Current Psychology
https://doi.org/10.1007/s12144-022-03974-3

Fractionation of executive functions in adolescents from Iran: invariance across age and socioeconomic status

Isis Angelica Segura[1] · Sareh Panjeh[2] · Hugo Cogo-Moreira[3] · Ali Nouri[4] · Monica Carolina Miranda[5] · Fattaneh Esmaeili[6] · Heliya Seyedi[6] · Sabine Pompéia[1]

Accepted: 30 October 2022

Abstract

Executive functions (EFs) are cognitive skills that regulate thoughts and behavior. The seminal EF unity and diversity theoretical framework proposes the existence of three correlated EF latent domains (inhibition, updating, and switching) that become distinguishable from a certain moment during adolescence, but it is unclear how age and socioeconomic status (SES) affect these abilities. Here, we assessed 407 9-15-year-old Iranians of variable SES using an open-access battery of executive function tests that includes two tasks of each EF domain and allows for sociocultural adaptations regarding language and stimuli. Various EF model configurations proposed in the literature were tested (one, two and three EF latent factor, nested and bifactor-S-1 models) using confirmatory factor analyses. In addition, to explore the unbiased effects of age and SES, we performed invariance testing (across age and SES) using multiple indicators multiple causes (MIMIC) model to the best fitting model solution. The three-correlated EF factor model had the best fit and was mostly invariant across age and SES, with all three EF latent traits improving with age, while SES exerted only minimal positive effects on shifting and updating. We concluded that the three separable EF domains, found in adults and adolescents of other ages from different populations, can already be detected from the beginning of adolescence when culturally and psychometrically appropriate EF tasks are used. Additionally, these abilities continue to improve with age and are little affected by SES, suggesting that the unity and diversity framework is useful to study the cross-country generality of EF development.

Keywords Executive functions · Human development · Socioeconomic factors · Inhibition · Switching · Updating

Introduction

Executive functions (EFs) encompass several cognitive skills associated with the concept of controlled attention that are responsible for regulating thought and behavior to reach goals that are in people's minds at a given moment (Friedman & Miyake, 2017). EFs are of interest to many fields because they are associated with current and future physical and mental health and well-being, as well as other issues such as academic and professional success (Moffitt et al., 2011).

The unity and diversity model of EF

A seminal theoretical framework in the study of EF, the unity and diversity model of EF (Miyake et al., 2000), initially suggested the existence of three correlated ("unity") executive domains that were also statistically separable ("diversity") in adults. These domains are: (1) inhibition (of prepotent responses), the ability to inhibit automatic behaviors; (2) updating, the ability to continuously replace the content in working memory that is no longer relevant with new information necessary to achieve given goals; and (3) switching (or shifting), the ability to switch between different tasks. The unity and diversity aspect of this model was confirmed

✉ Hugo Cogo-Moreira
hugo.c.moreira@hiof.no

1 Departamento de Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil

2 Departamento de Psiquiatria, Universidade Federal de São Paulo, São Paulo, Brazil

3 Department of Education, ICT and Learning, Østfold University College, Halden, Norway

4 Department of Education, Malayer University (UM), Malayer, Iran

5 São Paulo, Brazil

6 Malayer, Iran

Published online: 01 December 2022

using confirmatory factor analysis (CFA), which requires data from two or more tasks that tap each of the three domains. This allows latent scores for each domain to be obtained, which represent the variance that is shared between tasks, and thus circumvents problems of task impurity (Friedman & Miyake, 2017), that is, the contribution of other, non-executive cognitive abilities to test scores, which vary from task to task.

The same team of researchers who put forward this EF model found in subsequent studies that another model specification (see Friedman & Miyake, 2017), which has been called a "nested factor" model (Karr et al., 2018), had a better explanatory power in adult samples. In this alternative model, the common EF variance (that is, the shared variance among the three domains in the initial three factor model) was almost perfectly correlated with the inhibition latent factor. In this new bifactor specification, all tasks load on this common factor, akin to inhibition ability, which purportedly represents the ability to form, maintain, monitor, and apply goals (Friedman & Miyake, 2017). However, performance in switching and updating tasks in this later model load on specific factors that are not correlated with the common factor, capturing the remaining variance that is not accounted for by the common EF factor. There is, however, a great deal of controversy surrounding which model structure best explains the unity and diversity of EFs, as explained below.

Diversity of EF factor structures in the literature and the effects of age

A meta-analysis by Karr et al. (2018) showed that the most replicated factor solutions vary from age to age. When the three domains are tested, the most commonly replicated structure in adults are the three-correlated and the nested factor solutions. In contrast, in children (preschool) only one factor is usually observed, indicating that the EF domains are not yet separable at this age (Karr et al., 2018). During adolescence, which is the age of interest here, the three factors have been found to be psychometrically separable (Duan et al., 2010; Engelhardt et al., 2015; Hartung et al., 2020; Lee et al., 2013; Wu et al., 2011; Xu et al., 2013; Zanini et al., 2021; for a review see Karr et al., 2018). However, it is unclear if this fractionation is already present at the onset of adolescence, if it occurs later, in the first half of adolescence which encompasses the pubertal transition (until around 15 years of age), or if it is found only later on, close to adulthood. This is because studies vary widely in terms of the age range of their samples. Also, some studies that include adolescents report only a single-factor (e.g., Xu et al., 2013) while others two-factor models (inhibition merged with switching: Lee et al., 2013; updating/working memory merged with inhibition: Huizinga et al., 2006; van der Sluis et al., 2007).

These inconsistencies among previous publications in terms of the factor structure of EF might be related to theoretical misconceptions. The development of EFs is often demonstrated with the use of tasks that are not representative of the cognitive domains in the original model (Morra et al., 2018). For instance, working memory capacity tasks are often regarded as measures of updating, and cognitive flexibility tasks are misconstrued as assessing switching (e.g., Diamond 2013). In addition, the different age ranges investigated in each of these studies may be the reason for the discordant findings regarding factor structure. Intense cognitive maturation, and structural and functional brain transformations take place during adolescence, partly due to pubertal changes (Goddings et al., 2019). Therefore, it is possible that the three domains can only be captured as separable after a particular stage in adolescence is reached, but this has not yet been determined. There is previous evidence of the fractionation of the three-correlated EFs in early adolescence (up to 15 years of age) using the Free Research Executive Function Evaluation (FREE) (Zanini et al., 2021), an open-access battery of executive function tests, which includes tasks drawn from the literature and adapted for use in diverse, non-English speaking populations to measure the unity and diversity of the EFs as originally proposed by Miyake et al. (2000). However, the effects of age are not clear in this literature. To establish this, invariance across age must first be determined, and this has not been tested for in models with more than one EF factor. Evidence of invariance across age means that the constructs that are under investigation, in the present case, EF, are equivalent regardless of the participants' age (Brown, 2014). This is, therefore, a mandatory step (Meredith, 1993) before analyzing whether and/or how EF develops as adolescents become older.

Other limitations in the literature

Even if tasks are adequately selected and age is found to be invariant, advances in this field have other constraints such as psychometric issues that can impact the factor structure that is found. Karr et al. (2018) point out that there are many limitations in published studies not only regarding the large variety of tested factor structures, but also relating to the low rates of convergence of the models and the dearth of models that meet acceptable fit thresholds. These limitations are generally due to low power and/or variety in analytical modelling decisions, such as whether or not to include residual correlations and other modifications that can improve convergence and model fit. Most of the studies reviewed by Karr et al. (2018) also presented only one or a few possible factor structures, so it is impossible to tell how well other model specifications might have accounted for the data. Another issue that can, in part, explain the variability in results across studies (Karr et al., 2018) is that publications in this field differ in the choice of the use of speed or accuracy as the dependent variable in different tasks. These analytical decisions are not usually based on a clear rationale, so that differences in the variability of these types of data and task-specific and/or individual differences in speed-accuracy trade-off (Zanini et al., 2021) could potentially interfere in the determination of the latent factors.

Cultural and socioeconomic influences on EFs

Importantly, to establish the best EF factor structure in samples from different countries, tasks and stimuli must be adapted to the sample's sociocultural context. The reason for this is that cultural practices likely influence information processing and responses to stimuli (Henrich et al., 2010) due to differences in language, numeral systems, social norms and various other factors. This must be considered because classic tasks used to measure EFs were created for samples from developed countries which, besides presenting intrinsic cultural specificities (Haft & Hoeft, 2017; Henrich et al., 2010), often do not include participants from a wide range of different backgrounds and/or who have fewer opportunities in life. One of the many ways of measuring this is to consider socioeconomic status (SES), which varies a great deal in developing countries, such as Iran, where the present study took place. It is essential that this variability is taken into account when investigating the effects of SES on EF (see Lawson et al., 2018). Low SES has been shown to negatively affect the development of EFs and other cognitive abilities across childhood and adolescence via a variety of mechanisms (Foulkes & Blakemore, 2018; Haft & Hoeft, 2017), including impaired structural and functional brain maturation and neuroendocrine profiles due to exposure to factors such as chronic stress and poor health. Apart from these possible negative physiological effects as a result of being from an underprivileged family, low SES is also related to insufficient cognitive stimulation, usually secondary to an inadequate school environment (Foulkes & Blakemore, 2018; Haft & Hoeft, 2017). The effects of insufficient cognitive stimulation on EFs can be associated with having fewer opportunities to employ EF (e.g., in academic contexts) and/or because highly cognitively stimulated individuals may have better developed lower-level cognitive skills that can influence performance in (higher order) EF tasks, such as a larger vocabulary (Lurie et al., 2021). Furthermore, parental factors such as education and/or positive guidance (Foulkes & Blakemore, 2018; Haft & Hoeft, 2017) also influence the development of EFs. For example, lower levels of parental schooling have been associated with difficulties in inhibition and switching (McNeilly et al., 2021) and working memory task performance (Montroy et al., 2019). This factor is believed to have a significant impact on cognition, to the extent that parental schooling has often been used as a proxy for SES in studies with underaged samples (Foulkes & Blakemore, 2018; Last et al., 2018; Montroy et al., 2019). Therefore, ideal EF tasks should minimize the possibility of bias that result from participants being or not from more affluent families. Additionally, as mentioned in respect of the effects of age, invariance across SES must be ensured before the extent of the effects of SES at the EF latent trait level are established.

