undifferentiated up to 9.5 years of age. The factors became separable in 9.5 to 14.5 year olds. Although the differentiation hypothesis is useful for framing developmental differences in executive functioning, it also produces a similar difficulty as that which befell research on intelligence. In the context of intelligence, *g* is defined as the shared variance amongst measures of different aspects of intelligence. Although the notion of *g* has been replicated in a large number of studies, its substantive meaning is still a point of some controversy. Similarly, in the context of executive functioning, the nature of the shared variance amongst executive functioning tasks remains unresolved. In their seminal paper, Miyake et al. (2000) suggested two candidates: the controlled attentional aspect of working memory and inhibitory control. A third possibility, processing speed, was proposed in a more recent study (Rose, Feldman, & Jankowski, 2011).

In this study, we were primarily concerned with whether there are age-related differences in the structure of EFs. Miyake et al.'s (2000) tripartite model provided an initial framework for analysis. As noted above, there is some evidence pointing to qualitative invariance. That is, the organisation of EFs (the number of EFs and their inter-relations) seems to vary with age. If this is the case, there are two questions that need to be addressed. First, how is EF fractionated during childhood and the mid-adolescent years? Second, how do the relations between each function vary with age? These questions can perhaps be understood as addressing a strong and a weak version of the differentiation hypothesis; with the weak version specifying changes in the strength of correlations between abilities, and a strong version specifying changes in qualitative relations. In addition, we will examine the proposition that processing speed is a candidate that contributes to the efficiency of all executive functions. Although developmental differences in the structure of EF had previously been studied, the findings are inconclusive. We have identified several issues that may have contributed to the inconsistencies.

Limited Age Range. A number of studies examined the structure of EF in only one age group. In examining the psychometric properties of an EF test battery, Willoughby et al. (2010) tested 3 year olds and found a single factor model best accounted for the data (a more recent study with the same children tested one and two years later produced the same findings, Willoughby, Wirth, & Blair, 2012). Wiebe at al. (2011) also found evidence for an undifferentiated structure in 3 year olds (see Table 1 for more details on the design and tasks used in each confirmatory factor analytic study). Using an exploratory approach, Monette, Bigras, and Guay's (2011) data from 5 year olds conformed to a three factor solution with significant cross-loadings between the working memory and inhibitory measures. Focusing on 11 year olds, Rose et al. (2011) also found support for a three factor structure

<Insert Table 1 about here>

Some recent works have ventured beyond single age groups. Focusing on the preschool and kindergarten years, a number of studies (Hughes et al., 2010; Miller, Giesbrecht, Müller, McInerney, &

Kerns, 2012; Wiebe et al., 2008) found evidence for a single factor model . Shing et al. (2010) reanalysed Davidson, Amso, Anderson, and Diamond's (2006) data and found evidence of differentiation only amongst the 14 year olds. Huizinga, Dolan, and van der Molen's (2006) data on 7 to 21 year olds conformed to a two factor structure with measures loading onto a working memory and a shifting measure. McAuley and White (2011) also found support for a two factor solution across a similar age range. Unfortunately, a small sample size rendered their lack of age-related finding difficult to interpret. Furthermore, in this and several other studies (e.g., Agostino, Johnson, & Pascual-Leone, 2010; Brocki & Bohlin, 2004; Brookshire, Levin, Song, & Zhang, 2004; Klenberg, Korkman, & Lahti-Nuuttila, 2001; Lehto et al., 2003; Wu et al., 2011) different age groups were collapsed into one. This practice runs the risk of masking developmental change, especially during times when considerable maturation of EF skills can be expected.

Although comparing across studies provide some tantalising hints of developmental change, differences in tasks, constructs, and sample characteristics make it difficult to draw definitive conclusions. Furthermore, many studies deployed cross-sectional designs, from which it is difficult to draw strong conclusions about developmental changes. In this study we used a cohort-sequential design to study changes from kindergarten (6 year olds) to Secondary 3 (15 year olds). Children were recruited into four age based cohorts, two years apart, and were retested every year.

Task Complexity. Many studies that examined the structure of EF have used tasks that are complex and multi-dimensional (e.g., Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Brocki & Bohlin, 2004; Hughes et al., 2010; Lehto et al., 2003). These tasks make it difficult to pinpoint the specific skills being assessed. The Wisconsin Card Sort (WCST), and the Towers of London or Hanoi (TOL, TOH), for example, require participants to engage different EFs and auxiliary processes at different stages of the test (e.g., S. Graham et al., 2009). There may also be developmental differences in cognitive resources used for solving these tasks. Senn, Espy, and Kaufmann (2004), for example, found TOH performance of children younger than 4 was best predicted by inhibition, whilst performance of children older than 4 was best predicted by working memory. Bull, Espy, and Senn (2004) reported that the skills required for TOL changed with increasing complexity of the task, from simpler inhibitory skills at less complex levels to switching ability at higher levels, where multiple moves away from the end-goal state are required.

There is also conceptual ambiguity in how behaviour exhibited in a particular task should be classified. One of the key measures of the WCST, for example, is perseveration. Participants are said to have perseverated when they give a response that is appropriate for a previous set of rules, but no longer appropriate under the prevailing rule. Such perseveration can be understood as either failure to inhibit information that is no longer relevant, or a failure to switch cognitive strategies. By using less complex EF tasks whose performance are likely to be dominated by a singular or a smaller set of EFs (basic updating, inhibition, and switching processes), we would be in a better position to address how these fundamental skills might change with development.

Reliance on Correlational and Exploratory Techniques. Miyake et al. (2000) criticized studies which employed exploratory factor analytics to study EF without clear theoretical foundations. Although exploratory techniques help in identifying factors that provide the best separation amongst variables and that account for the most variance, the data driven nature of such analyses do not always lend themselves to clear interpretation or the explicit testing of alternative theoretical positions.

In Brocki and Bohlin (2004), for example, one of the identified factors included loadings from working memory span tasks, a Stroop-like task, and a fluency task. This factor was labelled Working Memory or Fluency. Brocki and Bohlin recognised that not having the Stroop task load with the other inhibitory measures was incongruent, but argued that it coalesced with the working memory measures because it required both inhibition of a response and a switch to a more appropriate response, both of which were argued to be inherent parts of working memory. What exploratory techniques lack is a mechanism for testing this explanation against another which specifies that the Stroop task should load with the inhibitory measures. Although the exploratory solution may present the best fit, it provides no information on the extent to which its fit differs from other theoretically meaningful models or the model with the next best fit.

Insufficient Indicators. Performing most EF tasks requires lower-order processes (e.g., motor control, visual perception). In a factor analytic setting, using multiple measures of the same construct provide the best known method for accounting for the effects of such lower-order processes and for measuring variance attributable to the underlying EF constructs. Good examples of such efforts include van der Sluis, de Jong, and van der Leij (2007) and Huizinga et al. (2006), both of whom used not only multiple indicators, but also additional tasks to index individual differences in the lower-order processes. The importance of having sufficient indicators was highlighted by Willoughby et al. (2012), who showed that only a small amount of observed variation in an EF task represents true EF ability. They argued that by using multiple tasks for each construct, the common variance of those tasks can be extracted to minimise the task impurity problem (Rabbitt, 1997).

Of more importance is that using only one task per constructs precludes the possibility of finding a multifactor solution. This is especially the case when there are only two or three constructs under examination: A minimum of two indicators are needed to achieve statistical identification for each factor, and this is only if factors are correlated. Having only one indicator also runs the risk of focusing the solution on incidental similarities across tasks (e.g., all the tasks being administered in the same format) rather than variance attributable to the construct that each task purportedly measures.

Interpretation of SEM findings. In some studies, conclusions regarding the structure of EF are drawn without consideration of alternative configurations. Hughes et al. (2010), for example, concluded that their findings supported an undifferentiated EF structure in pre-schoolers. However, they did not compare their undifferentiated model with a two or three factor models. Fuhs and Day (2011) concluded that their data supported a unidimensional model. However, as acknowledged by the authors, their finding is inconclusive as several measures failed to load in both their unidimensional and bi-dimensional models.

In other studies, the rationale for asserting that specific models provided better fit than others was not made clear. Wu et al. (2011), for example, evaluated a one, two, and three factor model. Although they concluded for a three factor structure, no inferential results were provided. From an inspection of the chi-squared fit indices, it seems clear that differences in misfit between the three and the one factor model were not statistically significant. Coupled with the lack of changes in three of the other reported fit indices, their findings should be deemed as providing better support for an undifferentiated model.

Summary

Many studies on the influence of executive functions adopted a two or three factor model of executive control without examining whether they are appropriate for the age group being studied. A number of studies are also limited by the use of a restricted age range, task complexity, and an insufficient number of indicators. In this study, we used a cohort-sequential design to examine developmental changes in 6 to 15 year olds. We addressed the following questions: a) To what extent can EF be differentiated into separate functions in childhood? b) Where different functions are found, are there age-related differences in the relations between the various functions? c) Are there age-related differences in the way in which functions are constituted?

Method

Participants & Design

Participants were 668 children enrolled in 6 kindergartens and 5 public schools serving families with low to middle SES backgrounds in western Singapore. Because of grade progression, by the end of the study, the children were distributed across 81 schools. Only children with parental consent were included in the study. The children were mainly ethnic Chinese (64%), with the remainder being of Indian (9%) and Malay (24%) ethnicities. Children were recruited as part of a larger study that examined the cognitive underpinnings of mathematical achievement. In this paper, we report only data from the EF measures.

