The unity and diversity of executive functions across adulthood in a diverse Brazilian sample with varying educational attainment

Sabine Pompeia[1], *Tamara Pessanha Taporoski*[2,3]*, *Isis Angelica Segura*[1], *Luz Marina Gómez*[3], *Hugo Cogo-Moreira*[4], *Andrea Roseli Vançan Russo Horimoto*[5], *Alexandre da Costa Pereira*[3], *Homero Vallada*[3], *Andre Brooking Negrao*[3]

[1]Departamento de Psicobiologia, Universidade Federal de São Paulo, Brazil; [2]Harvard Center for Population and Development Studies Harvard University Cambridge Massachusetts, United States of America; [3]Insituto de Psiquiatria do Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo, Brazil; [4]Department of Education, ICT and Learning, Østfold University College, Halden, Norway; [5]Universidad Nacional de Colombia, Manizales, Colombia; [6]Division of Aging, Brigham and Women's Hospital, Boston, United States of America.

Sabine Pompeia http://orcid.org/0000-0002-5208-8003

Tamara Pessanha Taporoski https://orcid.org/0000-0001-8783-4013

Isis Angelica Segura *https://orcid.org/0000-0003-0162-2066*

Luz Marina Gómez *https://orcid.org/0000-0001-8641-4009*

Hugo Cogo-Moreira *https://orcid.org/0000-0001-9411-9237*

Andrea R.V.R. Horimoto *https://orcid.org/0000-0002-8573-5158*

Alexandre da Costa Pereira *https://orcid.org/0000-0002-7782-5540*

Homero Vallada *https://orcid.org/0000-0001-5123-8295*

Andre Brooking Negrao No ORCID

*Corresponding author:

Tamara P. Taporoski

Harvard Center for Population and Development Studies

Harvard University, Cambridge Massachusetts USA

ZIP: 02138

E-mail: ttaporoski@harvard.edu

Abstract

Executive functions (EFs) are a set of cognitive abilities associated with many social and clinical outcomes which are believed to be sensitive to culture, age and educational attainment. Yet, theoretical models of EF and tasks used to measure them were developed mainly for highly educated young adults from developed countries. Here we assessed EFs in a diverse cohort from a rural town in Brazil (Baependi Heart Study: N=1253; 753 females), in which age (18-88 year) and schooling (0-23 years) varied widely. Based on a popular theoretical model of EF (the unity diversity EF framework), which describes the pattern of inter-associations of three specific types of EFs (inhibitions, shifting and updating), we developed a short, nonautomated, open-access EF test battery adapted for use in samples with varying schooling and age that included EF indicators obtained from three tasks (Plus-Minus, Trail Making and Random Number Generation tasks). Using Confirmatory Factor Analyses we found a good fitting three-correlated factor solution confirming the fractionation and inter-association of the three EF domains that was invariant (except for one of eight indicators) to age, schooling and sex. More years of schooling was related to higher latent traits in all domains, older age to worse shifting and women scored lower in latent shifting and inhibition. We conclude that the unity-diversity EF framework can be determined across adulthood in diverse samples with tasks adapted for these populations, allowing the theory-oriented investigation of sociocultural and demographic factors that influence EFs.

Key words: executive functions, educational status, age, sex, factor analysis.

INTRODUCTION

Executive functions (EFs) are a set of cognitive capacities responsible for forming, planning, maintaining and implementing goals which are temporarily held in working memory (Friedman & Miyake,2017). Because these capacities underly essential human abilities such as abstraction, reasoning, making decisions, controlling thought and action (Friedman & Miyake, 2017), it stands to reason that there should be aspects of executive functioning that are universal to humankind and that are similarly altered as people grow older, irrespective of their backgrounds (Barrett, 2020; Liebal & Haun, 2018).

Another important aspect regarding EFs' central role in behavior is that they are associated with many outcomes such as academic and professional success, interpersonal and social interactions and physical and mental health (see Baggeta & Alexander, 2016; Dias et al., 2024). However, this was determined mainly in samples from high income countries. This is a limitation of the literature because, across development, many aspects of behavior are influenced by sociocultural factors which involve different languages, types of mental demands undertaken in people's daily lives, their social values and norms (see Barrett, 2020). This also applies to EFs (Schirmbeck et al., 2020). It is therefore paramount to determine whether EF performance is similar in diverse samples so that a more inclusive and comprehensive portrait of human EF abilities can be established. Although most of the literature recognizes that EFs are multidimensional (i.e. include many different abilities), determining common aspects of EF across cultures has been hampered not only by a lack of diversity in the samples used in EF studies, but also by the differences in theoretical perspectives that were adopted across publications and the ambivalence in how the tested EF measures map onto specific EF constructs (Dias et al., 2024). Hence, for a better understanding of EFs it is necessary that studies be based on sound theoretical models and be assessed in different populations. This requires the adaption of tasks to ensure that the population under study can successfully understand and carry them out, which are dependant on their native language and culture and the extent to which test-takers were cognitive stimulated throughout life.

There are many theoretical models of EF that are based on abilities that do not relate to emotional and/or social stimuli (termed cool executive functions). Probably the most influential one (Baggeta & Alexander, 2016; Dias et al., 2024) is the unity and diversity framework of EFs (Friedman & Miyake, 2017; Miyake et al., 2000), which we based the present study on. This framework proposes that EF abilities are correlated (unity of EF), yet separable (diversity or fractionation) at the level of latent variables, that is, variables that represent the shared variance in scores (indicators) of more than one task that measures similar executive processes. The unity and separability of EF was initially explored considering three (among the many types) of existing executive abilities (Miyake et al., 2000): a) inhibition of automatic or prepotent responses (henceforth called inhibition), that is, the ability to suppress habitual behavior; b) switching or shifting, the ability to alternate between two or more goals; and c) updating, the capacity to update the contents held in working memory in order to temporarily maintain only information that is relevant to reach a given goal. Miyake et al. (2000) found these three latent EF domains to be separable but interrelated in a three-correlated latent factor configuration/solution.