The present study

Considering all the above, the best structure that explains the fractionation of EF in the first half of adolescence and the effects of age and SES can only be determined if: (a) tasks are carefully chosen so that they tap inhibition, updating and switching as conceptualized by the authors of the model, and there are at least two tasks of each domain to enable the determination of EF latent scores; (b) the tasks are appropriate in respect of the characteristics of the tested samples in terms of age, culture/language and SES; and (c) data are analyzed considering different model specifications with up-to-date psychometric models and invariance testing (Meredith, 1993) ascertained across age and SES. A test battery that fits these criteria is the Free Research Executive Function Evaluation (FREE; Zanini et al., 2021). The FREE is a non-automated open-access battery that contains representative tests of the three domains of EF proposed by Miyake et al. (2000) and can be adapted for use in samples from distinct cultural and socio-economic backgrounds. Moreover, the tasks and instructions are simple and include easily-recognizable stimuli that can be chosen to fit the characteristics of the to-be-tested populations in terms of language and familiarity with the stimuli. Furthermore, responses are vocal and the tasks are self-paced, with no time limit to respond to each stimulus. This helps to avoid statistical distortions due to the interference from inter-individual variability in psychomotor and perceptual speed that are observed when key presses and/or time limits are imposed (see Zanini et al., 2021). The adequacy of the FREE battery has been assessed in a sample of Brazilian adolescents (Zanini et al., 2021) and it was shown to reflect a three-correlated factor solution that was also found in adult and some other adolescent samples (Karr et al., 2018). However, other model specifications and invariance to age and SES were not tested.

The aim of the present study, therefore, was to determine which among many alternative model specifications (reviewed by Karr et al., 2018) is the best fitting one in the first half of adolescence in a sample from Iran, a developing country with high variability in terms of SES and low quality education (Programme, 2020) compared to developed nations. We also investigated whether the best fitting model was invariant across age and parental schooling. We hypothesized that the three-correlated factor solution would be the best fitting model because it was the one that has been most often reported at this age (Karr et al., 2018), especially in studies that used representative tasks of the EF domains of interest (Duan et al., 2010; Engelhardt et al., 2015; Wu et al., 2011; Xu et al., 2013; Zanini et al., 2021). We also expected to obtain evidence of invariance in terms of age and SES because the tasks were devised to use highly familiar stimuli that should not be more easily processed by older and more highly cognitively stimulated individuals. We also hypothesized that it would be possible to show age-related EF improvement because these abilities mature at this age (Foulkes & Blakemore, 2018) but that SES would have no or minimal effect due to the selection of highly familiar stimuli.

Materials and methods

Participants

We tested a sample of 407 (170 girls) 9 to 15-year-old native Persian-speaking Iranians who were drawn from public and private schools in Tehran, Iran. Because paying for private, usually higher quality schools in Iran, is only possible for families with higher SES, we included students from both types of schools in order to achieve a more representative range of SES in the sample. Participants were enrolled in the local equivalent of the United States grades 4 through 9 and had normal or corrected vision. They were included if their guardians provided informed consent and reported that they had no neurodevelopmental and/or mental disorders.

Procedures

All the procedures of this cross-sectional study were approved by the Institutional Ethics Committee of the Education Office of Tehran (Approval code: D/100/10,247; DATE: 2018-11-01). Participants were randomly selected by multi-stage cluster sampling based on city districts and private/public school attendance as this reflects higher/lower SES. To this end, the city of Tehran was divided into four districts - north, south, east and west. From each of these four districts, eight schools were selected: four primary schools of grades 4, 5, and 6 (two private schools, one for girls and one for boys; and two public schools, one for girls and one for boys) and four secondary schools of grades 7, 8, 9 and 10 (one private and one public school for girls, and one private and one public school for boys). Fifteen students from each school were tested (a total of 480 students, 240 boys and 240 girls). Students were provided with a brief explanation about the aims of the study. Following local ethical guidelines, selected individuals whose parents gave their written consent and who met the eligibility criteria were included in the study (if not, another individual with similar characteristics was randomly selected). Of the initial 480 students selected through this cluster sampling technique, 53 were excluded due to problems with data collection. Participants were tested individually at their schools and provided information on their age, sex and their guardians/parents' schooling. The examiners were trained to administer and correct the tasks following a test manual based on the procedures described by Zanini et al. (2021) (available at https://osf.io/px84t/?view_only=c42ee8e677e94f85a618bb2640c12b5ct) which was initially written in Portuguese by Brazilian researchers, translated by a native speaker into English and, from English, into Persian (Farsi) by the Iranian researchers. Due to school scheduling, the time of day in which the tests were administered varied from student to student. This was not controlled for because circadian effects on EF are not clear in this field.

Demographic measures

Age and parental schooling

The participants reported their date of birth, which was used to determine their age in months at testing. They also selected their parents' level of education following alternatives based on Last et al. (2018): 1 = did not complete high school; 2 = finished high school; 3 = completed two years of higher, tertiary education; 4 = completed tertiary education; 5 = master's degree or above. The mean value for both parents/guardians (or value of one if there was only one caregiver) was used as a proxy for SES (Last et al., 2018).

Cognitive measures

FREE battery

The FREE battery contains six tests that were adapted versions of tasks mainly drawn from international studies that assessed the unity and diversity framework, two for each of the three domains (inhibition, switching, and updating). This battery of tasks is non-automated and self-paced, and can be adapted in terms of instructions and choice of stimuli so as to be appropriate for samples with different SES, schooling levels, and cultural contexts. By cultural contexts, we mean that not only can the language regarding the task instructions be changed, but the order that the text/stimuli is presented (e.g., from left-to-right as in European languages to right-to-left as in Middle Eastern languages such as Persian/Farsi) and the stimuli themselves can be changed so that they represent concepts familiar to the tested population. Details on the rationale for the choice of tasks, the stimuli used, a description of how to administer and to score performance are available in Zanini et al. (2021). A description of the tasks can be found in Table 1;

Table 1 Description of the self-paced executive function tasks per domain and scoring methods (based on Zanini et al., 2021)

Domain (Task)	Paradigm	Scoring
Inhibition		
(Victoria Stroop)	Contains 3 blocks, each of which consists of 24 stimuli (color patches or words) displayed on a single screen. Participants name the ink color of patches (block 1), words that are not color names (block 2) and words that are color names written in incongruous ink colors (block 3). Block 1 is the control block, measuring speed to name colors. Blocks 2 and 3 depend on participants' reading skills. Block 2 involves some inhibition (avoiding reading the words to name their ink colors), and block 3 involves the most inhibition (naming ink colors instead of reading color names due to competition of lexical activation of word and color hues). Adapted from Strauss et al. (2006).	Inhibition cost: RCS of block 2 minus RCS of block 1 (non-incongruous inhibition cost), and RCS of block 3 minus RCS of block 1 (incongruous inhibition cost).
(Happy-Sad Stroop)	Contains 3 blocks, each of which consists of 20 facial emotions displayed on a single screen. The first block is just to automatize emotion naming and is not scored. In block 1, participants name the emotions (happy or sad). In block 2, participants inhibit naming the emotion they see and must instead name the opposite one (happy as sad or vice-versa). Adapted from Kramer et al. (2015) and Lagattuta et al. (2011).	Inhibition cost: RCS of block 2 minus RCS of block 1
Updating		
(2-Back)	Each screen contains 10 square outlines in fixed locations, one of which is filled in with black ink. As participants pass from screen to screen, they answer if the black square location they see is in the same or a different location as the black square two screens back, continuously updating these positions in working memory as the task progresses. The total number of updating opportunities is 66. Adapted from Friedman et al. (2008).	Total RCS (for accuracy in this case only: hits minus false alarms) (no control block)
(Number Memory)	Each screen contains a single digit number (1 to 9). As participants pass from screen to screen, they report the last three seen digits (trios), in the same order as they were presented, having to continuously update information held in working memory (discarding the first digit of the trio and adding the new digit that appears next, and so forth). The total number of updating opportunities is 24. Adapted from Miyake et al. (2000).	Total RCS (no control block)
Switching		
(Color-Shape)	Contains 3 blocks in which single-colored geometric pictures are presented on each screen (trial). As participants pass from screen to screen, pictures must be classified by shape (squares/circles) (block 1: 20 trials), by color (black/gray) (block 2: 20 trials) or alternating (switching) classifications (block 3: 40 trials) according to cues presented on top of the pictures (abstract shape for shape, rainbow for color). Adapted from Miyake et al. (2004).	Switching costs: RCS in block 3 minus the sum of RCS in blocks 1 and 2
(Category Switch)	Contains 3 blocks in which single pictures are presented on each screen (trial). As participant pass from screen to screen, each picture must be classified as living or non-living (block 1: 20 trials), big or small (block 2: 20 trial) or alternating (switching) classifications without cues (living/non-living, then big/small, and so forth) (block 3: 40 trials). Adapted from Friedman and Miyake (2004).	Switching costs: RCS in block 3 minus the sum of RCS in blocks 1 and 2