We used a cohort-sequential design with four cohorts whose starting grades were Kindergarten II (M_{age} = 5.72, *SD* = 0.34, 98 boys, 88 girls), Primary 2 (M_{age} = 7.85, *SD* = 0.32, 87 boys, 80 girls), Primary 4 (M_{age} = 10.05, *SD* = 0.30, 78 boys, 83 girls), and Primary 6 (M_{age} = 12.32, *SD* = 0.29, 67 boys, 85 girls). Children were tested once per year over four years. With the exception of the two youngest (i.e., 6 and 7 year olds) and the two oldest age groups (i.e., 14 and 15 year olds), we collected data from two cohorts for each age group. Due to attrition, the sample size varied with each wave of data. The final sample contained 156, 143, 130, and 119 children, respectively.

Materials and Procedure

Each year, we administered a battery of tasks that were divided into five sets. Each set took 45 to 60 minutes to administer. Within each set, the tasks were scheduled in a fixed order, with all task sets administered over 2 days to several weeks, depending on the schools' logistical requirements. For the younger children, the tasks were administered individually. For the older children, the computerised tasks were administered in small groups of two to three.

Updating and working memory. We used three tasks that indexed children's updating and working memory capacity. In the Listening Recall task (Alloway, 2007), children listened to a series of sentences and assessed whether each sentence was true or false. At the end of each trial, children had to recall the last word of each sentence in the order presented. Trials contained one to six sentences. The total number of points received from recalling the final word in each sentence in the correct sequence served as the dependent measure.

In the Mister X task (Alloway, 2007), two figures were presented on-screen. Children had to decide whether each figure held a ball in the same hand and to remember the position at which the target figure held the ball. At the end of each trial, the child had to recall the position of each ball in the correct sequence. Each block had six trials progressing from a block with one pair of figures to a block with seven pairs of figures. The dependent measure was the total number of positions recalled in the correct serial order.

In the pictorial updating task, pictures of animals were shown one at a time on the computer screen. Children had to recall the identities of a specified number of animals at the end of each trial. They were not told how many animals to expect to ensure that updating was being used in the task. The number of animals presented was varied randomly across trials (Min=3, Max=11). In the easiest trials, children had to recall the last two animals; in the most difficult trials, the last four. One point was given for every animal recalled correctly.

Inhibition & Switching. Eight conditions from four computerised tasks were used to assess efficiency in inhibition and switching. In the Flanker task (modified from Fan, McCandliss, Sommer, Raz, & Posner, 2002), children were presented with a row of five fish facing either left or right with the target fish in the centre of the computer screen. The target fish appeared on its own (neutral condition), or was flanked on either side by two fish facing the same or the opposite direction (congruent or incongruent conditions respectively). In each trial, children were asked to identify, by key press, the direction the target fish was facing. The first block consisted of 20 neutral trials, followed by two pure blocks of 20 congruent, 20 incongruent trials, or vice versa to counterbalance possible order effects. This was followed by three blocks of 28 trials, in which children have to switch between congruent and incongruent trials in approximately one-third of the trials.

In the Simon task (modified from Davidson et al., 2006), a colour picture of either a frog or a butterfly appeared on the left or right side of the computer screen. Children were asked to take the butterfly home by pressing a button on the left, and likewise for the frog by pressing a button on the right. A block of 25 congruent trials was followed by a block of 25 incongruent trials, or vice versa to counterbalance possible order effects. These were followed by four blocks of 21 mixed congruent and incongruent trials. Similar to Davidson et al., we used a longer response window for the younger children.

Efficiency in switching was also assessed using a Picture-Symbol task (based on the Number-letter task, Miyake et al., 2000). In each trial, a bigram consisting of a picture and a symbol appeared in one of four quadrants on the computer screen. When the bigram appeared in the top quadrants, children were asked to identify whether the picture was an animal. When the bigram appeared at the bottom quadrants, they were asked to identify whether the symbol was a number. To reduce working memory demands imposed by having to remember which rule to apply, each time a bigram was shown, the children would hear a voice saying, “Animal?” or “Number?” In the first block of 21 trials, the bigrams appeared only in the top quadrants. This was followed by another block of 21 trials where the bigrams appeared only in the bottom two quadrants. Presentation of these two blocks was counterbalanced. In the last two blocks of 33 trials each, the bigrams appeared in all four quadrants. Order of presentation followed a clockwise rotation pattern.

We initially used a modified anti-saccade task (Nieuwenhuis, Ridderinkhof, Kok, & an der Molen, 2000) to measure inhibitory abilities. The task proved insufficiently sensitive to age-related differences and was replaced by another that required similar inhibitory responses in Wave 2 of data collection. In the Mickey Task, children were asked to look at a central fixation point. When the fixation point disappeared, squares were presented briefly on both or either side of the screen. After 200 ms, a picture of Mickey Mouse was presented on either the right or left side of the screen. Participants were instructed to ignore the initial squares and to respond by pressing a button on the same side as Mickey. The task consisted of two blocks. The first block contained congruent and neutral trials. In congruent trials, the initial squares were presented on the same side as Mickey. In neutral trials, the squares appeared on both sides of the screen. The second block contained neutral and incongruent trials. In incongruent trials, the initial squares were presented on the opposite side of Mickey. In total, there were 20 neutral, 16 congruent and 16 incongruent trials.

Reaction time (RT) and accuracy measures were collected from all inhibition and switch tasks. Inhibition RT for the Flanker, Simon, and Mickey tasks was computed from all correctly answered trials in the incongruent blocks. Switch RT were computed from trials involving conditional switches (i.e., congruent to incongruent, incongruent to congruent, animal to number, or number to animal) in the mixed blocks. For the Flanker task, we also computed separate RT measures from correctly answered trials in the (a) congruent, (b) neutral, and (c) mixed blocks (non-switch trials). For the Simon and Mickey tasks, we computed parallel RT measures from the congruent blocks and the non-switch trials.

Results

We screened all measures for missing values, outliers, and normality of distribution, with data from different conditions of each task screened separately. We first cleaned each child's trial-by-trial RT data using each child's mean and *SD* on each particular task. This approach is less arbitrary than using a fixed threshold and does not result in the wholesale exclusion of children who were particularly fast or slow overall. Data from a particular trial were deleted if they differed by more than three *SD* from that child's mean. We then computed for each child means and standard deviations for each condition. RT and accuracy scores that were more than 3 *SD* from the sample mean were replaced by values at 3 *SD*. This affected 1.4% of the data. To ensure that the data were representative of children who were performing the task properly, we included only RT data from children who were accurate on 7 or more trials per condition (i.e., a minimum of 33% to 58% of all trials, depending on task). Means and *SD* of all variables can be found online, in Supplementary Table 1.

At the end of the fourth wave, we had 83.4% of the original sample. Excluding the Mickey data from Wave 1, which was administered only from Wave 2 onwards, 14.3% of data points were missing because participants dropped out, failed to attend the whole set of assessment during a particular wave, or because of the data cleaning process described above. These missing data points were spread across different measures and different waves of data, resulting in only 245 participants with complete data. To check for systematic patterns in the missing values, we examined the number of missing values across cohorts and grades. Overall the oldest cohorts exhibited more missing values (17.8%) than did the 6 (13.1%), 8 (13.3%), or 10 (12.1%) year old cohorts. The higher value for the older children probably reflected increased school demands and higher dropout amongst that group. We also examined the frequency of missing RT data resulting from children who were unable to perform the task or who failed to attain the 7 trial criterion. Here, the younger children did exhibit more difficulties. In total, children whose data were classified as missing because they had fewer than 7 accurate trials constituted 0.7% of all data points. For these children, some of the tasks were probably too difficult. For the youngest children in particular, the data should perhaps be interpreted as representative of those who can perform the task.

Because listwise deletion of data results in both bias in analyses (J. W. Graham, 2009) and a reduction in power, tests for age-related differences were conducted with missing values imputed using the expectation-maximization algorithm as implemented in SPSS 19. A second set of analyses was conducted with only data from complete cases to check for divergence in findings. In the confirmatory factor analyses, models were estimated using data from both complete and incomplete cases with the full information maximum likelihood approach as implemented in Mplus 6.11 (Muthén & Muthén 2011).

Age-related differences

To examine age-related differences in accuracy and RT, we first conducted a series of multivariate analyses of variance. For the inhibitory and switch measures, the analyses were based on a 4 (waves: timing of data collection) x 4 (cohort: grade level at entry to the study) x 2 (task condition: congruent vs. incongruent or switch vs. no switch) split-plot design. Data from each inhibitory and switch tasks served as dependent variables. Analyses on the updating and working memory measures were based on a 4 (waves of data collection) x 4 (cohort: grade level at entry to the study) split-plot design.

Inhibitory Measures. For the Flanker and Simon tasks, both the multivariate RT, F (18, 1864.42) = 13.60, $p < .001$, $\eta_p^2 = .11$, and accuracy measures, F (18, 1864.42) = 7.77, $p < .001$, $\eta_p^2 = .07$, were affected by tertiary interaction effects. The overall findings showed reduction in RT on both a longitudinal (across waves) and a cross-sectional (across cohorts) basis (see Figure 1; descriptive statistics can be found online, in Supplementary Table 1). On the Flanker task, all children were faster in the congruent than in the incongruent condition. For each cohort, this difference reduced significantly across the four waves. For the oldest age group, though the difference between the congruent and incongruent conditions was still statistically reliable, its magnitude was negligible (M_{diff} = 10 ms).