Since this initial proposal, the unity and fractionation of executive domains into three-correlated factors was replicated in samples with different demographic characteristics, but many studies also found alternative model configurations, such as only two-factors, in which two of the three domains were merged (Karr et al., 2018).

The most common selected configural solution in adults takes in a bifactor configuration (also called "nested"), in which the variance of the latent variable obtained from performance in the inhibition tasks is completely explained by a factor that represented the common variance in scores obtained from all tasks that measured inhibition, shifting and updating (Friedman et al., 2016; Friedman & Miyake, 2017). Both the three-correlated and bifactor models have also been found to have distinguishable physiological underpinnings (Feng et al., 2022; Friedman & Miyake, 2017; Friedman & Robbins, 2021; Ma et al., 2023; Rodríguez-Nieto et al., 2022), that is, all EF domains involve similar processes (unity) although each one shows specific effects as well (diversity).

The abovementioned studies on the EF unity and diversity framework in adulthood involved mostly samples from high income countries that either included only young individuals or older ones but with restricted age ranges (see Karr et al., 2018; Wray et al., 2020) and used tasks that were developed for these populations (see Dias et al., 2024). This limits the generalization of this framework to more diverse samples from low- and middle-income countries, where living standards and life opportunities tend to be much less homogenous. This is a considerable hindrance when studying EFs because there is a strong body of evidence that low socioeconomic status affects cognitive development (e.g. Amin et al., 2022; Ardila et al., 2010; Franzen et al., 2020; Lavrencic et al., 2018; Lawson et al., 2014; Wray et al., 2020) since before birth, persist at later ages and is intertwined with low cognitive stimulation (often indexed by limited educational attainment) in predicting outcomes such as physical and mental health (Hackman et al., 2015; Farah, 2017). Furthermore, even in developed nations in which there are small disparities among people in terms of years of education, Amin et al. (2022) found improvement on cognitive abilities caused by increasing schooling from secondary to tertiary levels.

EFs also change with age over adulthood. A meta-analysis showed that compared to 18-35-year-olds, people over the age of 65 years have worse performance in all three EF domains, although results were not uniform in terms of variability across tasks and domains, updating have been the least affected. Moreover, some studies propose that, in later adulthood, there is a tendency of de-differentiation among EFs domains, that is, they tend to become less separable as people age, although the exact time of life when this occurs is still unknown (e.g. Glisky et al., 2021). Sex effects on EFs are much less clear and no consistent pattern has been found (Gaillard et al., 2021; Grissom & Reyes, 2019).

This overview of the literature shows that data is still lacking to confirm the applicability of the EF unity diversity framework in adults from low- or middle-income countries who have different academic opportunities and to establish to what extent schooling, age and sex influence the pattern of inter-associations among inhibition, shifting and updating. It is therefore timely to investigate this theoretical framework in diverse samples spanning all of adulthood. To this end we assessed EFs of an adult cohort from a small rural town in Brazil including individuals of varying age and educational attainment (from illiterate to post-graduate degrees).

To reach our objective, the first step was to select EF tasks that are representative of the three EF domains proposed by the proponents of the framework mentioned above (Friedman & Miyake, 2017; Miyake et al., 2000) and that could be carried out by people of vastly different levels of schooling and ages. This was a challenging endeavor as there are very few cognitive tasks that are adequate for low-educated populations from different cultural settings (Ardila et al., 2010; Dias et al., 2024; Franzen et al., 2020). The usual tasks employed in this literature mostly involve complex instructions and words or similar verbal stimuli, which are not always adequate for people who speak languages other than English and that also put those with lower schooling at a disadvantage seeing that their vocabulary is more restricted (see Ardila et al., 2010). Additionally, most studies use computerized tasks that involve key presses to make responses, so performance of test takers who are unfamiliar with computers can be negatively affected. For a comprehensive description of these issues when testing EF in diverse populations, see Zanini et al. (2021).

To select EF measures we thus prioritized tasks that have numbers as stimuli, which are recalled and manipulated more easily than other types of verbal stimuli by illiterate/barely literate populations (Ardila et al., 2010; Deloche et al., 1999). We also sought to use tasks that could be completed quickly to avoid fatigue (see Balconi et al., 2023), which could more heavily tax individuals with lower education, with the proviso that the battery had to yield at least two measures (indicators) of each EF domain so that we could calculate a latent factor for each one of them. To avoid differential familiarity with computers among participants, we used non-automatized tasks, which are not necessarily less valid for cognitive assessment than computerized ones (Kessels, 2019). Additionally, to facilitates replications in other diverse samples, the selected tasks can be administered by non-specialists with only a little training and are in the public domain, making them more affordable to researchers worldwide who do not have access to specialists in neuropsychology, nor funds to pay for tests, hardware and proprietary software.

Our EF task battery was comprised of three tasks that can be completed in around 10 minutes. The Random Number Generation task (RNG; Towse & Neil, 1998) was used to assess updating and inhibition. This task consists of naming digits 1 to 9 in a random order for a couple of minutes at a fixed pace. This can only be successful if test takers manage to keep in mind the prior numbers that were generated (continuously update the content in working memory), which is necessary to avoid (inhibit) habitual, stereotyped, non-random patterns (e.g. sequentially naming digits of increasing or decreasing magnitude, producing sequences of even or odd numbers and such like; see Oomens et al., 2015). Indeed, specific randomization indexes obtained from RNG task responses (for detail, see Methods) were found to load mainly on two of the three EF domains of interest, updating and inhibition, in the studies by Miyake et al. (2000) and Friedman and Miyake (2004). Some of the same RNG indexes also loaded on separate principal component factors in Towse and Neil's (1998), Oomens et al.'s (2015) and Chatzopoulos et al.'s (2021) studies (see also Audiffren, 2009), indicating that they measured different cognitive abilities.