Participants themselves passed from screen to screen using the arrow on the keyboard (self-paced tasks) and their answers were vocal. The examiner wrote down the vocal answers and recorded the time taken to complete the task using a stopwatch. Scores used in the analyses were Rate Correct Scores (RCS) and were obtained by dividing accuracy (vocal responses) per the total time taken to complete blocks/task timed by the experimenter. See Fig. 1 for a visual illustration of the tasks

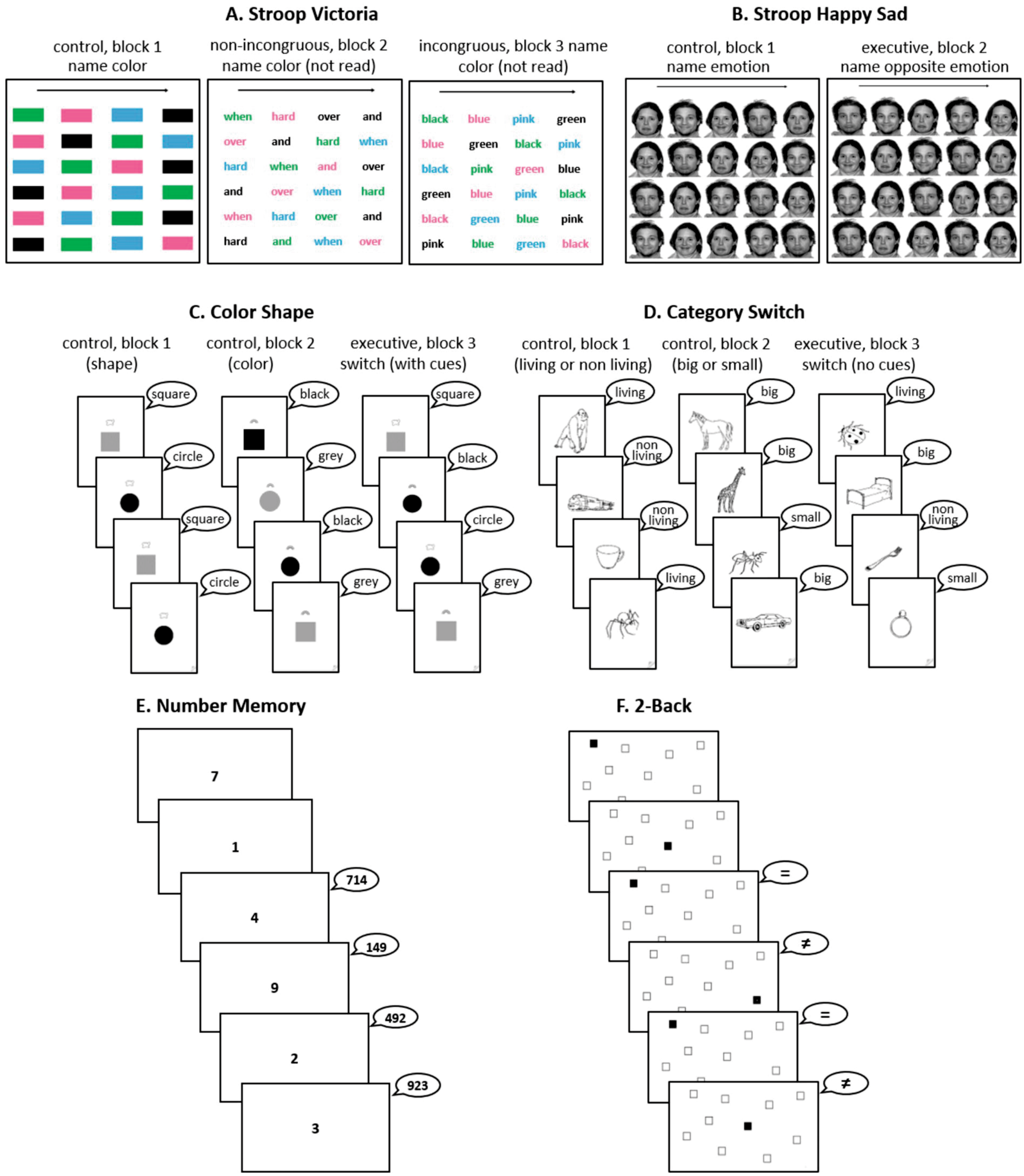

Fig. 1 Illustration (in English, although tasks were administered in modern Persian (Farsi) of the two tasks of each executive functions domain: inhibition (**A** and **B**), switching (**C** and **D**) and updating (**E** and **F**). See Table 1 for details of the tasks

Fig. 1 (for additional details, see https://osf.io/eg4vc). All the stimuli used were the same as those used with Brazilian adolescents in the study by Zanini et al. (2021), except for some color changes in the Victoria Stroop Task. In other words, no other adaptations had to be made regarding the stimuli, which were found to be appropriate in pilot studies with Iranian adolescents. The instructions were written in Persian and the stimuli and text were presented right-to-left.

All task instructions and written stimuli were presented in Persian and are available at https://doi.org/10.17605/OSF.IO/2BX8N, together with details of their administration and scoring in Persian, English, and Portuguese.

The executive tasks were administered using laptops in four different pseudorandom orders alternated with other tasks and behavioral questionnaires which will not be addressed here (e.g., Nouri et al., 2021). To carry out the EF tasks, the participants read the instructions or had them read to them if they preferred. The examiner clarified any questions that arose. The students then briefly practiced four trials for each block of the switching tasks and two blocks with five trials regarding the updating tasks to ensure they understood the instructions (except for the inhibition tasks, for which there were no practice trials, following methods used in other studies, although sample stimuli were shown in the instructions). Practice trials were kept to a minimum to ensure participants understood the tasks and yet did not become proficient, which reduces the use of EF. If the instructions were understood, the participant carried out the task. If not, the instructions were explained again until the students were able to perform the practice trials correctly or reported having understood what they were supposed to do. Participants themselves passed from screen to screen using the arrow on the keyboard (self-paced tasks) and their answers were vocal. They were asked to complete the tasks as quickly as they could while avoiding mistakes. The blocks/tasks ended after the response to the last stimulus (there were no interruption criteria). During the tasks, the examiner wrote down the vocal answers and recorded the time taken to complete the task using a stopwatch. Sessions were recorded with the participants' consent and erased once adequate scoring was ensured. Breaks between blocks and tasks were offered and taken if the students asked for them.

Instead of using reaction time or accuracy measures as done by the majority of prior publications in this field, we used the Rate Correct Score (RCS) (for details, see Zanini et al., 2021). In this measure, the number of correct answers is divided by the time (in seconds) taken to complete each block (for inhibition and switching tasks) or task (updating tasks). This was done because the RCS controls for speed-accuracy trade-offs, and no guidelines are available in the literature on when and why to select accuracy or reaction times (see Karr et al., 2018).

There was only one difference in scoring compared to Zanini et al. (2021), namely that scoring for the 2-back updating task was altered because participants can guess (without actually updating the content in their working memory) by responding that all spatial configurations are either different or the same as the one presented two trials back (only 36% of trials are the same). Hence, accuracy in this case was calculated as hits minus false alarms (Hartung et al., 2020).

The variables used in the CFA models (see below) for the inhibition and switching tasks were RCS absolute executive costs, that is the extent to which adding executive requirements to the tasks in the last blocks reduced performance compared to the control blocks (in other words, "excluding" the use of other cognitive, perceptual and psychomotor abilities that contribute to performance). This was determined by subtracting the RCS in block 3 from the RCS in the control blocks. Hence, for inhibition and switching costs, a lower RCS indicated that the costs of executive requirement of the last block were smaller, and therefore, executive abilities were better. Conversely, a higher RCS indicated better performance in the updating tasks, because as these tasks did not include a control block, in these cases total scores were used as per the literature. To facilitate the interpretability of the results in the CFA models, we reversed the results signs of the inhibition and switching tasks so that higher scores would represent better performance, and correlations among latent factors would all be positive.

Statistical analyses

The sample size was determined based on data from Zanini et al. (2021), that is openly available, considering the three-correlated factor model which we hypothesized it would be replicable. A Monte Carlo simulation analysis evaluating the power (i.e., 1-β) and other parameters showed that a sample of 320 would be sufficient (for details see Supplementary File).