<Insert Figure 1 about here>

Similar to performances on the Flanker task, the amount of time participants required to perform the Simon task exhibited a gradual decline. Differences between the congruent and incongruent conditions also fluctuated with age, with a larger decline in differences between the two conditions (i.e., inhibitory cost) during K2 to P3, but a levelling off from P4 onwards. Nonetheless, significant inhibitory cost was observed at each and every time-point (M_{diff} = 83 ms for the oldest age group).

As the Mickey task was not administered in the first wave of data collection, its data were analysed separately. There were significant differences across waves, cohorts, and task conditions: cohort x waves, *F* (6, 1326) = 37.43, $p < .001$, $\eta_p^2 = .15$, cohort x congruency, *F* (3, 664) = 48.26, $p < .001$, $\eta_p^2 = .18$, and waves x congruency, *F* (2, 663) = 33.89, $p < .001$, $\eta_p^2 = .09$. Follow up t tests showed significant differences between the congruent and incongruent conditions in the first wave of all four cohorts. This decreased across cohorts. For the oldest cohort, though the difference was significant, the substantive difference was small (M_{diff} = 11ms). With the youngest cohort, differences were significant across all three waves. For the other cohorts, differences across the three waves were mainly small or non-significant.

Accuracy on both the Flanker and Simon tasks was high. On the Flanker task, even the youngest children achieved accuracy of 84% and 92% on the incongruent and congruent conditions respectively. Although the accuracy data revealed age-related differences in inhibitory cost for all four cohorts, follow-up tests showed that apart from the youngest cohort, which showed a 6% reduction in cost across the first two waves, differences across the various waves in other cohorts were small (< 1%) and mostly non-significant. Accuracy data from the Simon task exhibited larger inhibitory cost with significant differences between the incongruent and congruent conditions at every time point. However, the pattern of age-related changes was less consistent and showed no clear sign of reduction with age. With performances in the congruent conditions at or above 95% in many cases, the possibility of differences being masked by ceiling effects cannot be discounted. Although the Mickey task exhibited a 3-way interaction effect, *F* (6, 1326) = 2.39, $p = .027$, $\eta_p^2 = .01$, accuracy ranged from 92.4% to 97.9% with negligible differences between the congruent and incongruent conditions: 0.005% to 0.03%.

Switch Measures. Using the same analytic strategy as the inhibitory measures, both the RT, *F* (27, 1916.50) = 6.41, $p < .001$, $\eta_p^2 = .08$, and accuracy measures, *F* (27, 1916.50) = 4.33, $p < .001$, $\eta_p^2 = .06$, revealed reliable tertiary interaction effects. For the Flanker task, the two youngest cohorts exhibited a similar pattern of findings with significant differences in RT between the switch and non-switch conditions in the first wave, followed by a sharp decline and a small increase in the last wave. The two older age groups showed no difference in the first wave, but significant differences in the subsequent waves. However, with the exception of the younger children, differences between the two conditions were uniformly small (*M*s < 22ms). The Simon task exhibited larger switch costs at all time points with consistent differences between the switch and non-switch conditions. For each cohort, there was a decrease in switch cost across the four waves. For the second youngest cohort, this decline was more modest and failed to attain significance. For the Picture-Symbol task, all age groups showed significant differences in RT between the switch and non-switch conditions across the four waves. With the exception of the youngest cohort, who exhibited a slight increase in switch cost from the first to second wave, there was a steady decline in switch cost across the four waves for all cohorts.

For the accuracy measures, all three switch tasks exhibited tertiary interaction effects. Again, these effects were found in a context where performance accuracy was very high (*M* = .91). Follow-up tests showed that, on the Flanker task, only the youngest cohort exhibited an age-related reduction in switch cost. The magnitude of switch cost, even when statistically reliable, was negligible (≤ 1%; 2.8% for the youngest children). On the Simon task, there were age-related variations in switch costs for all but the 10 year olds, but the variations in cost across consecutive waves were small (< 4%). On the Picture-Symbol task, there was age-related variation in switch costs in all four cohorts; again, the largest variation in cost across waves was small (< 4%).

Updating and Working Memory measures. The updating and working memory measures exhibited reliable secondary interaction effect, $F\ (27, 1916.50) = 40.30, p < .001, \eta_p^2 = .36$. All measures showed significant increases in capacity at most time points (the only exceptions being the 8 year old cohort, Wave 2 vs. 3, on the Listening Recall task, the 12 year old cohort, Wave 1 vs. 2 on the Listening Recall task, the 10 year old cohort Wave 3 vs. 4 on the Mr X task, and the 12 year old cohort, Wave 1 vs. 2 on the Mr X task).

Missing Data. To check whether findings were influenced by our treatment of the missing data, we re-analysed the data using a listwise deletion method. The overall pattern of findings was the same, but some of the significant differences found in the accuracy data were found to be statistically non-significant. This is likely a consequence of the smaller sample size. We deem this difference to be of little consequence as performance on the accuracy data was close to ceiling and any differences found were generally small.

Factor Structure

The previous analyses used an analysis of variance approach to examine age-related differences. Although there were similarities in pattern of differences across measures that purportedly assessed the same constructs, there were also similarities across measures assessing different constructs. In this section, we focused on individual differences in performances.

Correlations between measures of the same construct tended to be moderate (see Supplementary Table 1). However, measures derived from the same task, but which gauged different constructs (e.g., Flanker incongruent versus Flanker switch) were often as highly or even more highly correlated. Correlations between the incongruent and congruent measures, or switch and non-switch measures, from each task were also notably high. To examine whether correlations fall into patterns of constructs as predicted by previous findings, we conducted a series of confirmatory factor analyses. Five models were tested. First, we tested the fit of a three factor model containing latent factors for Inhibition, Switch, and Updating. Second, we tested three two factor models containing all possible combinations of the three factors. Third, we included an undifferentiated model in which all manifest variables loaded onto the same latent factor.

Analytic Approach – Model Specification. Previous confirmatory factor analytic studies typically utilised differences in RT between incongruent versus congruent, or switch versus non-switch, as manifest measures (e.g., Miyake et al., 2000). As demonstrated in our age-related differences analyses, these measures typically show participants taking longer in trials requiring inhibition or switch than those that do not. However, subtraction scores typically have poor inter-item reliability (Lord, 1958). This is so even when the inter-item reliability of their constituent measures are high (all our RT data have moderate to high internal reliability, Cronbach $\alpha > .72$; see Supplementary Table 1), Poor reliability can result in low correlational findings even when, in reality, substantive relations exist.

To combat this problem, one approach that was used in previous studies (Huizinga et al., 2006; Lee et al., 2012; van der Sluis et al., 2007; Van der Ven, Kroesbergen, Boom, & Leseman, 2012) was to use RT measures from incongruent conditions as manifest variables for an inhibition latent factor. Performances in the congruent or neutral conditions are mapped onto a separate latent factor that reflects simple choice RT. The incongruent manifest is then allowed to cross-load onto this speed based latent factor. In these models, values taken by the incongruent measures are deemed to result from three sources: inhibitory processes, simple choice RT, and measurement error. Though these models provided an improvement on models using subtraction scores, the use of cross-loading can result in ambiguities in the meaning of the factors on which they load.

Here, we used a modified version of models used in previous study (see Figure 2). The major innovation was that we regressed the incongruent or switch measures to their congruent or non-switch counterparts. In effect, we treated the congruent or non-switch measures as covariates. Variances in the incongruent or switch measures that could be attributed to their congruent or non-switch counterparts were attributed to the latter and do not enter into the latent constructs. Because the regression was specified at the manifest level, task-specific variances common to the incongruent and congruent, or switch and non-switch, conditions were captured at the same time. As a result, both the inhibitory and switch latent variable should provide a purer measure of their underlying constructs.

<Insert Figure 2 about here.>

Analytic Approach – Model Evaluation. When models are nested, their relative adequacy in describing the data can be tested using a number of methods (Kline, 2005). Because a two factor model contains one less latent factor than a three factor model, strictly speaking, it is not nested. However, van der Sluis, Dolan, and Stoel (2005) showed that by constraining (a) correlations between two of the latent factors to one and (b) their correlations with the third factor to equality, a constrained three factor model is mathematically equivalent to a two factor model. We used this approach to provide a direct inferential test of the relative fit of the various models that were non-constrained (i.e., three factors) versus constrained to be equivalent to two factors or one factor models.

Stoel, Garre, Dolan, and van den Wittenboer (2006) reminded us that where parameters are placed at their boundary value, its density function will not follow a canonical chi squared distribution. They argued that in comparing a three versus a two factor model, for example, we should use a critical value derived from a weighted mixture of chi squared distributions with one and two degrees of freedom.

In our one and two factor models, whenever the correlation between two of the latent factors was constrained to one, corresponding relations between the two affected latent and each covariate were also constrained to equality (van der Sluis et al., 2005). As a result, there was a difference of eight degrees of freedom between the three and two factor models, and seven degrees of freedom between the two and one factor models. Using Monte Carlo simulations, the critical values for evaluating differences in chi squared fit between the three versus two factor models and two versus one factor models were estimated to be 16.27 and 14.85 (at $p = .05$), respectively. Models were fitted using the ML estimator in Mplus version 6.11 (Muthén & Muthén 2011).

Model Comparison. Due to power consideration, data from the same grades, across different cohorts, were combined for analyses. This was possible for all but the two youngest and two oldest grades, which were sampled only once. The combined data sets for the two factor models approximated the 5:1 sample-to-parameter ratio discussed in Bentler and Chou (1987).