To assess EF switching, we used the Plus-Minus task (Miyake et al., 2000), which involves sequentially switching between adding and subtracting three from a list of two-digit numbers, as well as the Trail Making Task part B (Buck, 2013), which requires switching between connecting letters in alphabetical order and numbers of increasing size which are displayed on a sheet of paper, a common measure of this EF domain (see Maldonado et al., 2020). With these three tasks, which yield two or more indicators for each of the three EF domains under study, it would thus be possible to obtain latent factors to determine the unity and diversity of EF. To test the effects of age, schooling and sex on the executive domains it was necessary to first ensure that these EF indicators measured the same constructs across these demographic variables (invariance testing: see Chen, 2008).

The specific aims of the present study were therefore: 1) to investigate whether performance metrics of inhibition and updating, obtained from the RNG task, and of shifting, obtained from the Plus-Minus and Trail Making B tasks, allowed the replication of a good fitting confirmatory factor analyses (CFA) model that differentiated latent inhibitions, shifting and updating or combinations of these factors in our sample across adulthood (see Karr et al., 2018); and 2) to test indicators for invariance to schooling, age and sex in the best fitting model configuration that was obtained and, if this was found, to describe schooling, age and sex effects on EF latent factors. We hypothesized that either the trifactor or bifactor (Friedman & Miyake, 2017) model, which tend to be the most replicated ones in adults (Karr et al., 2018), would be found to have adequate fit due to the: 1) care taken in selecting adequate indicators of the three EF domains of the unity diversity framework, the most seminal EF theoretical model to date (Baggetta & Alexander, 2016); and 2) fact that EFs are essential human abilities so similar patterns of inter-relations among domains should be present irrespective of people's sociocultural background (Barrett, 2020; Liebal & Haun, 2018). We predicted that EF latent performance in most of the tested domains would be lower in older (Maldonado et al., 2020) and less schooled individuals (Lavrencic et al., 2018; Wray et al., 2020), but expected no EF sex differences, which have not been consistently found (e.g. Gaillard et al., 2021; Grissom & Reyes, 2019).

METHODS

Participants: this study's data collection took place from March 2013 to March 2016, and involved 1253 (753 females) individuals, participants of the family-based Baependi Heart Study cohort (see Egan et al., 2016), who were aged 18 to 88 years, with educational attainment ranging from zero to 23 years of schooling and who underwent the EF test battery described below.

Procedure: this cross-sectional study was approved by the Ethics Committee of the Faculdade de Medicina of the Universidade de São Paulo, Brazil (#245-15) and all participants provided written informed consent. Sampling and other information on the tested population can be found in Egan et al. (2016). We undertook extensive piloting before establishing the parameters of the EF tasks/measures which will be detailed only regarding some specific aspects. Participants were assessed once (see test battery below) in a quiet environment at their own homes at the time of day that was to their convenience. They also provided information on years of schooling, age and sex. Other data were also obtained from participants and results have been reported elsewhere.

Executive functions test battery

- Random Number Generation (RNG; see Towse & Neil, 1998): this task was used to assess the executive domains of inhibition and updating. The volunteers were instructed to generate 110 digits in a random sequence containing numbers (from 1 to 9) following a tone presented at a rate of one digit every 1.5 s (rate determined in pilot studies) produced by a metronomer. The answers were vocal and taken down by the experimenter. To illustrate the concept of randomness the instructions used the analogy of picking numbers from a hat and then replacing the numbers in the hat before the next draw. Indices of randomness were calculated using the free software RGCalc (http://www.lancs.ac.uk/staff/towse/rgcpage.html). The first 10 responses were treated as practice trials and were not scored. Scores were RNG indices obtained from the remaining generated digits which were found to be related to inhibition and updating (Friedman & Miyake, 2004; Miyake et al., 2000).

The inhibition indexes were: 1) the Turning Point Index (TPI) measures the regularity of the sequence of generated numbers taking into account ascending and descending series. For example, in the sequence “1, 5, 7, 4, 2” the digit “7” represents a TPI, since from there the sequence starts to decrease; 2) the Runs index, which is very similar to the TPI and indicates the variability between the size of ascending and descending sequences; and 3) Adjacency, which corresponds to the percentage of occurrence of obvious pairs in the sequence (for example 1 followed by 2, 5 followed by 6, etc.).

The updating indexes were: 1) Redundancy, which reflects the frequency (in %) of use of each digit throughout the test (values close to zero indicate lower redundancy); 2) the Coupon indexes, which measures the number of digits generated before all possibilities (1 to 9) are used (a score of 9 would indicate that all numbers were mentioned for every 9 digits generated); and 3) the Mean Repetition Gap, which reflects the average number of digits generated until a given digit is repeated.

For Redundancy, Coupon, Adjacency and Runs indexes, the lower the score, the better the performance; for Mean Repetition Gap, the higher the score, the better the performance (Audiffren, 2009; Maes et al., 2011), while TPI scores are better when near 100 (Towse & Neil, 1998). Errors, such as saying 10 or zero, and failures to generate numbers at every tone were ignored in the sequences.

- Plus-Minus shifting task (modified from Miyake et al., 2000): this is a paper-and-pencil measure of executive shifting. Participants were asked to alternate between adding and subtracting three to a list of two-digit number and writing down the answers while keeping in mind what was the last operation to be carried out (there were no cues regarding this). This was done on an answer sheet presenting 30 two-digit numbers that were randomly picked from 10–99, except that we excluded numbers which ended in 8 and 9 when three was to be added (e.g. 19+3=21), and numbers ending in one and two when three was to be subtracted (e.g. 32-3=29), which were found to be more difficult for little schooled individuals in pilot studies. They were asked to begin by adding three to the first two-digit number and work on the task as quickly as possible, avoiding mistakes. We assessed the time to complete the task, errors in the mathematical operations and shifting errors (the number of times two sequential calculations were additions or subtractions). As very few calculation errors were made, the score was a rate correct score (which corrects for speed-accuracy tradeoffs: Vandierendonck, 2017): number of correct shifts, from a possible maximum of 30, divided by the time taken to complete the task.