Descriptive statistics were determined for all raw and RCS measures (RCS per block in each task and RCS inhibition and switching absolute costs) using IBM SPSS Statistics version 21 software for Windows (IBM Corp., Armonk, NY, USA). Data were inspected for outliers (values over three SD of the mean) which were substituted for values of the mean plus three SD (following Friedman et al., 2016; Miyake et al., 2000). To confirm that the inhibition and switching tasks actually displayed executive costs, that is, the last block had higher executive requirements, we used SPSS within-participant General Linear Models (GLM) with the factor Block. The level of significance was 5%.

Inferential analysis was carried out in three general steps. First, we determined how the scores in the tasks were interrelated considering the whole tested sample. To do this, we based the analyses not on the raw scores of the EF tasks but on the latent factor scores, which represent the common variance in scores obtained using different tasks which are regarded as purer EF measures. To find the best pattern of interrelations among the possible latent factors obtained from the collected data we sought to confirm (confirmatory factor analyses) the adequacy of seven different combinations of latent factors (models) that have been proposed in the literature. For example, there could be a single latent factor underlying performance in the scores for all the tasks, as found in children, or there could be three latent factors, each of which representing the common variance in performance in both tasks that measured each tested EF domain, as has been found in adults. To determine which of these 7 models (measurement model) best fitted (explained) the data we inspected their fit indices. The second step was to investigate whether the psychometric properties of the best fitting model was the same, even if the participants were of different ages and/or from different socioeconomic backgrounds (i.e., there was measurement equivalence, or measurement invariance of the model across these variables). Upon ascertaining the invariance across these variables, we proceeded to the third step, namely verifying to what extent the EF latent traits of the best fitting of the seven tested models varied across age and SES. Find below a detailed description of the analyses.

Measurement model

CFA using Mplus version 8.6 (Muthén & Muthen, 2017) was used to test various competing models that have been found to explain EFs in adolescents in the literature (see below). The latent factors were initially determined by performance in two tasks for each domain, following Zanini et al. (2021). As proposed by Miyake et al. (2000) (see also Engelhardt et al., 2015; Hartung et al., 2020; Zanini et al., 2021), in all factor analyses we used absolute executive cost measures for the inhibition (Victoria Stroop and Happy-Sad Stroop) and switching tasks (Color-Shape and Category Switch tasks), but for updating, only total scores were used (Number Memory and 2-Back tasks).

We tested seven model specifications under the maximum likelihood (ML) estimator. Six of these models were chosen from among those that were most explored in the literature according to the review by Karr et al. (2018): (a) a unidimensional model (all measures loading on a single factor); (b) two-factor models (1. updating and inhibition merged with switching; 2. inhibition and updating merged with switching; 3. switching and updating merged with inhibition); (c) a three-correlated factor model; (d) the 'nested' model, which is a bifactor model that includes a common factor among all variables and the two extra factors formed by the updating and switching indicators with no covariation with each other. We did not test the classic bifactor model explored by Karr et al. (2018) that includes all three factors of EFs plus a common factor because it was not accepted in any of the publications reviewed by these authors. The seventh model was a bifactor-(S-1) model (Eid et al., 2017), a new model specification that had not been previously explored in the literature in respect of the unity and diversity EF framework. This model was equivalent to the 'nested' model except that it included the covariation between the updating and switching factors. This bifactor-(S-1) model was included because Eid et al. (2017) showed that not adding covariance between the specific factors leads to anomalies in the patterns of factors loadings (Eid et al., 2017), which may alter the interpretability of the model. Furthermore, assuming that there is no interrelation between switching and updating factors seems to be too strong an assumption, especially as these domains do correlate in the three-correlated factor model.

To assess the quality of the models, we used the following fit indices according to the recommendations of Schermelleh-Engel et al. (2003) and Schreiber et al. (2006), namely the chi-square ($\chi 2$) test ($p \geq 0.01$), root mean square error of approximation ($RMSEA \leq 0.08$), with a corresponding p-value > 0.05, standardized root mean-square residual ($SRMR \leq 0.10$), the Tucker-Lewis Index ($TLI \geq 0.95$) and the comparative fit index ($CFI \geq 0.95$). To compare fit of different models we used the chi-square difference test when models were nested, with non-significant result p value indicating no significant increase in misfit from one model to another. When a model presented poor fit, modification indices (MIs) were inspected. MIs indicate parameters that can be added to be estimated to improve model fit. MIs were applied here when their values were higher than four and they were theoretically justifiable in terms of EF (Brown, 2014).

We also tested the seven abovementioned model specifications adding a third indicator in the inhibition latent factor. This was done because the authors of the FREE test battery acknowledge a possible limitation regarding the inclusion of the Victoria Stroop task, which relies on reading skills that may not be fully developed in young samples and can be impaired in people with inadequate schooling. Controlling for reading skill when using the Victoria Stroop may not be important in adult samples from developing nations, but must be considered at ages when this ability is still improving and/or for samples that may have low quality education, such as ours. Adolescents with reading difficulties (possibly associated with worse schooling and/or low SES) would potentially find it easier to name the color of the ink in which color names were written (block 3) and therefore present *lower* inhibition costs, which would not actually reflect better executive functioning, biasing our findings. This new variable (here called non-incongruent cost) was the RCS of the Victoria Stroop task in which the stimuli were non-color word names printed in different colors (here called Block 2) minus the RCS of the control block (Block 1), in which colors of rectangles were named. By including a correlation a priori, between residuals of both these Victoria

Stroop metrics (RCS in block 2 minus block 1 and RCS in block 3 minus block 1), we hoped to correct for any individual differences in reading abilities in the usual Stroop metric that is included in these models (performance in conditions in which color names and ink color are incongruous: block 3 minus block 1).

Measurement invariance testing

Invariance of the best-fitting measurement model of EFs (i.e., three-correlated factors) across age and SES (level of parental schooling) was investigated using "multiple indicators, multiple causes" (MIMIC; Brown, 2015). Models were estimated using the maximum likelihood estimator (Muthén & Muthen, 2017). In the MIMIC analyses, both the latent factors and indicators of the CFA solution are regressed onto continuous and categorical covariates [i.e., age (in months) and SES (level of parental schooling ranging from 1 = under diploma to 5 = post-graduation)]. A significant direct effect of covariates on a latent factor is evidence of population heterogeneity (i.e., the factor means are different at different levels of the covariates) and a significant direct effect of covariates on an indicator depicts measurement non-invariance (i.e., means of indicators are different at different levels of the covariates [differential item functioning (DIF)]). In other words, MIMIC modeling tests the invariance of factor means (population heterogeneity) and indicator intercepts (measurement invariance). All direct effects of the covariate on the EF measures (i.e., indicators) are fixed to zero in an exploratory fashion to investigate if salient direct effects may be present in data by inspecting MIs. Also, the residual variances of latent factors (inhibition, updating, and switching) are specified to be correlated as they are not completely orthogonal and this overlap cannot be fully accounted for by covariates (age and SES). MIMIC modeling has the advantages of having a smaller sample size requirements than multiple-group CFA, the availability of invariance testing for continuous covariates without having to use arbitrary cutoffs values and greater parsimony (i.e., fewer freely estimated parameters). Model fit and MIs were assessed following the same recommendations described above for the measurement model. MIs in the Mplus output were inspected for any direct effect that should be freely estimated in the model, which would mean that the indicator showed DIF (indicated by MIs higher than four for an indicator on the covariate; Brown, 2014). Evidence of measurement invariance was found when the inspection of MIs revealed no DIFs, which means that response probabilities to the items do not differ across the values of the covariate (with equivalent value of the factor trait).

Results

The databank is available at https://doi.org/10.17605/OSF.IO/D4Y36. Participants' ages ranged from 9 to 15 years and parental/guardian schooling from levels 1 to 5. Demographics, raw scores and RCS per task are shown in Table 2. The following number of outliers were found per variable: one in the Victoria Stroop task (incongruous inhibition cost), three in the Happy-Sad Stroop task, two in the 2-back, four in the number memory task, three for category switch and five in the color-shape task. These values were substituted for the mean ± 3 SD.

A GLM was conducted to ensure that the inhibition and switching tasks presented executive cost. Models included the factor block [two levels: executive blocks versus control blocks (sum of RCS of blocks 1 and 2 in the case of the shifting tasks). The exception was the Victoria Stroop, for which we ran two models - block 1 versus block 2 (non-incongruous) and block 1 versus block 3 (incongruous)]. The results of the GLM confirmed the higher executive requirement of the executive blocks in all models, with large effect sizes: switching domain [Category switch: $F_{(1,406)} = 2783.50$, $p < 0.001$, $p\eta^2 = 0.87$; Color-Shape: $F_{(1,406)} = 4311.16$, $p < 0.001$, $p\eta^2 = 0.91$] and inhibition domain [Happy-Sad Stroop: $F_{(2,812)} = 1912.44$, $p < 0.001$, $p\eta^2 = 0.82$; Victoria Stroop non-incongruous: $F_{(1,406)} = 600.31$, $p < 0.001$, $p\eta^2 = 0.60$; Victoria Stroop incongruous: $F_{(1,406)} = 2439.21$, $p < 0.001$, $p\eta^2 = 0.86$]. Note that the presence of words in the Victoria Stroop task made it harder to name colors, both when the words were color names incongruous with the color they were printed in and when the words were not color names.