Data from the 6 year olds showed that though the three factor model provided a relatively good fit, it produced a solution that was inadmissible: the estimated correlation between Update and Switch was greater than one. Of the three two factor models, the one that combined the inhibitory and switch measures (IS) provided the best model fit (see Tables 2 and 3; the upper diagonal of Supplementary Table 1 provides the residuals for the best-fitting model). Though a comparison of the chi squared fit index showed significantly poorer fit as compared to the three factor model, the Akaike Information Index (AIC) showed little deterioration. Furthermore, as mentioned earlier, the three factor model was inadmissible. Comparing the two and the one factor models, all fit indices from the latter showed poorer fit. The two factor (IS) model showed that all but the Flanker inhibitory and Picture-Symbol switch measures loaded significantly onto their respective latent variables. The two latent factors, Updating and the combined Inhibition-Switch factor, were negatively and significantly correlated ($r = -.58$). Children with larger updating capacity exhibited faster RT in the inhibitory-switch conditions. In other words, children with larger updating capacity were less affected by inhibitory-switch demands. Of interest was that all covariates were strongly correlated with the combined Inhibitory-Switch factor ($r > .72$), suggesting that children with slower non-inhibitory/non-switch RT were also slower in inhibition or switching.

< Insert Tables 2 & 3 about here >

Data from the same children, tested one year later, showed a similar pattern of findings. The three factor solution was inadmissible and the two factor (IS) model provided the best fit with similar chi squared, Comparative Fit Index (CFI), and AIC values. Updating and the combined Inhibition-Switch variables were again negatively and significantly correlated (r = -.56).

For the combined cohorts, data from 8 to 12 year olds showed a similar pattern: the two factor (IS) model provided the best fit. Correlations between Updating and the combined Inhibition-Switch factors ranged from moderate to strong. One finding of note came from the 11 year olds. The three factor model was admissible and the fit indices indicated a good fit. However, the estimated correlation between inhibition and switch was very strong (r = .86). Although there was a significant chi squared difference between the three factor and the two factor (IS) models, other fit indices for the two models were near identical. For this reason, we judged the two factor (IS) model as being more appropriate. Another finding of interest was that the correlations between the covariates and the combined Inhibition-Switch factor were less consistent with the older participants. Although all correlations were still significant, most were in the weak to moderate range, with only two covariates from the Flanker task remaining very strongly correlated (r > .83). This suggests that for the older children, the inhibitory-switch processes are less strongly associated with their non-inhibitory/non- switch RT.

Findings from the 13 to 15 year olds were mixed. Fit indices from the three and two factor (IS) models showed that the three factor model provided a better fit to the 13 year olds' data. However, the correlation between the Inhibition and Switch factors was very strong (r = .85). Data from the 14 year olds showed similar fit indices for the two (IS) and three factor models, with no significant differences in their chi squared fit indices. Correlation between the Inhibition and Switch factors, though remaining strong, was more moderate (r = .72). Data from the 15 year olds showed much clearer differentiation between the three and two factor models. All indices showed better fit for the three factor model and indicated a good fit to the data. Correlation between all three latent variables ranged from small to moderate. Factor loadings for all manifest variables were significant, though modest. Correlations between the covariates and the latent variables were also smaller than those found in younger children: correlations between the covariates and Switch failed to attain significance (see Supplementary Table 2 for all parameter estimates and Supplementary Table 3 for variance explained, both available online).

Age differences. A question that we have not answered explicitly is whether there are age-related differences in the structure of EF. The findings presented so far show that there are qualitative differences in the structure of EF, with a three factor structure becoming increasingly viable from 11 years of age and which becomes more clearly differentiated at 15 years of age. A further question that can be asked is whether there are quantitative differences amongst age groups that share a similar structure. We examined this issue by conducting a test for longitudinal factorial invariance (Brown, 2006) amongst the 8 and 12 year olds.

We conducted the analysis in several stages (Byrne, 2011). First, we computed the configural model by including data from both age groups. Using the confirmatory model described earlier, the overall model fit of both age groups was estimated simultaneously. At this stage, no equality constraints were applied. This served as a baseline for comparison with models in which cross-age equality constraints were applied. In the second step, we applied equality constraints on all factor loadings across the two groups. This provided a test for whether the latent constructs (i.e., Updating, Inhibition-Switch) were constituted in the same way across the two age groups. The third step involved holding constant across age groups the covariance between the two latent constructs. This tested whether there were age-related differences in the relations between the constructs. In the fourth step, we also constrained the relations between conditions requiring versus not requiring inhibition or switch. In essence, this was a test of age-related differences in the extent to which variance in the inhibitory or switch conditions could be attributed to variance in the covariates. In the final model, we further constrained the covariances between the latent constructs and the covariates. This provided a very stringent test of the extent to which the latent constructs and the covariates varied with age.

The configural model revealed a reasonable fit to the data (see Table 4). Inspection of the modification indices revealed several sources of misfit, but none that could be implemented given the logic of the model. Imposing constraints on factor loadings produced a significant deterioration in chi squared fit, but only marginal differences on the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Cheung and Rensvold (2002) argued that the use of chi squared differences to evaluate factorial invariance is subject to the same sensitivities that prompted researchers to devise alternative measures of model fit. Based on their simulation, they recommended a .01 margin on the comparative fit index (CFI) as an alternative criterion to evaluate invariance. Using this criterion, the .006 difference in CFI between the configural and the loadings constrained model is not of significance. In the third step of invariance testing, we imposed equality constraints on the covariance between the two latent constructs: Updating and Inhibition-Switch. All indices showed minimal further deterioration to model fit. This suggests that there are no age-related differences in the relation between Updating and the combined Inhibition-Switch factor.

< Insert Table 4 about here >

Imposition of constraints on the regressive relations between measures requiring versus not requiring inhibition or switch (i.e., the covariates) resulted in more substantive differences. Inspection of the modification indexes shows several sources of misfit. For both age groups, model fit would be improved if the factor loadings of the Flanker and Simon inhibitory measures were not constrained to be equalled across groups. The constrained values were too low for the 8 year olds and marginally too high for the 12 year olds. These differences are of interest because they only appeared after the regression parameters onto the covariates had been constrained. They can thus be viewed as the influence of constraining these parameters on the factor loadings. In addition, the modification indices showed that the parameter involving Flanker inhibition being regressed onto Flanker congruent was substantially underestimated for the 8 year olds. This finding suggests that variance resulting from task specific demands and choice RT had a greater influence on the children's performance in the inhibitory condition when children were younger than when they were older.

Constraining the covariances between the latent constructs and the non-inhibitory/non-switch measures produced the greatest discrepancies. Of particular interest was that using a constant value for the two age groups underestimated the relations between the combined latent and the non-switch measures for the 8 year olds, but overestimated them for the 12 year olds. This reinforces our previous interpretation and suggests that the 8 year olds exhibited a higher correlation between the non-inhibitory/non-switch measures and inhibitory or switch abilities than did the 12 year olds.

Discussion

Updating and working memory capacity showed steady increase from 6 to 15 years with little sign of a plateau even amongst the oldest children. Previous studies have reported a plateau in performance at 14 years of age on both the phonological and executive components of working memory (Gathercole, Pickering, Ambridge, & Wearing, 2004); data from our listening recall task showed a similar pattern. The other two working memory tasks showed a decrease in growth amongst the oldest cohort, but significant year-on-year improvement was detected all the way through to the oldest age group.

Age-related changes in both inhibitory and switch cost showed significant task specific differences. On both the Flanker and Mickey tasks, we observed a rapid reduction in inhibitory cost with age. In contrast, the magnitude of inhibitory cost was larger and relatively stable across all age groups on the Simon task. Our adaptation of the Flanker task exhibited only non-significant or small differences between the switch and non-switch conditions. However, small but stable switch cost was observed on the Simon task. Much larger switch costs were observed on the Picture-Symbol task, with noticeable reduction in cost amongst the older children.

The Organisation of EF

Although age-related differences in costs and capacities show some similarities in trajectories, they do not provide a direct test of the relations amongst the measures. A unique contribution of this study is that it shows how the structure of EF varies over a ten year period from childhood to the mid-adolescent years. The confirmatory factor analyses point to a process of differentiation from a two factor structure in early childhood to a three factor structure in the teenage years. In additional to qualitative changes in whether a two or three factor model better describe the data, there are also more nuanced differences in the relation between the identified factors and the covariates (i.e., the non-inhibitory and non-switch measures). For younger children, correlations between the factors tended to be moderate, with strong to very strong correlations between the factors and the covariates. In the adolescent years, correlations with the covariates remain in the moderate to high range, but there is a marked reduction as compared to the younger children. Consistent with our finding on changes in the structure of EF, these findings suggest that efficiency in executive control becomes increasingly specialised and independent. Nonetheless, it should be noted that even among the 12 year olds, the correlations remained relatively high. Only amongst the oldest children do we see a marked reduction, especially for the switch factor. Data from the 11 and 14 year olds suggest a period of transition, during which the two and three factor models vacillated in their abilities to provide the best description for the data. It was not until the children were 15 years old when a three factor structure emerged as being clearly preferable.