- Trail Making Test (TMT), part B (see Bucks, 2013): this paper-and-pencil test involves a sheet of paper on which 25 scattered circles in which a single digits (from 1 to 12) or consecutive letter (from A to M) is printed. The task consists of using a pen to connect the circles, beginning with the first number, shifting to the first letter, then shifting back to the second number, and so forth (i.e. 1-A, 2-B, 3-C, etc.), as quickly as possible, avoiding mistakes and not lifting the pen. So that scores were comparable to those used in the Plus-Minus task, the Trail Making rate correct scores were also used: the number of circles connected in the right order divided by the total time taken to complete the task except that, following the standard administration procedure, participants were excluded if they took more than 300 s to complete the task.

Statistical analysis

Determining EF model configuration

Data were inspected for outliers (values above or below three standard deviations of the mean), which were excluded from the analysis. The inhibition indicators were TPI, Runs and Adjacency RNG indexes, the updating ones were Redundancy, Coupon and Mean Repetition Gap index scores of the RNG task and the shifting indicators were the Plus-Minus and Trail Making rate correct scores. We tested various configural factor structures that are common in the literature (Karr et al., 2018): single factor model, two-factor models combining domains two-by-two, a three-correlated factor and the classic bifactor model [with two specific latent factors for shifting and updating plus a reference factor (composed of all indicators including those that assess inhibition)] (see Friedman & Miyake, 2017). These models were obtained using CFA with the Mplus version 8.6 software (Muthén & Muthén, 1998-2017) estimated using the Robust Maximum Likelihood (MLR) because it is robust for normality violations and is acknowledged as a sophisticated method for dealing with missing data (Brown, 2015), which were not imputed here. We also tested an additional model configuration [bifactor-(S-1) model] that has been seldom used in the EF literature (for an exception, see Segura et al., 2023) because it has recently come to light that the bifactor model used by Friedman and Miyake (2017) has psychometric limitations (Eid et al., 2017): it lacks a correlated path between the non-reference factors (shifting and updating), the reference factor.

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): chi-square ($\chi 2$) test ($p \geq 0.01$), Root Mean Square Error of Approximation (RMSEA, which should be $\leq .08$), Standardized Root Mean-square Residual (SRMR $\leq .10$), Tucker-Lewis index (TLI $\geq .95$) and Comparative Fit Index (CFI $\geq .95$). If an acceptable model fit was not found, Modification Indices (MI) were inspected, which indicate parameters that, if included, can improve model fit. MI values higher than 10 were considered when including the parameter was theoretical plausible. To compare fits of different acceptably nested models, because of the use of MLR estimator, we computed the Satorra-Bentler scaled chi-square difference test (Satorra & Bentler, 2010).

Invariance across age, schooling and sex

Invariance to schooling, age and sex was tested in the best fitting model configuration using Multiple Indicators, Multiple Causes (MIMIC; Brown, 2015) models with MLR estimator (Muthén & Muthén, 1998-2017). In MIMIC analyses, factors of the CFA solution are regressed onto continuous covariates [i.e. age (in years), sex and years of schooling] and all direct paths between the covariate and the EF indicators are fixed at zero (allowing correlation between factors to be freely estimated). Model fits were assessed following the same recommendation described above for the measurement model: MIs in the Mplus output were inspected for any direct effect of the covariates on indicators that should be freely estimated in the model, which would mean that the indicator showed Differential Item Functioning (DIF, indicated by a MI > 4 for an indicator on the covariate; Brown, 2015). The final model was found when the inspection of MIs retrieved no DIFs, which means that response probabilities to the remaining items do not differ across the values of the covariate (with equivalent value of the factor trait). Population heterogeneity was also interpreted in MIMIC models, which are the effects of the covariate (i.e. schooling, age and sex) on the latent traits. Because we tested categorial effects of sex, for descriptive purposes we determined Cohen *d* effect sizes comparing mean±SD) schooling and age of men and women using a free software (https://www.socscistatistics.com/effectsize/default3.aspx).

RESULTS

The databank is available upon request to the corresponding author (see Data Availability Statement). The statistical scripts and preprint can be found at https://doi.org/10.17605/OSF.IO/83YHU. Descriptive statistics of all variables can be found in Table 1. The following number of outliers were found in the data per indicator and were excluded: four each for the Plus-Minus task, Adjacency and Mean Repetition Gap RNG indices, five for the Trail Making Task and the TPI index, 12 for the coupon index, 20 for the runs index, 21 for the Redundancy index. There were many patterns of missing data which were dealt with by MLR estimator. Among these there were 31 participants who were excluded from the databank because they exceeded the 300 s limit to complete the Trail Making Task.

<Table 1 near here>

Determining EF model configuration

Model fit of the tested model configurations can be found in Table 2. The unifactorial model and two-factor models combining latent domains two-by-two converged but had low fit, even after considering MIs, with the exception of the two-factor model with inhibition merged with switching, which presented good fit with the inclusion of a correlation between the residuals of the TPI and Runs indicators and between Trail Making Task and Plus-Minus indicators indicated by inspection of MIs. This solution yielded a significant but low correlation between the two factors (r=0.12). Factor loadings on individual tasks ranges were: inhibition-shifting: λ=0.23 to λ =-.95 and updating: λ=.51 to λ=.87).

<Table 2 near here>

The three-correlated factor model converged with a non-acceptable chi-square, TLI and CFI fit indices. Inspection of MIs showed that including a correlation between the residuals of the two inhibition scores (Runs and TPI) would represent an increase in model fit (MI = 33.88). The final model (Table 2; Figure 1) including this correlation showed a good fit for all indices except for the chi-square, which is not a limitation because it is sensitive to sample size. The solution retrieved significant correlations between the factors (ranging from r=-0.10 to r=0.39). Factor loadings ranges on individual tasks were all above 0.40: inhibition: λ=-0.41 to λ =-.95; shifting: λ=.50 and λ=.89 and updating: λ=.51 and λ=.85). The bifactor solutions [bifactor and bifactor-(S-1)] did no converge (Table 2).