Measurement model

We tested the seven different CFA models specifications (see Table 3) with two indicators per factor, as proposed by Zanini et al. (2021), and also with an added indicator only in the inhibition domain: non-incongruous inhibition cost in the Victoria Stroop (block 2 minus block 1). We included a covariation a priori between residuals of the two measures of the Victoria Stroop because both are from the same task and are influenced by reading abilities. The syntax for all tested models can be found in the Supplementary File.

We chose to focus on the models with three inhibition indicators despite the fact that the non-incongruous measure of the Victoria Stroop did not significantly load onto the inhibition factor in any of the models that converged. This is because, for these solutions, model fits were better compared to models with only two inhibition indicators (for all models except for the three-factor solution) and because, without the non-incongruous Victoria Stroop

Table 2 Demographics and descriptive statistics of raw score and Rate Correct Scores (RCS: accuracy divided by the total time in seconds) in all blocks per tasks and absolute executive costs (RCS of executive blocks minus RCS of control blocks) according to each executive domain

Variables		Raw scores		RCS
	Mean (±SD)	Speed (s) [mean±(SD)]	Accuracy (no) [mean±(SD)]	Mean (±SD)
Demographics (N = 407, 170 girls)				
Age (months)	144.55 (20.42)			
Mean parental schooling (level)	3.31 (1.20)			
Inhibition tasks				
Victoria Stroop – Block 1 (control: name color patches)		15.00 (3.77)	23.97 (0.22)	1.68 (0.35)
Victoria Stroop – Block 2 (executive: name ink of non-color words)		18.84 (5.09)	23.97 (0.19)	1.36 (0.34)
Victoria Stroop – non-incongruous inhibition cost (block 2 minus block 1)		——	——	0.32 (0.26)
Victoria Stroop – Block 3 (executive: name ink of color words)		27.21 (7.94)	23.71 (0.72)	0.94 (0.26)
Victoria Stroop – incongruous inhibition cost (block 3 minus block 1)		——	——	0.74 (0.30)
Stroop Happy-Sad – Block 1 (control: name emotion)		11.52 (2.37)	19.98 (0.29)	1.37 (0.30)
Stroop Happy-Sad – Block 2 (executive: name opposite emotion)		21.53 (5.95)	19.49 (1.02)	0.97 (0.26)
Stroop Happy-Sad – Inhibition cost (block 2 minus block1)		——	——	0.40 (0.24)
Switching tasks				
Color-Shape – Block 1 (control: classify by shape)		18.24 (4.29)	19.95 (0.34)	1.14 (0.26)
Color-Shape – Block 2 (control: classify by color)		17.14 (3.74)	19.96 (0.26)	1.22 (0.27
Color-Shape – Sum of control blocks		——	——	1.18 (0.24)
Color-Shape – Block 3 (executive: switch classification)		69.04 (15.62)	39.84 (0.71)	0.60 (0.12)
Color-Shape – Switching cost (sum of block 1 and 2 minus Block 3)		——	——	0.57 (0.18)
Category Switch – Block 1 (control: classify as living/non-living)		19.47 (4.47)	19.97 (0.20)	1.08 (0.25)
Category Switch – Block 2 (control: classify as big/small)		22.92 (4.45)	19.80 (0.56)	0.91 (0.22)
Category Switch – Sum of control blocks		——	——	0.98 (0.22)
Category Switch – Block 3 (executive: switch classification)		78.82 (24.10)	39.80 (0.61)	0.55 (0.16)
Category Switch – Switching cost sum of block 1 and 2 minus Block 3)		——	——	0.43 (0.17)
Updating tasks				
Number Memory – (total correct score) *		113.54 (36.82)	21.78 (2.27)	0.21 (0.07)
2-Back – (total correct minus false alarms) *		95.60 (29.63)	62.49 (4.39)	0.67 (0.21)

For number of trials per task, see Table 1. *Absolute costs in the updating tasks were not obtained, following the literature. Rate Correct Scores (RCS) of the inhibition and switching tasks are reversed scored so higher values represent better performance (see text)

indicator, some expected correlations between factors were not significant (i.e., between updating and inhibition in the three-factor solution). It therefore seemed that controlling for reading skills in the Victoria Stroop task was important to show the unity and diversity of EF.

Table 3 contains the fit indices for all the tested models that reached convergence. We found that the models with one and two factors did not present adequate model fit indices, even when considering theoretically acceptable MIs, except for the two-factor solution with merged inhibition and switching domains. The three-correlated factors model also showed a good fit. The CFA nested and bifactor-(*S*-1) models only converged if manipulations of the variances and indicators were included and, even then, both models presented non-positive residual covariance matrices, so were deemed inadequate (see Supplementary File for details). Figure 2 shows model configurations, standardized factor loadings, and fit indices for the two adequately fitting models (two-correlated factors with merged inhibition and switching and the three-correlated factors). See Supplementary Files for the diagrams of the remaining five tested models.

The two models with good fit (see Fig. 2) were compared using a chi-square difference test that showed no significant increase in misfit between the models ($\chi(2)^2 = 3.84$, $p = 0.15$).

The two-factor solution with merged inhibition and switching had significant correlations between factors (r = 0.75), which was also found for the three-factor solution (ranging from r = 0.38 to r = 0.78). Significant factor loadings on individual indicators for this two-factor solution were λ = 0.68 to λ = 0.75 for updating and λ = 0.20 to λ = 0.78 for the merged inhibition-switching factor. For the three-factor model, factor loadings of the inhibition domain

Table 3 Model fit for the 7 tested confirmatory factor analyses models

Model	χ2 (df)	CFI	TLI	RMSEA	90% C.I.	SRMR
One factor	48.55 (13)*	0.95	0.91	0.082	[0.058, 0.107]	0.056
Two factors (updating merged with switching)	46.70 (12)*	0.95	0.91	0.084	[0.060, 0.110]	0.054
Two factors (updating merged with inhibition)	39.02 (12)*	0.96	0.93	0.074	[0.049, 0.101]	0.055
Two factors (inhibition merged with switching)	**29.05 (12)***	**0.97**	**0.95**	**0.059**	**[0.032, 0.087]**	**0.047**
Three factors	**25.02 (10)***	**0.98**	**0.95**	**0.061**	**[0.031, 0.091]**	**0.041**
Nested	No convergence					
Bifactor-(S-1)	Residual matrix non positive					

Models in bold presented equally good fit. χ2: chi-square (p ≥ 0.01), RSMEA: Root Mean Square Error of Approximation (≤ 0.08), SRMR: Standardized Root Mean-square Residual (≤ 0.10), TLI: Tucker-Lewis index (≥ 0.95), CFI: Comparative Fit Index (≥ 0.95). *$p < 0.05$