Findings from the current study stand in contrast to recent studies, conducted with younger children, that pointed towards unidimensionality (e.g., Hughes et al., 2010; Wiebe et al., 2008; Willoughby, Blair, Wirth, & Greenberg, 2012). Our findings are similar to van der Ven, et al. (2012), and notably Huizinga et al.'s (2006) in which they found separation between a working memory and a task switching factor across three child age groups: 7, 11, and 15. Furthermore, their data suggest an age-related decrease in the correlation between working memory and switching. Several other studies found evidence of differentiated structures across a similar age range (e.g., Agostino et al., 2010; Brocki & Bohlin, 2004). St Clair-Thompson and Gathercole (2006) found inhibition and working memory measures loaded on separate factors, with switch measures cross-loading on these two factors. Their findings echoed our findings, which showed that in addition to the best fitting model with Inhibition and Switch combined, the second best fitting model is often the model with Updating and Switch combined.

Age- and ability-related patterns of differentiation or specialisation has long been proposed in developmental literature on intelligence (Garrett, 1946). Our findings expand upon these studies and join an increasing number of EF studies that point to an undifferentiated structure in the preschool and early kindergarten years. The transition to formal schooling sees the emergence of early differentiation. One unique contribution of our study is that it suggests that differentiation into a three factor structure occurs over a protracted period, with signs of early differentiation emerging at age 11 and reaching some stability at age 15.

There are two questions that remain unanswered: the causes of differentiation and the nature of the shared variance amongst different EFs. Although the timing of differentiation suggests that both cortical maturation and schooling may play a part, the present study provides no direct evidence on this issue. One reason that was suggested by a recent review is that working memory underpins switching efficiency (Vandierendonck, 2012). In order to switch between tasks efficiently, task sets and rules need to be instantiated and maintained in an active form. This perhaps requires a working memory capacity that is not attained until the early adolescent years. Of interest is that there seems to be a period between 11 and 15 years, during which there seems to be some vicissitudes in the factor structure. Using techniques such as latent transition analysis, it may be possible for future studies to ascertain variables that contribute to individual differences in the rate of differentiation.

Regarding the nature of the shared variance, Rose, Feldman, and Jankowski (2011) argued that processing speed is a candidate that contributes to the efficiency of all EFs. Our data showed that the EF factors are correlated very highly with some of the choice reaction time measures. As the EF factors were derived from measures from which the influence of choice reaction time had been removed, any association between the choice reaction measures and the EF factors should be due to real associations between the two constructs. Our model places all the exogenous variables, manifest or latent, on an equal footing and does not allow us to further disambiguate the relations between the choice reaction and EF measures. As it stands, the data should be interpreted as being consistent with the proposition that processing speed is a candidate process underlying the association between the various EFs.

In support of this proposition, other studies examining the relation between processing speed and EF have revealed that age-related improvements in response inhibition and working memory are largely mediated by concomitant improvements in processing speed (e.g., McAuley & White, 2011). However, having processing speed as the commonality between EFs is not uncontroversial. Other work has demonstrated that processing speed is separable from response inhibition, switching, and working memory (Span, Ridderinkhof, & van der Molen, 2004). This is an issue that requires further investigation. It may be fruitful for future studies to follow the paradigm used in the intelligence literature and investigate this question using a second level latent construct to examine the manner in which each of the EFs relate to a common contributor.

Interpretative Challenges

One challenge to our findings is that the undifferentiated Inhibition-Switch factor is an artefact of the way in which we measured the two constructs. Specifically, the two constructs fail to differentiate earlier because they are measured by a similar set of tasks. A related challenge is that the Flanker and Simon switch measures were not based on significantly varied cognitive processes; that they merely reflected whether inhibition was needed in alternating trials (Garon, Bryson, & Smith, 2008).

On the first criticism, it is important to note that the switch and inhibitory measures were generated not from the same data computed in different ways, but from independent trials presented in different blocks of the respective tasks. Although both constructs were computed by measures from the Flanker and Simon tasks, we also used non-shared tasks -- the Mickey and Picture-Symbol tasks -- to compute each construct. Nonetheless, it is possible that the use of these measures may have underestimated the age at which differentiation occurs: shared variance resulting from different measures from the same task environment may have masked differences between Inhibition and Switch. If we further assume that separation between Inhibition and Switch is smaller in earlier childhood and increases gradually with age, this mask may be sufficiently strong to occlude differences in younger children. We had acted against this possibility using three lines of control.

First, the switch and inhibitory measures were regressed onto baseline measures from the same tasks. This should have captured variance attributable to both simple choice-decision time and task based effects. Second, we included in the models shared task-based covariances. This specification serves the same goal as the correlated-uniqueness specification in multitrait-multimethod designs (Byrne, 2011) in which variances attributable to the targeted constructs versus the method with which they are measured are estimated separately. Indeed, given the possibility that shared variance may have been inflated by the use of RT measures for the inhibitory and switch tasks, we also allowed covariances amongst all the exogenous RT measures. These two features of our models should have captured much of the variance resulting from the measures being from the same tasks. Third, because the different age groups were modelled separately in the main analyses, any age-related differences in the magnitude of task based variances should also have been captured. In our modelling, we assumed independence between such task-based similarity effects on the one hand and correlations between the inhibitory and switch constructs on the other. One possibility is that they interact. Although this possibility exists even if we used tasks that differed, the likelihood is perhaps accentuated by the use of the same tasks. To test this proposition empirically will require a much larger sample size and the specification of complex interactions between the various tasks and construct related parameters. This is an issue that future study may need to consider.

On the second criticism, we acknowledge that the switch trials in the Flanker and Simon tasks are likely to evoke a smaller switch cost than tasks in which more varied cognitive tasks are used. This is borne out by the data: switch costs are substantially larger in the Picture-Symbol than in the Flanker or

Simon tasks. However, we conceptualise the two types of task as existing on a continuum that differs in the extent to which ability to switch influences performance. In the context of local switch cost, tasks that require predictable on-off switches are likely to occupy one end of the continuum, with tasks requiring random switches between different but overlapping cognitive processes on the other. In the context of age-related factor differentiation, one question that arises from this concern is whether differentiation would have been detected earlier if tasks involving larger switch costs were used. If a larger switch cost alone is indeed a prerequisite for factor differentiation, the factors should be better differentiated in the youngest children as they exhibited the largest switch cost. This was not the case.

Perhaps a more fundamental concern is whether switch conditions in the Flanker and Simon tasks are fundamentally capable of yielding measures of switch cost. Although we agree that the Flanker and Simon tasks may have underestimated switch costs, we think they are capable of yielding viable measures of switch cost. We say "capable" because data from the 15 year olds, and to a lesser extent, the 11 and 13 year olds, did conform to a 3 factor structure; the latter, alas, with strongly correlated inhibition and switch components. Looking at the issue from a task analysis perspective, the inhibition and switch measures were computed in different ways. Fundamental to the present discussion is that the switch and non-switch measures were derived from trials with the same inhibitory demands. In both the Flanker and Simon tasks, incongruent and congruent trials contributed equally to the computation of the switch and non-switch data. The only difference between the switch and non-switch measures was that for the switch measure, we selected incongruent trials that were preceded by congruent trials (or vice versa). For the non-switch measure, we selected incongruent trials that were preceded by incongruent trials and congruent trials preceded by the congruent trials. If all that is happening is differing susceptibility to interference in the congruent versus incongruent trials, and there is no switch cost associated with moving from the congruent to incongruent trials (or vice versa), the switch versus non-switch measures should exhibit similar distributions with no mean differences. Furthermore, with inhibitory trials contributing to the computation of both switch and non-switch measures, any age-related effects on susceptibility to interference should affect the constituent trials of the switch and non-switch conditions equally. There should be no age-related differences in this contrast. The finding of age-related differences in switch cost argues against this and suggests that the switch measures reflect something other than just susceptibility to interference.

A third challenge is that the differentiation of factors into Updating versus Inhibition-Switch is an artefact of differences in measurement modality: accuracy scores were used for the former, reaction time for the latter. In our model, we used a regression approach to reduce task based effects by regressing the incongruent or switch measures onto their task specific congruent and non-switch counterparts. Unfortunately, the updating tasks did not have equivalent measures that could be used for this method of control. To gain more insight into the likelihood that the findings are affected by this this problem, we performed a set of confirmatory factor analyses in which we replaced the reaction time measures with accuracy measures from four age groups that differ most widely in age (6, 9, 12, and 15 year olds). Given the ceiling effect that affected the accuracy data, it was not surprising that model fit was relatively poorer even with the best fitting models (.86 < CFI < .95). However, in all cases, the completely undifferentiated model was untenable. Even with only accuracy data in the model, we obtained separation between the Updating and Inhibition-Switch factors. Correlation between two factors ranged from relatively strong for the youngest age group (r = .79) to fairly modest in older age groups (.18 < r < .33).

Conclusions

We found the process of EF differentiation across 6 to 15 year olds to be slow, with the change from a two to three factor structure occurring over a protracted period. Our findings also show that much of the variance in our manifest measures could be attributed to processing speed. In addition, for much of childhood, inhibitory or switch ability was closely associated with processing speed. As children mature, EF and processing speed became more distinct. From an applied perspective, our findings show that only a relatively small amount of variance in each task can be attributed to EF. Furthermore, the various tasks followed different developmental time courses. These findings reinforce the view that EF tasks should not be used in isolation for discerning proficiency.

References

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology, 105*, 286-305.

Alloway, T. P. (2007). The automated working memory assessment. London: Psychological Corporation.

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 20*, 385-406.

Atkin, R., Bray, R., Davison, M., Herzberger, S., Humphreys, L., & Selzer, U. (1977). Ability factor differentiation, Grades 5 through 11. *Applied Psychological Measurement, 1*, 65-76.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78-117.