Because both the two-factor (inhibition merged with shifting) and three-factor models presented acceptable fit, we compared them with a qui-square difference test. A significant decrease in the two-factor model fit ($\Delta\chi^2(1)=10.29$, p=0.001) was verified so we used the three-factor solution to explore invariance to demographic effects.

< Figure 1 near here>

Invariance across age, schooling and sex

Regarding invariance testing for the good-fitting trifactor model, the MIMIC analysis including years of schooling, age and sex as covariates showed poor fit, but inspection of MIs indicated DIF of the Trail Making indicator on sex. The model freeing the estimation of this path showed acceptable fit but still showed DIF of the Trail Making on age. By freeing the direct path of Trail Making on age the model had good fit, but the Trail Making Task showed again DIF on schooling. The model including a path of the Trail Making Task indicator on schooling converged with a good fit (see Figure 2) and retrieved no more DIFs. Inspection of the final model showed that performance in the Trail Making task increased with schooling (β=0.29, $p<0.001$), decreased with age (β=- 0.31, $p<0.001$) and the sex effect was not significant (β=0.12, p=0.17). Hence, this model presented partial invariance regarding only one of the eight used indicators. Note that at each of these steps the Plus-Minus task indicator also showed the same MI values as that of the Trail Making task. The reason we chose to free the paths of the latter was that inspection of the standardized expected parameter change index (StdYX E.P.C.), which reflects the expected parameter value in the absence of the equality constraint, revealed it to be lower for the Plus-Minus task for all three covariates.-

<Figure 2 near here>

As to population heterogeneity (Figure 2B), a higher number of years of schooling was positively and significantly associated with scores of all the three latent factors: shifting (β=0.48, $p<0.001$), inhibition (β=0.17, $p<0.001$) and updating (β=-0.08, p=0.044). Linear age effects were only found for the shifting latent factor (decreased with age: β=-0.18, $p<0.001$) and women had lower scores in the inhibition (β=-0.30, $p<0.001$) and shifting (β=-0.61, $p<0.001$) latent factors than men. This occurred despite women being slightly more educated [women = 9.53±4.37 years of schooling; men (mean±SD)= 8.61± 4.13 years of schooling; Cohen *d*=0.22) and less than a year younger (women: 41.79±14.26 years; men: 42.59±14.99 years; Cohen *d*=0.05), on average, but these differences were of very small effect sizes.

DISCUSSION

In this study, which involved a diverse Brazilian sample spanning all of adulthood that varied widely in terms of educational attainment, we obtained a good fitting three-factor model, successfully replicating one of the most common configural solutions of the EF unity and diversity framework in adults (see Karr et al., 2018). The fact that this was found before mostly in samples from higher income nations suggests commonalities in the pattern of fractionation and inter-associations of EF domains across countries in adulthood (Barrett, 2020; Liebal & Haun, 2018), despite the influence of sociocultural and other environmental factors in the development of these cognitive abilities (Gustavson et al., 2018; Schirmbeck et al., 2020). Furthermore, as predicted, schooling was positively associated with all three tested EF latent domains, confirming the well-known negative impact of low education (and/or adverse socioeconomic environments) on EF in adults (Lavrencic et al., 2018; Roldán-Tapia et al., 2012; Wray et al., 2020). We also found negative age effects, but only for shifting, and an unexpected sex effect on shifting and inhibition favoring men. Next, we detail our results.

Latent EF factor model configuration

We tested many EF latent factor model configurations which have been reported in the literature, only some of which usually converge and present adequate fit to the data in each publication (Karr et al., 2018). Differences among studies in this respect can be explained by many features that influence CFA results which are: 1) analytic in nature (e.g. choice of model parameters/identification, the number of factors/subfactors to be confirmed, cutoff values of fit indices (Karr et al., 2018); 2) related to the particular set of EF indicators that are used (see Glisky et al., 2021; Karr et al., 2018; Yangüez et al., 2023); and 3) associated with sample characteristics (Karr et al., 2018). We will address each of these features in turn to discuss our results.

First, rates of acceptance (convergence and adequate fit) of different types of model configurations in the EF unity-diversity literature are overall quite low (Karr et al., 2018), particularly convergence of bifactor models which, when they do converge, tend to have an overfitting tendency (Yangüez et al., 2023). Low convergence of this configuration explains why our bifactor was non-convergent, even when we added correlations between the shifting and updating factors in the bifactor-(S-1) model, which is today deemed to be psychometrically necessary (Eid et al., 2017). The latter model, to our knowledge, was only tested before in typically developing adolescents and also failed to converge (see Segura et al., 2023). On the other hand, the fact that converging models overfit to data can account for why EF bifactor models are often selected as the best fitting ones in many papers that tested adult samples (see Karr et al., 2018).

Regarding EF indicators, many publications in this field investigate EF domains that do not map onto the ones proposed by the proponents of the theoretical framework under study (Miyake et al., 2000). Examples are testing distinct types of inhibition, working memory capacity instead of updating, and/or including extra domains such as access to long term memory (for a discussion on this issue, see Zanini et al., 2021). Hence, the fact that studies differ in terms of model configurations should not come as a surprise because they may have tested different abilities that inter-associate in different ways. Nonetheless, it is notable that in our diverse sample the good fitting three-correlated factor solution corroborated the interrelation and separability of inhibition, shifting and updating that was proposed in the initial version of the unity-diversity framework (Miyake et al., 2000) and that was also found in many other studies that considered the *same* EF constructs as we did, yet were obtained in widely different adult populations that varied in terms of sample size, sociocultural aspects and age (e.g. Feng et al., 2022; Fisk & Sharp, 2004; Friedman et al., 2016; Glisky et al., 2021; Ma et al., 2023; Miyake et al., 2000; Vaughan & Giovanello, 2010). This is even more noteworthy considering another issue that influences model configurations and can contribute to variability in latent models among studies (see Glisky et al., 2021; Karr et al., 2018; Yangüez et al., 2023): despite the similarity in EF constructs tested across these studies and their use of similar sets of EF tests (see Ma et al., 2023; Maldonado et al., 2020), which were not the same as ours, these publications often varied in terms of EF *indicators* obtained from the tasks (e.g. reaction time, accuracy, trial-by-trial variability, or other metrics).