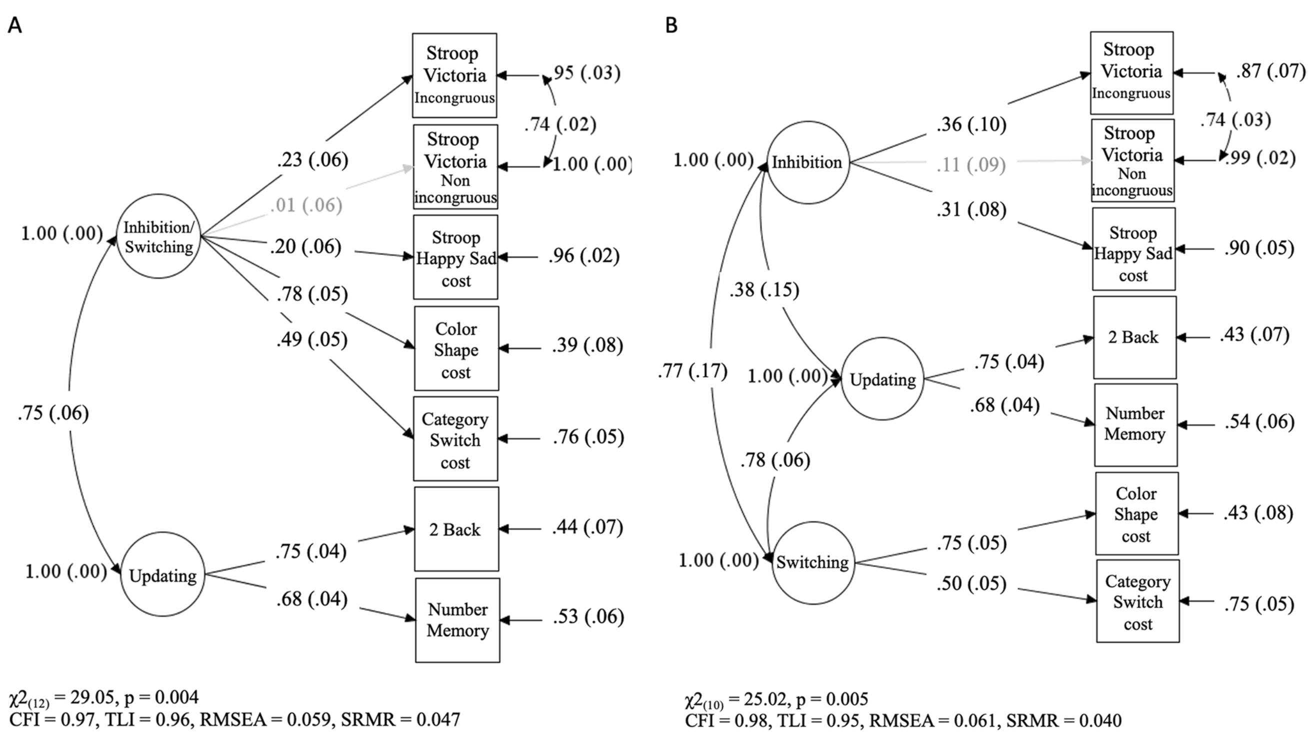

Fig. 2 Diagrams of the two best solutions among the 7 tested Confirmatory Factor Analyses models. Diagrams include factor loadings (values on the linear arrows) of executive measures (squares) on the executive latent variables (ovals). Note: Standard errors of residuals for each task are represented by the numbers at the end of the arrows pointing towards each square. Double-headed arrows represent correlations of residual variances between the latent variables. Grey arrows represent non-significant paths (p > 0.05). Fit indices of each psychometrically adequate model are presented below each diagram. Acceptable fits for each index are: χ2: chi-square test ($p \geq 0.01$), RSMEA: Root Mean Square Error of Approximation (≤ 0.08), SRMR: Standardized Root Mean-square Residual (≤ 0.10), TLI: Tucker-Lewis index (≥ 0.95), CFI: Comparative Fit Index (≥ 0.95). Rate Correct Score costs of the inhibition and switching tasks are reversed scored so that higher values for all indicators represent better performance (see text). See the Supplementary File for the syntax and diagrams of other tested models.

were low (λ = 0.31 to λ = 0.36) and moderate to high for switching (λ = 0.50 to λ = 0.75) and updating (λ = 0.68 to λ = 0.75). In both cases, the third added indicator in the inhibition latent factor was the only one to have non-significant factor loadings.

Despite having two good and equivalent model solutions, we selected the three-correlated factors model as our measurement model and used it to test measurement invariance. We did this because the latter shows that the three EF domains of the unity and diversity EF framework can already be dissociated at the tested age, which corresponds to the model solution that is found later in life, in young adults (e.g., Miyake et al., 2000). Therefore, we considered it the model of greater interest.

Measurement invariance testing

The analysis of invariance, based on the three-correlated factor model, using the MIMIC model, converged when including age (in months) and mean parental level of schooling (SES) as covariates concomitantly, but presented a non-positive covariance matrix when direct paths from the covariates to the indicators were all constrained to 0. Inspection of MIs indicated DIF for the Victoria Stroop non-incongruous inhibition cost on age (MI > 4). The model freeing this path to be estimated converged, and it had acceptable fit ($\chi^2_{(18)} = 44.06$, $p = 0.001$, CFI = 0.97, TLI = 0.94, and RMSEA = 0.06) (see Fig. 3). Inspection of MIs still showed DIF for both the Victoria Stroop cost measures for SES. As the MI value was higher for the non-incongruous inhibition cost of the Victoria Stroop task, we decided to free this path first. The model showed an acceptable fit ($\chi^2_{(17)} = 35.75$, $p = 0.005$, CFI = 0.98, TLI = 0.95 and RMSEA = 0.052). Inspection of the MIs suggested no more DIFs. At the level of the indicators, the model thus presented partial invariance, with only one non-invariant indicator.

Regarding population heterogeneity, older age predicted improvement in latent EF scores: lower costs in the inhibition (β = 0.36, $p = 0.001$) and switching (β = 0.44, $p < 0.001$), as well as better updating (β = 0.56, $p < 0.001$). Having parents who had completed higher levels of schooling also positively and significantly predicted improved switching (β = 0.16, $p = 0.006$) and updating (β = 0.16, $p = 0.003$), but the effects were minimal and did not affect the inhibition latent trait (β = 0.026, $p = 0.80$). Based on Cohen's d, all the effects of the covariates on the EF factors are small, except for the effect of age on updating, which was a medium effect (Cohen's d = 0.56). In this model, however, the correlation between the residual variances of inhibition and updating was no longer significant (although this type of change is not usually addressed in the literature, we have included a brief discussion about this in the Supplementary file).

Discussion

In this study, we tested which among a number of factor structures best reflected executive functioning in the first half of adolescence, and also the effects of age and SES using a culturally adaptable test battery based on the EF unity and diversity framework of Miyake et al. (2000) that proposes the existence of three separable EF (inhibition, updating and shifting). We had four main findings. First, the three-correlated factor structure had adequate psychometric properties and best explained our data considering that it conformed with most prior studies. This is important because we were able to show that from the first half of adolescence it is already possible to differentiate

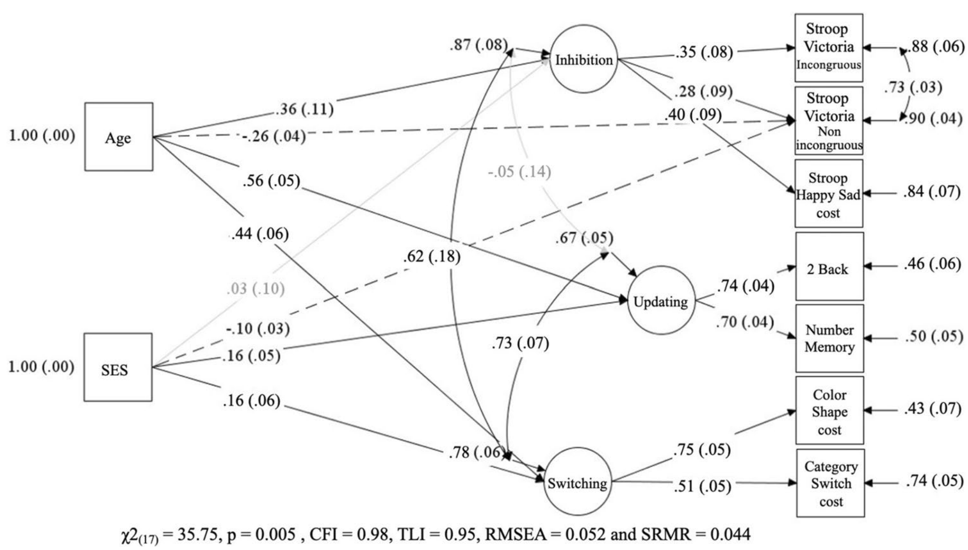

Fig. 3 Multiple indicators, multiple causes (MIMIC) model showing the effect of the continuous covariates (age in months and mean level of guardian schooling as a proxy for socioeconomic status - SES) on the three latent factors (ovals) of the three-correlated factor solution. The diagram also displays factor loadings (values on the linear arrows) of executive tasks (squares) on the executive latent variables. Note: Standard errors of residuals for each task are represented by the numbers at the end of the arrows pointing to each square. Double-headed arrows represent correlations between residuals of variances of the latent variables. Grey arrows represent non-significant estimates ($p > 0.05$). Dashed arrows represent the significant direct paths (effects) of the covariates on the indicator with differential item functioning (non-invariant indicator: only for the non-incongruous inhibition cost of the Victoria Stroop test). Fit indices found for the model is at the bottom of the diagram (for symbols, see Fig. 2). Rate Correct Score costs of the inhibition and switching tasks are reversed scored so higher values for all indicators represent better performance (see text)

(diversity) the three tested EF latent domains (inhibition, updating and switching) that are also interrelated (unity), as it was unclear from which phase in adolescence this fractionation emerged. Second, these findings are in agreement with results obtained in samples that included adolescents with different age ranges and were from other countries/cultural contexts (China: Duan et al., 2010; Xu et al., 2013; Hong Kong: Wu et al., 2011; Brazil: Zanini et al., 2021), which speaks to the cross-cultural generality of the development of these abilities in adolescence. Third, we provided evidence of invariance of the tested tasks in terms of age and SES, which allowed us to ascertain that the constructs being measured (three latent executive domains) were the same irrespective of the participants' age or their parents' schooling, which is a novel finding in the literature. Lastly, this evidence of invariance allowed us to show that the three EF domains are differentiable right from the onset of adolescence, although they continue to improve with age, and also that being from a more privileged family had only a minor impact on these executive latent traits when tasks are designed to minimize these effects. A detailed account of the results is given below.

Factor structure that best represents data in the first half of adolescence

We found that EF abilities in the three tested domains become psychometrically differentiable during the first half of adolescence, which was confirmed by the good model fit for the three-correlated factor solution. This corroborates data from the few other studies on adolescent participants from other countries that used tasks that represent the framework of interest and tested only these three domains, although the age ranges in these studies were different from ours and included older adolescents (Duan et al., 2010; Wu et al., 2011; Xu et al., 2013; Zanini et al., 2021), and also studies in adults (see Karr et al., 2018). Other studies with adolescent samples from the United States that used comparable tasks, but also included other EF domains (e.g., working memory), also found these three domains to be interrelated and separable (Engelhardt et al., 2015; Hartung et al., 2020). Two-factor domain combinations (inhibition combined with updating and updating with switching), the unidimensional model, and the nested solution did not present adequate adherence to data, nor did the bifactor-(S-1) configuration. The latter was tested here for the first time in the literature, which corresponds to the nested factor with an added correlation between specific factors in order to circumvent psychometric distortions of the nested model (Eid et al., 2017).