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*, 296-312.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*, 571-593.

Brookshire, B., Levin, H. S., Song, J. X., & Zhang, L. (2004). Components of executive function in typically developing and head-injured children. *Developmental Neuropsychology, 25*, 61-83.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry, 45*, 743-754.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205-228.

Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Taylor & Francis.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2 to 3. *Developmental Psychology, 40*, 1105-1122.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

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.

Espy, K. A., Sheffield, T. D., Wiebe, S. A., Clark, C. A., & Moehr, M. J. (2011). Executive control and dimensions of problem behaviors in preschool children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 52*, 33-46.

Facon, B. (2007). On the differentiation of cognitive abilities of the child. *European Review of Applied Psychology, 57*, 213-223.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340-347.

Fitzgerald, J. M., Nesselroade, J. R., & Baltes, P. B. (1973). Emergence of adult intellectual structure - prior to or during adolescence. *Developmental Psychology, 9*, 114-119.

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology, 47*, 404-416.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist, 1*, 372-378.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576.

Graham, S., Phua, E., Soon, C. S., Oh, T., Au, C., Shuter, B., et al. (2009). Role of medial cortical, hippocampal and striatal interactions during cognitive set-shifting. *Neuroimage, 45*, 1359-1367.

Happe, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain and Cognition, 61*, 25-39.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*, 1447-1459.

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*, 20-36.

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*, 2017-2036.

Hülür, G., Wilhelm, O., & Robitzsch, A. (2011). Intelligence differentiation in early childhood. *Journal of Individual Differences, 32*, 170-179.

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20*, 407-428.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: NY: Guildford Press.

Lee, K., Ng, S. F., Bull, R., Pe, M. L., & Ho, R. H. M. (2011). Are patterns important? An investigation of the relationships between proficiencies in patterns, computation, executive functioning, and algebraic word problems. *Journal of Educational Psychology, 103*, 269-281.

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology, 82*, 82-99.

Lehto, J. E., Juujaervi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59-80.

Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and Psychological Measurement, 18*, 437-451.

McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology, 108*, 453-468.

Michel, E., & Roebers, C. M. (2008). Children in regular and special needs classes: Cognitive and non-cognitive aspects. *Swiss Journal of Psychology, 67*, 249-259.

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition & Development, 13*, 395-423.

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.

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence, 38*, 611-624.

Monette, S., Bigras, M., & Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology, 109*, 158-173.

Muthén, B. O., & Muthén , L. K. (2011). Mplus (Version 6.11). Los Angeles: Muthén & Muthén

Nieuwenhuis, S., Ridderinkhof, K. R., Kok, A., & an der Molen, M. W. (2000). Inhibitory inefficiency and failures of intention activation: Age-related decline in the control of saccadic eye movements. *Psychology and Aging, 15,* 635-647.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental Psychology, 46*, 530-544.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1-38). Hove, East Sussex, U.K.: Psychology Press.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: the role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science, 14*, 1161-1175.

Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology, 26*, 445-464.

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S. C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology, 35*, 679-697.

Span, M. M., Ridderinkhof, K. R., & van der Molen, M. W. (2004). Age-related changes in the efficiency of cognitive processing across the life span. *Acta Psychologica, 117*, 155-183.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*, 745-759.

Stankov, L. (1978). Fluid and crystallized intelligence and broad perceptual factors among 11 to 12 year olds. *Journal of Educational Psychology, 70*, 324-334.

Stoel, R. D., Garre, F. G., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods, 11*, 439-455.

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*, 343-379.

Tamnes, C. K., Østby, Y., Walhovd, K. B., Westlye, L. T., Due-Tønnessen, P., & Fjell, A. M. (2010). Neuroanatomical correlates of executive functions in children and adolescents: A magnetic resonance imaging (MRI) study of cortical thickness. *Neuropsychologia, 48*, 2496-2508.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*, 1097-1118.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427-449.

van der Sluis, S., Dolan, C. V., & Stoel, R. D. (2005). A note on testing perfect correlations in SEM. *Structural Equation Modeling, 12*, 551-577.

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*, 100-119.

Vandierendonck, A. (2012). Role of working memory in task switching. *Psychologica Belgica, 52*, 229-253.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575-587.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology, 108*, 436-452.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment, 22*, 306-317.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment, 24*, 226-239.

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive Function in Early Childhood: Longitudinal Measurement Invariance and Developmental Change. *Psychological Assessment 24*, 418-431.

Wu, K. K., Chan, S. K., Leung, P. W., Liu, W. S., Leung, F. L., & Ng, R. (2011). Components and developmental differences of executive functioning for school-aged children. *Developmental Neuropsychology, 36*, 319-337.

Table 1

Confirmatory Factor Analytic Studies of Executive Functioning in Children

Study	*N*	Age (years)	Design	Factors and tasks	Findings
Willoughby, Blair, Wirth, & Greenberg (2010)	1,123	3	single time point	Working memory: Working Memory Span Inhibition: Spatial Conflict, Silly Sounds Stroop, Animal Go/No-Go Attention shifting: Item Selection	Single factor
Willoughby, Wirth, & Blair (2012)	1,123	3, 4, 5	longitudinal	Working memory: Working Memory Span, Pick the Picture Inhibition: Spatial Conflict, Silly Sounds Stroop, Spatial Conflict Arrows, Animal Go/No-Go Attention shifting: Something is the Same	Single factor
Wiebe et al. (2011)	228	3	single time point	Working memory: Nine Boxes, Nebraska Barnyard, Delayed Alternation Inhibition: Big Little Stroop, Go/No-Go, Shape School, Snack Delay	Single factor
Rose et al. (2011)	134	11	single time point	Working memory: Spatial Working Memory, Listening Span, Counting Span Inhibition: Go/No-Go, Rapid Visual Information Processing Shifting: Trail-making, Dimensional Shift	Three factors
Hughes, Ensor, Wilson, & Graham (2010)	191	4 - 6	longitudinal	Working memory: Beads Inhibition: Day-Night Planning: Tower of London	Single factor
Miller, Giesbrecht, Müller, McInerney, & Kerns (2012)	129	3 - 5	cross-sectional[a]	Working memory: Backward Digit & Word Span, Boxes, Continuous Performance task: omissions Inhibition: Dimensional Change Card Sort, Go/No-Go: hit Set shift:: Continuous Performance Task: commissions, Boy-Girl Stroop, Tower of Hanoi, Go/No-Go: commissions	Single (WM & Inhibition only)/Two-factors
Wiebe, Espy, & Charak (2008)	243	2 - 3 4 - 6	cross-sectional	Working memory: Six Boxes, Delayed Alternation, Digit Span Inhibition: Delayed Response, Whisper, Statue, Visual Attention, Shape School, Tower of Hanoi, Continuous Performance Test	Single factor
Shing, Lindenberger, Diamond, Li, & Davidson (2010)	263	4 - 6 7 - 9 10 - 15	cross-sectional	Working memory: 2 Abstract Shape, 6 Abstract Shape Inhibition: Pictures, Arrows, Dots Incongruent, Dots Mixed	Single factor (younger); two factor (oldest)
Huizinga, Dolan, & Van Der Molen (2006)	384	7, 11, 15, 21	cross-sectional	Working memory: Tic Tac Toe, Mental Counters, Running Memory Inhibition: Stop Signal, Flankers, Stroop Shifting: Local-Global, Dots-Triangles, Smiling Faces	Two factors
McAuley & White (2011)	147	6 - 8 9 - 12	cross-sectional	Working memory: Digit Span Forward/Backward, Recognition Span (shape/location), 2-back (location/letter)	Two factors (excluding speed)

		13 - 17		Inhibition: Stimulus Response Compatibility	
		18 - 24		Processing Speed: Simple Reaction Time, Stimulus Response Compatibility – compatible trials, Go/No-Go go trials	
Agostino, Johnson, & Pascual-Leone (2010)	155	8 – 11	cross-sectional [a]	M-capacity: Mental Attention Memory, Direction Following, Figural Intersection Inhibition: Antisaccade, Colour Stroop, Number Stroop Updating: Letter Memory, n-back Shifting: Contingency Naming, Trails	Four factors
Lee, Ng, Pe, Ang, Hasshim, & Bull (2012)	163	6	single time point	Working memory: Mister X, Listening Recall, Pictorial Updating Inhibition: Simon, Flanker Shift:: Simon (switch), Flanker (switch), Picture-symbol	Two factors
van der Ven	211	7 - 8	longitudinal	Working memory: Digit Span Backwards, Odd One Out, & Keep Track Inhibition: Animal Stroop, Local Global, & Simon Shift:: Animal Shifting, Trail Making Test in Colours, &Sorting Task	Two factors
Lehto et al. (2003)	108	8 - 13	cross-sectional [a]	Working memory: Auditory Attention and Response, Spatial Span, Spatial Working Memory, Mazes Inhibition: Tower of London, Matching Familiar Figures Shifting: Word Fluency, Trail Making	Three factors
Wu et al. (2011)	185	7-14	cross-sectional [a]	Working memory: Code Transmission Inhibition: Sky Search Attention, Stroop Shifting: Creature Counting, Contingency Naming, Opposite World	Three factors
van der Sluis, de Jong, and van der Leij (2007)	172	9 - 12	cross-sectional [a]	Updating: Keep Track, Letter Memory, Digit Memory Inhibition: Quantity Stroop, Object Inhibition, Stroop, Numerical Size Shifting: Object Shift, Symbol, Place, Making Trails	Two factors
Fuhs and Day (2011)	132	3 - 5	cross-sectional [a]	Inhibition: Head/Feet, Day/Night, BRIEF-P (inhibition) Shifting: Flexible Item Selection, Spatial Reversal, BRIEF-P (Shift)	Single factor