However, the correlations among the three factors in our study were much lower than those reported in many previous publications in young (e.g. Friedman et al., 2016; Friedman & Miyake, 2017; Miyake et al., 2000) and older adults (e.g. Ma et al., 2023; Vaughan & Giovanello, 2010) that explored the same types of EF domains as we did. This was by no means exclusive to the present study, nor to specific ages in adulthood. For example, lower correlations were also found by Glisky et al. (2020) in 18-32-year-olds from the United States, and by Fisk and Sharp (2004), whose sample was from the United Kingdom and ranged in age from 20 to 81 years [although they used a different psychometric approach (principal component analysis) and included one more EF domain (access to long-term memory)]. Feng et al. (2022) had a mixed pattern of results in young, highly educated Han Chinese: a high correlation between inhibition and updating but low ones between the other combinations of latent factors.

Although the trifactor model had the best fit, we also obtained a good fitting two-factor configuration (with inhibition and shifting merged), equivalent to the model reported by Glisky et al. (2020) and Vaughan and Giovanello (2010) in elderly from the United States of America, although merging inhibition and updating provided a similar (Vaughan & Giovanello, 2010) or better fit to the data in young (Feng et al., 2022) and older adult (Ma et al., 2023) Chinese. As we found both these configurations, which mapped onto those of other studies that used different EF measures of the same domains, point to de adequacy of our choice of tasks/indicators to test the theoretical framework of Miyake et al. (2000) for our sample. This was reinforced by the finding of invariance to demographic variables at the task level (except for one of eight indicators), signaling that the same EF constructs (Chen, 2008) were measured irrespective of participants' schooling, age and sex, which is seldom tested for in this particular literature. Previous attempts to propose EF test batteries based on the same theoretical framework for samples from low- and/or middle-income countries did not include enough measures to compute latent variables for the three EF domains and involved paying for access to tasks (e.g. Wray et al., 2020), or were openly available, but very time consuming (e.g. Zanini et al., 2021), unlike our 10 minute-long, non-automated and free access battery.

The third aspect that influences CFA model configurations (i.e. demographic differences among samples) must also be considered in explaining our results, particularly regarding age. Some studies suggest that the tested latent factors may operate more independently in young versus older adults, with a de-differentiation of these abilities as people grow older, possibly because they develop different strategies to optimize performance (Glisky et al., 2021). Although we tested 18-88-year-olds, unlike other samples which varied much less in age ranges, we did not test different model configurations for different age groups. The reason was that this would have depended on establishing arbitrary age cutoffs (which differ from study to study, likely introducing bias in age effects) and because this would not have allowed us to determine demographic effects for the whole sample, which are described in the next section. Notwithstanding, the separability of the three domains was observed considering our whole sample, similarly to reports in other studies on older adults, which also found similarly-fitting two-model configurations with different combinations of latent EF domains (e.g. Ma et al., 2023; Vaughan & Giovanello, 2010).

Therefore, the hypothesis of age-related de-differentiation is in need of further support, as claimed by Karr et al. (2018).

Hence, our finding of a trifactor model in a diverse sample in terms of age and schooling speaks to the usefulness of the unity-diversity framework across adulthood applied to a broader set of samples, at least when it comes to considering the specific EF domains following the theoretical perspective of Miyake et al. (2000).

Demographic effects

The fact that we found invariance to demographic effects at the EF indicator level provided evidence that the same theoretical constructs were assessed across participants of different ages, who had different educational attainment and were of either sex, conferring a good foundation for testing the influences of these demographic effects on EFs at the latent trait level.

In spite of the wide age difference among participants, only shifting was linearly negatively related to age, so we did not find the age-related decline in all EFs domains reported in a meta-analyses of age effects on EFs (Maldonado et al., 2020). This meta-analysis, however, contrasted performance of young (18-35-year-olds) versus that of older individuals (>60 years) separated into groups based on arbitrary age cutoffs (as in Glisky et al., 2021) and considered EF scores across studies without taking their inter-associations and invariance testing into consideration, differently from our approach. Our lack of effects regarding updating concurred, nonetheless, with the fact that this domain seems to be the least affected by age (Maldonado et al., 2020). As to inhibition, our results are more in line with the idea that performance in this domain may not be linearly impaired as people age, as observed in a study published after the abovementioned meta-analyses. Verissimo et al. (2021), who used a large sample of 58-98-year-old Taiwanese, found that performance that indexes efficiency in inhibitory control (using the classic Flanker task) actually improved from the age of 58 to mid/late 70s, after which it declined (Veríssimo et al., 2021). Because a considerable portion of our sample was younger than 58 years-old, it is possible that this non-linear effect in older individuals was not picked up in our linear analysis.

Concerning schooling, we found that higher education was related to significantly better latent shifting, inhibition and updating traits, corroborating the positive link between cognitive stimulation/socioeconomic factors and EF performance in adults (Fisk & Sharp, 2004; Lavrencic et al., 2017; Wray et al., 2020), which is also described for other ages and/or cognitive abilities in samples from different nations (Amin et al., 2022; Ardila et al., 2010; Farah, 2017; Franzen et al., 2020; Hackman et al., 2015; Lawson et al., 2014; Segura et al., 2023; Schirmbeck et al., 2020). Because EFs are related to many outcomes in life and susceptible to enhancement following many types of manipulations, researchers have proposed that educational initiatives and/or interventions that stimulate EFs may help improve health (Baggeta & Alexander, 2016; Dias et al., 2024) and break the intergenerational reproduction of poverty and deprivation (de Neubourg et al., 2018).