However, we also found a good fit for the two-factor model in which inhibition and switching were merged, which indicates that these two factors are highly correlated in the tested age range ($r = 0.75$), possibly because switching develops later than the other domains (Karr et al., 2018). Whether this particular two-factor model fits adult data well is unclear because most studies with older populations do not describe the adequacy of various alternative models. In order to provide evidence on this, a comprehensive study by Karr et al. (2018) undertook a bootstrap reanalysis of published data to test many possible alternative factor structures and their adequacy at different ages. The authors reported that for adults both the three-factor and 'nested' factor models have better fits (with the proviso that the latter has psychometric inadequacies: Eid et al., 2017), whilst in children/adolescents, models with less factors tend to be more acceptable. However, Karr et al. (2018) analyzed data of children together with adolescents. Hence, the results of their bootstrap reanalysis are not comparable with ours because at some stage within this age-range individuals transition to displaying adult-like differentiable EF domains. The data from our study and from other studies (Duan et al., 2010; Engelhardt et al., 2015; Hartung et al., 2020; Xu et al., 2013; Zanini et al., 2021) suggest that this takes place in the first half of adolescence. Another issue that was not taken into account by Karr et al. (2018) is whether tasks in the reanalyzed studies were representative of the domains proposed in the unity and diversity model of interest here (see Morra et al., 2018), so their findings in respect of the most reliable factor structures may not be generalizable when it comes to testing this framework as it was originally conceptualized (see Supplementary File for alternative models with tasks that represent other EF domains in the age range tested here).

Invariance testing and effects of age

The three-correlated factor model was invariant across age in 6 of 7 scores/measures. Non-invariance was found only for the non-incongruous inhibition cost of the Victoria Stroop task score. This indicator was expected to vary as it depends on reading skills (Rasinski et al., 2009), which improve as adolescents become older (particularly in populations with low quality schooling such as part of our sample: Buckingham et al., 2013). This was corroborated by the negative direct effect of age on the non-incongruent measure. Hence, it can be said that the approximate invariant three-correlated model found here demonstrates that the improvement with age seen in all explored domains of EFs can be attributed to increases in EF latent traits and not to other cognitive abilities that are recruited to carry out the tasks (residuals). This improvement was highest in updating, followed by switching and lowest in inhibition. We stress, however, that the effects sizes of improvement were small or barely medium at the latent level, confirming results from other studies that also showed invariance across age in samples of similar age using scores similar to ours (efficiency scores, or accuracy divided by reaction time to correct trials), but that only tested invariance regarding a general EF latent trait based on tasks that mainly measured other EF domains than those assessed here (Xu et al., 2020).

It is possible that larger age-related effects would have been found had we analyzed raw scores at the level of individual tasks instead of latent scores. This is because raw scores reflect the use of EF together with task-specific lower-order non-executive cognitive abilities that are recruited when people carry out EF tasks, all of which continue to mature across adolescence. This can be supported by the higher improvement with age observed for the updating factor. This may have occurred because the scores used for tasks that tap updating do not include a control for speed of processing, which likely improves as adolescents age, while this is controlled for in the case of the shifting and inhibition measures (executive costs relative to control tasks) (Zanini et al., 2021). These results showing improvement of EFs with age add to the literature as this effect is not consistently found in adolescent samples from developing countries (see Schirmbeck et al., 2020), especially as prior studies that tested the unity and diversity framework with adequate tasks usually overlooked measurement invariance, which is mandatory to allow comparisons of EF performance between ages (Meredith, 1993).

Indeed, to our knowledge, no study to date has explored invariance across age as a continuous variable of the three-correlated factor solution in adolescents, or alternative configurations with tasks that represent the EF unity and diversity framework (see Supplementary File for studies on invariance in adolescents considering other EF domains). Some studies that did explore invariance across age found partial invariance (Engelhardt et al., 2015; Latzman & Markon, 2010; McAuley & White, 2011), but they are not directly comparable to ours, either because they used tasks that measure other types of EF that are not inhibition, switching and updating, or because their factor structures were different from ours. Additionally, these studies considered age as a categorial factor, with arbitrary cutoffs that vary from study to study, which is of little use to track developmental changes in EF, and is not recommended because linear relations can fail to be detected when doing so (MacCallum et al., 2002). Full invariance across age as a continuous variable was only found in one study, but in a unidimensional model (a model with a single general EF latent factor) (Xu et al., 2020), which does not provide information on the development of the diversity of EF.

Invariance testing and effects of socioeconomic status

Apart from the age-related developmental improvement in EF, another covariate that impacts these abilities is SES (Montroy et al., 2019), which varies widely in non-developed countries such as Iran, where the study took place. Ideally, these effects should be assessed using tasks that are designed to reduce possible performance advantages of having a better SES or having had better schooling but that do not relate to EF themselves (e.g., having a larger vocabulary, which can help improve performance in EF tasks). This was controlled for in our choice of test battery (FREE). In this respect, we found invariance across SES for all measures except for the same one that was not invariant to age that depends on reading skills (non-incongruous inhibition cost score of the Victoria Stroop task), which was expected to be lower in participants with parents who have lower schooling and/or SES (Thomson, 2018).

Testing for invariance across SES in models that show unity and diversity of the tested EF domains had not been undertaken before, so this was a novel result (see Supplementary File for studies that investigated invariance of models that explored other EF domains). Based on our data, it can be concluded that SES has a slight positive effect on the EF latent traits of updating and switching; however, it was about three times smaller than the effect of age. This corroborates findings in the literature showing that the effects of SES are mostly small on EFs, even when raw scores are considered instead of latent scores (Lawson et al., 2018). Differently, the inhibition factor was not affected by SES once reading skills were controlled for in the model, corroborating some findings in the literature in respect of raw scores on inhibition measures which showed no effect of SES on the Stroop Color Word test in a sample of 6 to 25-year-olds (Last et al., 2018). It should be noted, however, that only our study ascertained invariance of the model for SES (that is, ensured the EF construct being measured did not change according to SES) to explore these effects on the three separable *latent* EF domains. This suggests that SES effects can be minimal if tasks are built to minimize the contribution of other lower order cognitive abilities that contribute to performance in EF tasks but are not EF themselves. Exactly how SES can elicit these small effects is unclear, but this may stem from its negative impact on brain development/physiology (e.g., Foulkes & Blakemore, 2018) associated with systems involved in EF.

Limitations

The current study has some limitations that should be noted: (1) - because EFs develop during adolescence, a phase of life marked by accelerated brain changes in areas that support self-regulation (Foulkes & Blakemore, 2018; Goddings et al., 2019), a longitudinal study might have better captured developmental changes in these abilities. However, EF tasks have low test-retest reliability, which would be an advantage (Hedge et al., 2018) of conducting cross-sectional studies such as ours. (2) - because varying SES is a characteristic of samples from non-developed countries, a composite measure of SES that included variables other than parental schooling, such as family purchasing power, income and parental occupation, would have been ideal. The measure of parental schooling, however, is extensively used (Foulkes & Blakemore, 2018; Last et al., 2018; Montroy et al., 2019) and has been shown to correlate with EF abilities (e.g., Last et al., 2018; Montroy et al., 2019). Therefore, we assumed that our measure was a fair representation of what we could find with a composite measure of SES. (3) - the use of only two tasks per domain in our study, which is common practice in underaged samples (see Karr et al., 2018), could be regarded as a limitation, because more indicators are often needed to reach acceptable model fit (Karr et al., 2018). However, with only two tasks per domain we were able to differentiate all types of EF in a good fitting model, suggesting that the use of appropriate tasks, rather than the number of tasks, might be more important to show the diversity of EF. Nonetheless, including more indicators per domain is advisable and might have resolved the convergence problems of the nested and bifactor-(S-1) models. Additionally, this could have allowed us to circumvent the limitation of using the Victoria Stroop task that relies on reading skills, which can be criticized when testing disadvantaged populations (see Zanini et al., 2021) and was indeed found here to be non-invariant to age and SES. (4) - the FREE does not include tasks that control for psychomotor speed in the updating measures as it does for the other EF domains. Although this is in line with the great majority of studies in the literature (see Karr et al., 2018), it could well be that accounting for this might have led to different model configurations. (5) there was no validation procedure for the translation of the tasks despite extensive testing in pilot studies. The battery, however, is almost language free, as the only task that includes words (four color names) is the Victoria Stroop task. Additionally, this does not seem to have negatively impacted the results because we were able to psychometrically replicate a previous study using the same battery of tasks in Portuguese-speaking adolescents (Zanini et al., 2021) and also the seminal model that the present study was based on, which involved American adults (Miyake et al., 2000). (6) - we did not control for intelligence (e.g., unlike Wu et al., 2011), because many intelligence measures correlate differently with inhibition, shifting and updating (see Friedman et al., 2006), so this control would probably have distorted the results. (7) - measurement invariance in the tested framework for pubertal status and sex were not tested as there is no data in the former case, or consistent indication, in the latter (e.g., Xu et al., 2013) that these variables affect EF. (8) - we only explored the unity and diversity framework as reflexive and not as formative models, which Willoughby et al. (2014) suggest may be a better alternative.