Note. [a] Age groups were collapsed for the purpose of determining the measurement model or factor structure

Table 2

Fit Indices for Measurement Models

Model	χ^2	DF	CFI	RMSEA (90% CI)	AIC	SRMR	r_{IU} or $r_{S(IU)}$	r_{IS} or $r_{U(IS)}$	r_{SU} or $r_{I(SU)}$
Kindergarten 2/6 year olds									
One factor (1)	119.62	56	0.944	0.079 (0.059 - 0.098)	7176	0.055			
Two factor (IU)	102.15	49	0.954	0.077 (0.056 - 0.098)	7172	0.053	-1.20		
Two factor (IS)	84.87	49	0.969	0.063 (0.040 - 0.085)	7155	0.042		-0.58	
Two factor (SU)					No Convergence				
Three factor (3)	67.73	41	0.977	0.060 (0.033 - 0.084)	7154	0.040	0.24	-0.82	-1.04
Primary 1/7 year olds									
One factor (1)	135.68	76	0.953	0.069 (0.050 - 0.088)	7044	0.049			
Two factor (IU)	125.57	68	0.955	0.072 (0.052 - 0.092)	7050	0.048	1.10		
Two factor (IS)	119.04	68	0.960	0.068 (0.047 - 0.088)	7043	0.042		-0.56	
Two factor (SU)	125.24	68	0.955	0.072 (0.052 - 0.091)	7049	0.044			0.70
Three factor (3)	107.59	59	0.962	0.071 (0.049 - 0.092)	7050	0.041	-0.53	1.09	-0.64
Primary 2/8 year olds									
One factor (1)	229.70	76	0.942	0.078 (0.067 - 0.090)	11894	0.059			
Two factor (IU)	220.63	68	0.943	0.082 (0.070 - 0.094)	11901	0.059	1.20		
Two factor (IS)	145.47	68	0.971	0.059 (0.045 - 0.072)	11826	0.050		-0.41	
Two factor (SU)	183.68	68	0.957	0.072 (0.059 - 0.084)	11864	0.052			0.50
Three factor (3)	133.91	59	0.972	0.062 (0.048 - 0.076)	11832	0.048	-0.41	1.25	-0.41
Primary 3/9 year olds									
One factor (1)	235.81	76	0.933	0.083 (0.071 - 0.095)	9242	0.060			
Two factor (IU)	218.24	68	0.937	0.085 (0.073 - 0.098)	9241	0.058	0.60		
Two factor (IS)	168.73	68	0.958	0.070 (0.057 - 0.083)	9191	0.046		-0.46	
Two factor (SU)	192.14	68	0.948	0.077 (0.065 - 0.091)	9215	0.049			0.43
Three factor (3)	154.87	59	0.960	0.073 (0.059 - 0.087)	9195	0.046	-0.41	0.65	-0.38
Primary 4/10 year olds									

One factor (1)	205.20	76	0.946	0.074 (0.062 - 0.086)	7389	0.051			
Two factor (IU)	191.07	68	0.949	0.076 (0.064 - 0.089)	7391	0.045	-0.86		
Two factor (IS)	135.65	68	0.972	0.057 (0.043 - 0.070)	7336	0.040		-0.55	
Two factor (SU)	163.89	68	0.960	0.067 (0.054 - 0.081)	7364	0.041			0.74
Three factor (3)	125.30	59	0.972	0.060 (0.046 - 0.075)	7343	0.035	-0.70	1.01	-0.50
Primary 5/11 year olds									
One factor (1)	195.75	76	0.948	0.073 (0.060 - 0.086)	6911	0.045			
Two factor (IU)	177.83	68	0.953	0.074 (0.061 - 0.087)	6910	0.042	0.85		
Two factor (IS)	139.25	68	0.969	0.059 (0.045 - 0.074)	6871	0.036		-0.67	
Two factor (SU)	164.99	68	0.958	0.069 (0.056 - 0.083)	6897	0.038			0.87
Three factor (3)	121.01	59	0.973	0.060 (0.044 - 0.075)	6871	0.034	-0.68	0.87	-0.63
Primary 6/12 year olds									
One factor (1)	241.71	76	0.934	0.085 (0.073 - 0.097)	6602	0.057			
Two factor (IU)	219.71	68	0.939	0.086 (0.074 - 0.099)	6596	0.053	-1.20		
Two factor (IS)	156.31	68	0.965	0.066 (0.052 - 0.079)	6533	0.043		-0.49	
Two factor (SU)	188.40	68	0.952	0.077 (0.064 - 0.090)	6565	0.052			0.37
Three factor (3)	133.12	59	0.970	0.065 (0.050 - 0.079)	6528	0.043	-0.47	1.15	-0.65
Secondary 1/13 year olds									
One factor (1)	303.01	76	0.899	0.108 (0.095 - 0.121)	5116	0.063			
Two factor (IU)	244.98	68	0.921	0.101 (0.087 - 0.115)	5074	0.059	0.83		
Two factor (IS)	190.11	68	0.946	0.084 (0.070 - 0.098)	5020	0.032		-0.30	
Two factor (SU)					No Convergence				
Three factor (3)	142.58	59	0.963	0.074 (0.059 - 0.090)	4990	0.027	-0.33	0.85	-0.24
Secondary 2/14 year olds									
One factor (1)	177.09	76	0.917	0.102 (0.083 - 0.122)	2106	0.069			
Two factor (IU)	160.88	68	0.924	0.104 (0.083 - 0.124)	2106	0.064	0.75		
Two factor (IS)	126.09	68	0.953	0.082 (0.059 - 0.104)	2071	0.047		-0.34	
Two factor (SU)	151.79	68	0.932	0.099 (0.078 - 0.120)	2097	0.057			0.65

Three factor (3)	114.74	59	0.954	0.086 (0.062 - 0.110)	2834	0.043	-0.36	0.72	-0.40
Secondary 3/15 year olds									
One factor (1)	186.27	76	0.891	0.110 (0.090 - 0.131)	1687	0.058			
Two factor (IU)	136.67	68	0.932	0.092 (0.070 - 0.114)	1653	0.057	0.27		
Two factor (IS)	140.55	68	0.929	0.095 (0.072 - 0.117)	1657	0.055		0.56	
Two factor (SU)	137.58	68	0.931	0.093 (0.070 - 0.115)	1654	0.050			0.56
Three factor (3)	87.72	59	0.972	0.064 (0.033 - 0.091)	1622	0.051	-0.54	0.15	-0.29

Note. DF = degree of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike information criterion; SRMR = standardized root mean residual. rIU: for the three factor models, correlation between two factors, in this case, Inhibition and Updating; rS(IU): for the two factor models, correlation between the combined and the remaining factor, in this case, Switch and the combined Inhibition and Updating factor.

* $p < .05$.

Table 3

χ^2_{diff} *for Measurement Models*

	Age									
Models	6	7	8	9	10	11	12	13	14	15
1 vs. IU	17.47*	10.11	9.07	17.57*	14.13	17.92*	22.00*	58.03*	16.21*	49.61*
1 vs. IS	34.75*	16.64*	84.23*	67.08*	69.55*	56.50*	85.41*	112.90*	51.00*	45.73*
1 vs. SU	—	10.43	46.02*	43.67*	41.30*	30.76*	53.31*	—	25.30*	48.70*
1 vs. 3	51.89*	28.09*	95.79*	80.94*	79.90*	74.74*	108.59*	160.43*	62.35*	98.55*
IU vs. 3	34.43*	17.98*	86.72*	63.37*	65.77*	56.82*	86.59*	102.41*	46.14*	48.95*
IS vs. 3	17.14*	11.45	11.56	13.85	10.35	18.24*	23.19*	47.54*	11.35	52.83*
SU vs. 3	—	17.66*	49.77*	37.26*	38.60*	43.98*	55.28*	—	37.05*	49.86*

Note. 1 = One factor; IU = Two factor, Inhibition and Updating combined; IS = Two factor, Inhibition and Switch combined; SU = Two factor, Switch and Updating combined; 3 = Three factors.
$^* p < .05$.