Regarding sex differences, we found that women had worse shifting and inhibition latent scores compared to men, which were unexpected effects considering that literature reviews point to no consistent sex/gender effects on EFs (Gaillard et al., 2020; Grissom & Reyes, 2019); when sex differences are observed, they are usually secondary to differential sex-related cognitive strategies and/or spatial abilities that are not executive in nature, despite being recruited to perform many types of EF tasks (Grissom & Reyes, 2019) that are quite different from those used here. Our sex effects can therefore be better accounted for by an important issue that is overlooked in studies from high income countries, namely occupational sex-related discrimination or segregation favoring men, which is rife in Brazil (King, 2009; Tonet Maciel, 2021), and could be more pronounced in rural areas (King, 2009), such as in Baependi where the participants of this study reside. Although these sex/gender occupational differences cannot usually be explained by differences in human capital (King, 2009), this type of segregation means that women often take on less cognitive demanding jobs, especially those who have lower schooling. As the shifting Trail Making and Plus-Minus metrics and the inhibition indexes of the RNG task depend on numeracy and literacy (sequencing numbers and letters, mental calculations and knowledge of ways in which numbers can be organized), we tentatively suggest that a portion of the tested women could have been more cognitive overloaded while doing these tasks, resulting in poorer scores. This is a form of lower cognitive reserve, which is found to affect cognitive performance across adulthood (Roldán-Tapia et al., 2012). Educational and age differences were probably not to blame for these sex-effects because the model controlled for these variables, their covariance was not significant and, on average, men and women had similar years of schooling and age. In contrast, participants of both sexes did not differ in the ability to keep the last generated numbers in mind, indexed by the updating indicators from the RNG task (Towse & Neil, 1998), a domain that was nonetheless sensitive to schooling, as mentioned above.

Overall, we found that the three tested EF domains were not equally sensitive to the tested demographic effects, shifting having been the most affected one by education, age and sex. We suggest some possible reasons for this. The first is that there seems to be something particular about shifting, as discussed at length by Karr et al. (2018). For instance, the gradual differentiation of EFs (from a single general domain in early life to three domains in early adulthood) shows shifting to emerge first as a specific factor around the passage from childhood to adolescence (Karr et al., 2018). A study in elderly participants also found that, among the three EF domains, only shifting had genetic markers linked to dopamine D2 receptor genes (Ma et al., 2023). Because changes in this receptor occur as people become older (Ma et al., 2023), this may make shifting more prone to pick up age effect, as we observed. Additionally, Feng et al. (2022) found other genetic markers for inhibition, updating and for a general EF factor that were not present for shifting, which is further evidence that the EF domains tested here have different biological underpinnings (Friedman & Miyake, 2017; Friedman & Robbins, 2021; Rodríguez-Nieto et al., 2022).

Another non-exclusive explanation for the higher sensitivity of shifting to demographic variables has to do with processing speed, which is related to but does not account for the totality of EF performance (Glisky et al., 2021; Maldonado et al., 2020; Wray et al., 2020). The reason is that the tasks used to assess this domain (Plus-Minus and Trail Making task), unlike the RNG task, were self-paced, meaning that participants could take as long as they needed to respond to each stimulus. A possible consequence of this is that performance may have been more varied among participants, allowing the shifting latent factor to better capture impairment in older individuals, partly due to age-related slowing of processing speed (Glisky et al., 2021; Maldonado et al., 2020; Roldán-Tapia et al., 2012). Whether speed effects could explain part of the negative effects of education on shifting is less clear because no consistent speed effect has been found for individuals of different schooling levels [Staff et al. (2016) found positive associations but Lavrencic et al. (2018) did not (see also Veríssimo et al., 2021)]. Speed is also unlikely to explain why women scored lower in inhibition because they had similar updating abilities to men, assessed with indicators obtained from the same nonself-paced RNG task. Furthermore, there is no general male advantage in processing speed across the lifespan (Camarata & Woodcock, 2006; Nooyens et al., 2022).

General results, limitations and future perspectives

In sum, we found that a short, nonautomated, copyright-free test battery that can be administered by researchers with little training fit the theoretical and psychometric requirements to test the unity and diversity of inhibition, shifting and updating in a sample from a developing country across adulthood and that varied in educational attainment. This battery may also prove to be adaptable for other populations that are seldom represented in EF studies, allowing: 1) cross-cultural studies which can inform on the role of cultural and sociodemographic factors on EFs in a theory-oriented manner (Dias et al., 2024) and; 2) studies that aim to develop appropriate socio-culturally sensitive interventions to diminish the effects of impaired EFs on various types of outcomes (see Baggeta & Alexander, 2016; Dias et al., 2024; de Neubourg et al., 2018).

As to limitations of this study, although our shifting indicators were obtained from commonly used tasks (Trail Making and Plus-Minus), a novelty that requires replication was our use of inhibition and updating indicators derived from a single test (RNG task). We do not believe this constitutes a limitation because moderate factor loading of these indicators on their respective domains was found and these latent factors were only very weakly intercorrelated, indicating that they measured different cognitive abilities, as suggested by others (Audiffren, 2009; Chatzopoulos et al., 2021; Friedman & Miyake, 2004; Miyake et al., 2000; Oomens et al., 2015; Towse & Neil, 1998). Our rate-correct (Vandierendonck, 2017) shifting score indicators controlled for speed-accuracy trade-offs (Yangüez et al., 2023), so this was not a limitation of our study, but no such measure could be derived from the RNG task that was not self-paced.