Conclusions and implications

In conclusion, we showed that inhibition, shifting and updating can already be differentiated at the onset of adolescence when EF tasks are based on the well-established EF unity and diversity theoretical framework, corroborating studies in samples of adolescents of other age ranges and adults from various different countries. We also showed that this fractionation of EFs was mostly invariant across age and parental schooling (SES), that performance at the latent level improved with age and was only very slightly positively associated with SES (except for having no effect on inhibition). Importantly, these results were found in a relatively diverse sample compared with those investigated in most cognitive studies (Henrich et al., 2010) as the participants were Persian-speaking adolescents from a widely different socio-cultural context that includes high variability in terms of SES and educational levels. The results of the present study, therefore, not only provide information on an under-represented population in the literature, but also speak to the cross-cultural generality of the development of EFs. Future cross-country studies that aim to confirm the universality of these findings may therefore consider using the FREE test battery as it exhibited adequate psychometric properties and its tasks were easily adaptable to the Iranian setting. Future studies would also benefit from the inclusion of more tasks to explore the development of other types of EF in the first half of adolescence (e.g., planning, dual tasking, access to long term memory, and reasoning).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12144-022-03974-3.

Funding Open Access funding provided by Ostfold University College This study was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Processes: due to fellowships to author IS: 2019/19709-6 and SP: 2016/14750-0 and SP/AN: 2017/02816-9), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (finance code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – #301,899/2019-3 and 302,511/2018-0 due to fellowships to author SP and MCM), Associação Fundo de Incentivo à Pesquisa (AFIP) and Iranian Cognitive Science and Technologies Council, Tehran, Iran (Project code: 4758), all non-profit organizations that sponsor research.

Data availability The datasets generated and/or analyzed during the current study are available in the Open Science Framework repository at https://doi.org/10.17605/OSF.IO/D4Y36.

Declarations

All procedures of this cross-sectional study were approved by the institutional ethics committee of the Education Office of Tehran (Approval code: D/100/10,247; DATE: 2018-11-01). Participation informed consent was provided by legal guardians.

Competing interests The authors declare no commercial or financial relationships that could be construed as a potential conflict of interest.

References

Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). Methodology in the Social Sciences, (pp. 462). The Guilford Press.

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education, 57*(3), 190–213. https://doi.org/10.1177/0004944113495500

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Duan, X., Wei, S., Wang, G., & Shi, J. (2010). The relationship between executive functions and intelligence on 11- to 12-year-old children. *Psychological Test and Assessment Modeling, 52*(4), 419–431.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083

Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P., & Tucker-Drob, E. M. (2015). Genes unite executive functions in childhood. *Psychological Science, 26*(8), 1151–1163. https://doi.org/10.1177/0956797615577209

Foulkes, L., & Blakemore, S. J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience, 21*(3), 315–323. https://doi.org/10.1038/s41593-018-0078-4

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology, 52*(2), 326–340. https://doi.org/10.1037/dev0000075

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Goddings, A. L., Beltz, A., Peper, J. S., Crone, E. A., & Braams, B. R. (2019). Understanding the role of puberty in structural and functional development of the adolescent brain. *Journal of Research on Adolescence, 29*(1), 32–53. https://doi.org/10.1111/jora.12408

Haft, S. L., & Hoeft, F. (2017). Poverty's impact on children's executive functions: Global considerations: Poverty's impact on children's executive functions: Global considerations. *New Directions for Child and Adolescent Development, 2017*(158), 69–79. https://doi.org/10.1002/cad.20220

Hartung, J., Engelhardt, L. E., Thibodeaux, M. L., Harden, K. P., & Tucker-Drob, E. M. (2020). Developmental transformations in the structure of executive functions. *Journal of Experimental Child Psychology, 189*, 104681. https://doi.org/10.1016/j.jecp.2019.104681

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Kramer, H. J., Lagattuta, K. H., & Sayfan, L. (2015). Why is happy–sad more difficult? Focal emotional information impairs inhibitory control in children and adults. *Emotion, 15*(1), 61–72. https://doi.org/10.1037/emo0000023

Lagattuta, K. H., Sayfan, L., & Monsour, M. (2011). A new measure for assessing executive function across a wide age range: Children and adults find happy-sad more difficult than day-night: Children and adults find happy-sad more difficult. *Developmental Science, 14*(3), 481–489. https://doi.org/10.1111/j.1467-7687.2010.00994.x

Last, B. S., Lawson, G. M., Breiner, K., Steinberg, L., & Farah, M. J. (2018). Childhood socioeconomic status and executive function in childhood and beyond. *PLoS One1, 13*(8), e0202964. https://doi.org/10.1371/journal.pone.0202964

Latzman, R. D., & Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment, 17*(2), 172–184. https://doi.org/10.1177/1073191109356254

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science, 21*(2), e12529. https://doi.org/10.1111/desc.12529

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development, 84*(6), 1933–1953. https://doi.org/10.1111/cdev.12096

Lurie, L. A., Hagen, M. P., McLaughlin, K. A., Sheridan, M. A., Meltzoff, A. N., & Rosen, M. L. (2021). Mechanisms linking socioeconomic status and academic achievement in early childhood: Cognitive stimulation and language. *Cognitive Development, 58*, 101045. https://doi.org/10.1016/j.cogdev.2021.101045

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

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*(3), 453–468. https://doi.org/10.1016/j.jecp.2010.08.009

McNeilly, E. A., Peverill, M., Jung, J., & McLaughlin, K. A. (2021). Executive function as a mechanism linking socioeconomic status to internalizing and externalizing psychopathology in children and adolescents. *Journal of Adolescence, 89*, 149–160. https://doi.org/10.1016/j.adolescence.2021.04.010

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/BF02294825

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*(2–3), 123–142. https://doi.org/10.1016/j.actpsy.2003.12.004

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*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Montroy, J. J., Merz, E. C., Williams, J. M., Landry, S. H., Johnson, U. Y., Zucker, T. A., Assel, M., Taylor, H. B., Lonigan, C. J., Phillips, B. M., Clancy-Menchetti, J., Barnes, M. A., Eisenberg, N., Spinrad, T., Valiente, C., de Villiers, J., & de Villiers, P. (2019). Hot and cool dimensionality of executive function: Model invariance across age and maternal education in preschool children. *Early Childhood Research Quarterly, 49*, 188–201. https://doi.org/10.1016/j.ecresq.2019.06.011

Morra, S., Panesi, S., Traverso, L., & Usai, M. C. (2018). Which tasks measure what? Reflections on executive function development and a commentary on Podjarny, Kamawar, and Andrews (2017). *Journal of Experimental Child Psychology*, *167*, 246–258. https://doi.org/10.1016/j.jecp.2017.11.004

Muthén, L. K., & Muthen, B. (2017). *Mplus User's Guide* (8th ed.). (944) Muthén & Muthén

Nouri, A., Esmaeili, F., Seyedi, H., Rezaeian, S., Panjeh, S., Cogo-Moreira, H., & Pompeia, S. (2021). Factor structure and psychometric properties of the Persian versions of the Pediatric Daytime Sleepiness Scale and Morningness–Eveningness Scale for Children. *Journal of Education and Health Promotion, 10*(1), 343–343. https://doi.org/10.4103/jehp.jehp_260_21

Programme, U. N. D. (2020). *Human development report 2020*. United Nations. https://www.un-ilibrary.org/content/books/9789210055161. Accessed 18 Feb 2022.

Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48*(4), 350–361. https://doi.org/10.1080/19388070802468715

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

Schirmbeck, K., Rao, N., & Maehler, C. (2020). Similarities and differences across countries in the development of executive functions in children: A systematic review. *Infant and Child Development*, *29*(1). https://doi.org/10.1002/icd.2164.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Strauss, E., Sherman, E. M. S., Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. (pp. 1216). Oxford University Press.

Thomson, S. (2018). Achievement at school and socioeconomic background—An educational perspective. *NPJ Science of Learning*, *3*(1). https://doi.org/10.1038/s41539-018-0022-0.

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*(5), 427–449. https://doi.org/10.1016/j.intell.2006.09.001

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research and Perspectives, 12*(3), 69–95. https://doi.org/10.1080/15366367.2014.929453

Wu, K. K., Chan, S. K., Leung, P. W. L., Liu, W. S., Leung, F. L. T., & Ng, R. (2011). Components and developmental differences of executive functioning for school-aged children. *Developmental Neuropsychology, 36*(3), 319–337. https://doi.org/10.1080/87565641.2010.549979

Xu, C., Ellefson, M. R., Ng, F. F. Y., Wang, Q., & Hughes, C. (2020). An East–West contrast in executive function: Measurement invariance of computerized tasks in school-aged children and adolescents. *Journal of Experimental Child Psychology, 199*, 104929. https://doi.org/10.1016/j.jecp.2020.104929

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS One, 8*(10), e77770. https://doi.org/10.1371/journal.pone.0077770

Zanini, G. A. V., Miranda, M. C., Cogo-Moreira, H., Nouri, A., Fernández, A. L., & Pompéia, S. (2021). An adaptable, open-access test battery to study the fractionation of executive-functions in diverse populations. *Frontiers in Psychology, 12*, 627219. https://doi.org/10.3389/fpsyg.2021.627219