Table 4

Test of age-related Differences in Factor Structure Amongst 8 and 12 Year Olds

Model	Sources of constraints across age groups	χ^2	df	p	$\Delta\chi^2$	Δdf	$p(\Delta\chi^2)$	RMSEA	CFI	SRMR
1	Configural: no constraints	301.77	136	<.001				0.062	0.968	0.047
2	Factor loadings	339.82	145	<.001	38.05	9	<.001	0.065	0.962	0.071
3	Covariances between the latent factors	340.49	146	<.001	38.71	10	<.001	0.065	0.962	0.071
4	Covariances between conditions requiring versus not requiring inhibition or switch	396.47	153	<.001	94.70	17	<.001	0.071	0.953	0.101
5	Covariances between the covariates and latent factors	478.14	167	<.001	176.37	31	<.001	0.077	0.94	0.24

Note. All constraints were cumulative: each higher up model contains constraints applied in additional to those applied in previous models

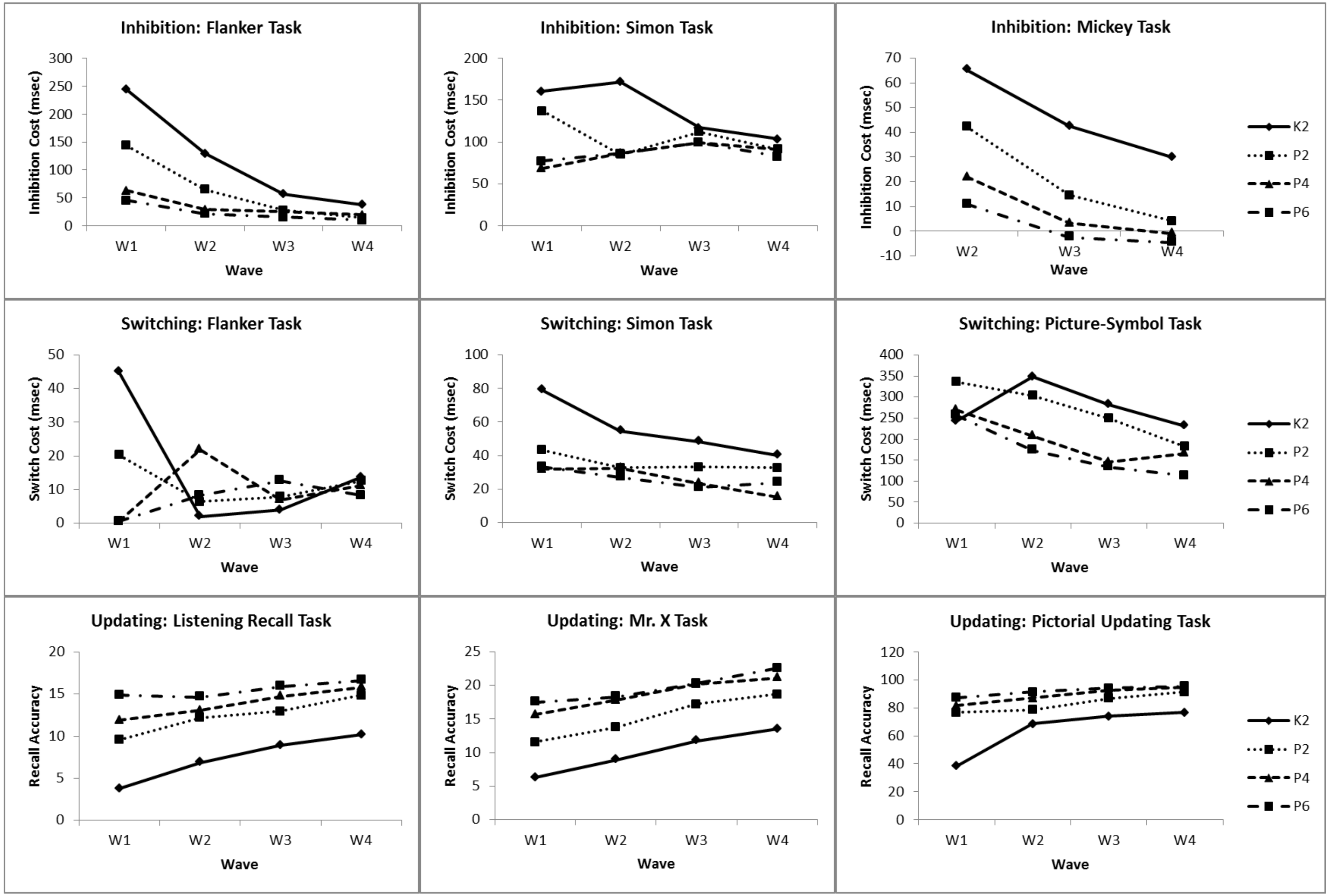
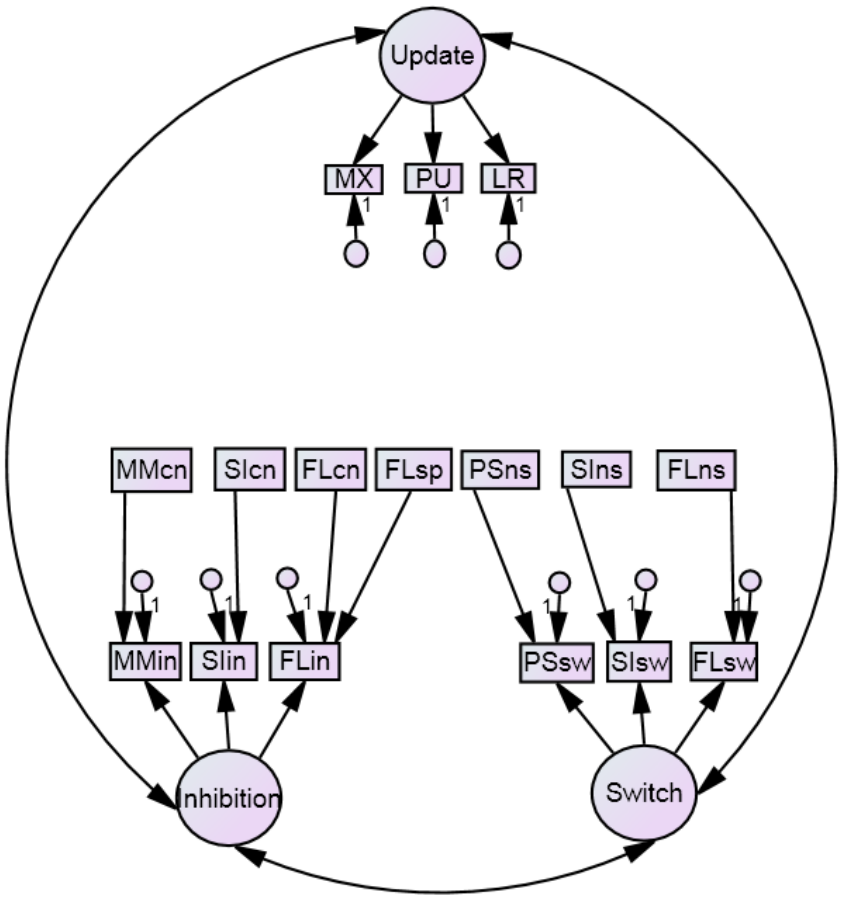


Figure 1. Age-related differences in mean performance on the inhibition, switch, and updating/working memory tasks. K2 = Kindergarten 2 cohort; P2 = Primary 2 cohort; P4 = Primary 4 cohort; P6 = Primary 6 cohort. Inhibition cost = Incongruent RT – Congruent RT, Switch cost = Switch RT – Non-switch RT



Parameter Estimates for Best Fitting Model

Parameter	6-year-old	7-year-old	8-year-old	9-year-old	10-year-old
Factor loadings[a]					
INH By FLin	0.115	0.936**	0.864***	0.613***	0.762***
INH By SIin	0.613***	0.096	0.206***	0.101	0.125*
INH By MMin	—	0.323***	0.473***	0.354***	0.396***
SW By FLsw	0.111*	0.184*	0.229***	0.516**	0.417**
SW By SIsw	0.332***	0.155***	0.163***	0.082**	0.136***
SW By PSsw	-0.016	-0.008	0.096**	0.075	0.101*
UPD By LR	0.365***	0.491***	0.689***	0.402***	0.576***
UPD By MX	0.656***	0.542***	0.509***	0.381***	0.534***
UPD By PU	0.403***	0.387***	0.526***	0.842***	0.687***
Correlations between latent factors					
INH With SW	1.000	1.000	1.000	1.000	1.000
INH With UPD	-0.581***	-0.564***	-0.407***	-0.455***	-0.552***
SW With UPD	†	†	†	†	†
Parameter	11-year-old	12-year-old	13-year-old	14-year-old	15-year-old
Factor loadings[a]					
INH By FLin	0.515***	0.728***	0.661***	0.611***	0.548***
INH By SIin	0.361***	0.143*	0.282***	0.185**	0.455***
INH By MMin	0.417***	0.483***	0.499***	0.539***	0.512***
SW By FLsw	0.072	0.316***	-0.073	0.866*	0.310***
SW By SIsw	0.302***	0.132***	0.703**	0.245***	0.140**
SW By PSsw	0.194***	0.184***	0.500***	0.116	0.481***
UPD By LR	0.572***	0.432***	0.720***	0.576***	0.819***
UPD By MX	0.469***	0.510***	0.533***	0.509***	0.640***
UPD By PU	0.652***	0.834***	0.662***	0.668***	0.492***
Correlations between latent factors					
INH With SW	1.000	1.000	0.846***	1.000	0.149
INH With UPD	-0.674***	-0.492***	-0.327***	-0.340**	-0.535***
SW With UPD	†	†	-0.244**	†	-0.289*

Figure 2. Specification of the 3-factor model. Latent factors (Update, Inhibition, and Switch) are presented in circles. UPD = Update; INH = Inhibition; SW = Switch. Manifest variables are presented in rectangles. For all manifest variables, the first two letters indicate task names, the following two letters indicate task conditions: FL = Flanker task; SI = Simon task; MM = Mickey task; PS = Picture-Symbol task; LR = Listening Recall task; MX = Mister X task; PU = Pictorial Updating task; in = incongruent condition; sw = switch condition; cn = congruent condition; ns = non-switch condition; sp = neutral condition; † = for these models, Inhibition and Switch were combined into one factor..

[a]Parameters are standardised factor loadings. The first variable refers to the latent, the second refers to the manifest variable. Not depicted in the figure are covariances between all exogenous variables, which were included in the model.

*p < .05. **p < .01. ***p < .001.