Ideally, all indicators should have been similarly controlled for processing speed effects, which is a general limitation of the literature that tested the EF unity diversity framework (Maldonado et al., 2020; Schweizer et al., 2021; Yangüez et al., 2023). We also did not analyze different model configurations considering different patterns of missing data, nor treated the data for omission errors (see Schweizer et al., 2021; Towse & Neil, 1998) because there is no consensual way to do so in the unity-diversity literature. Although schooling is regarded as a measure of cognitive reserve (Lavrencic et al., 2018; Staff et al., 2016), other variables associated with this construct like occupation (e.g. Roldán-Tapia et al., 2012) and engagement in leisure activities across life were not available for our sample and might have yielded additional explanations for our patterns of demographic effects. Our sample was also not screened for clinical conditions or cognitive impairment, which could have influenced EF performance, and was a family-based cohort, in which individuals were related to different extents. This was not controlled for either due to a lack of validated, widespread use of analytic confirmatory structural equation approached to deal with extended pedigree designs.

Finally, to improve upon our findings, some alterations to the test battery/indicators may increase reliability, especially for samples with low education. First, only the Trail-Making task score was not invariant across age/schooling, which was not entirely surprising: different levels of familiarity with numbers and the alphabet across people with different educational attainment, which is lower in Brazil in older individuals (Kang et al., 2021), influences performance differently when tasks involve these abilities (see Ardila et al., 2010; Deloche et al., 1999). Other Trail Making tasks that do not include letters are also available (Kim et al., 2014) and may be an alternative. Non-paced number generation (Oomens et al., 2015) can also substitute the self-paced version of the RNG administered here. Future studies should also strive to use more sophisticated analyses that take non-linear (e.g. Oomens et al., 2015) and/or trial-by-trial variability into account (e.g. Yangüez et al., 2023), only possible if tasks are automated, which was not the case of our measures that were designed to be feasible for researchers with low funding from developing countries.

FUNDING DETAILS:

This study was supported by FAPESP (fellowships to author IAS: #2019/19709-6), CNPq (fellowship to author SP: # 309584/2023-0), CAPES (finance code 001) and AFIP.

DATA AVAILABILITY STATEMENT

The data cannot be shared publicly due to potential identification of participants' information, constrained by the local Ethics Committee, even if participants are unidentified. The study protocol was approved by the Ethics Committee of the Faculdade de Medicina of the Universidade de São Paulo, Brazil (#245-15). Data access inquiries can be sent to ttaporoski@harvard.edu.

DISCLUSURE OF INTEREST

The authors declare that they have no competing personal relationships and/or financial interests that could influence the work reported here.

REFERENCES

Amin, V., Behrman, J., Fletcher, J. M., Flores, C. A., Flores-Lagunes, A., & Kohler, H. P. (2022). Does schooling improve cognitive abilities at older ages: causal evidence from nonparametric bounds.

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., ... & Rosselli, M. (2010). Illiteracy: the neuropsychology of cognition without reading. *Archives of clinical neuropsychology*, *25*(8), 689-712.

Audiffren M. (2009). Acute exercise and psychological functions: a cognitive-energetic approach. In McMorris T., Tomporowski P. D., & Audiffren M. (Ed). *Exercise and Cognitive Functions*, (pp. 3-39), John Wiley & Sons

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education*, *10*(1), 10-33.

Balconi, M., Acconito, C., Allegretta, R. A., & Crivelli, D. (2023). What Is the Relationship between Metacognition and Mental Effort in Executive Functions? The Contribution of Neurophysiology. *Behavioral Sciences*, *13*(11), 918.

Barrett, H. C. (2020). Towards a Cognitive Science of the Human: Cross-Cultural Approaches and Their Urgency. Trends in Cognitive Sciences, 24(8), 620–638. https://doi.org/10.1016/j.tics.2020.05.007

Brown, T. A. (2015). Confirmatory factor analysis for applied research (Second edition). The Guilford Press.

Bucks R.S. (2013) Trail-Making Test. In: Gellman M.D., Turner J.R. (eds) Encyclopedia of Behavioral Medicine. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1005-9_1538

Camarata, S., & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, *34*(3), 231-252.

Chatzopoulos, G., Dinou, M., & Proios, H. (2021). Performance of Children and Adolescents on Random Number Generation as a measure of Executive Functions. *Journal of Educational and Developmental Psychology*, *11*(1), 1-37.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

De Neubourg, E., Borghans, L., Coppens, K., & Jansen, M. (2018). Explaining Children's Life Outcomes: Parental Socioeconomic Status, Intelligence and

Neurocognitive Factors in a Dynamic Life Cycle Model. Child Indicators Research, 11(5), 1495–1513. https://doi.org/10.1007/s12187-017-9481-8

Deloche, G., Souza, L., Braga, L. W., & Dellatolas, G. (1999). A Calculation and Number Processing Battery for Clinical Application in Illiterates and Semi-Literates. Cortex, 35(4), 503–521. https://doi.org/10.1016/S0010-9452(08)70815-3

Dias, N. M., Helsdingen, I. E., Lins, E. K. R. M. D., Etcheverria, C. E., Dechen, V. D. A., Steffen, L., Cardoso, C. D. O., & Lopes, F. M. (2024). Executive functions beyond the "Holy Trinity": A scoping review. Neuropsychology, 38(2), 107–125. https://doi.org/10.1037/neu0000922

Egan, K. J., Von Schantz, M., Negrão, A. B., Santos, H. C., Horimoto, A. R., Duarte, N. E., ... & Pereira, A. C. (2016). Cohort profile: the Baependi Heart Study—a family-based, highly admixed cohort study in a rural Brazilian town. *BMJ open*, *6*(10), e011598.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. Psychological Methods, 22(3), 541.

Farah, M. J. (2017). The neuroscience of socioeconomic status: Correlates, causes, and consequences. *Neuron*, *96*(1), 56-71. Friedman NP, Miyake A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *J. Exp. Psychol. Gen.* 133(1):101-35.

Feng, J., Zhang, L., Chen, C., Sheng, J., Ye, Z., Feng, K., ... & Xue, G. (2022). A cognitive neurogenetic approach to uncovering the structure of executive functions. *Nature communications*, *13*(1), 4588.

Fisk, J. E., & Sharp, C. A. (2004). Age-Related Impairment in Executive Functioning: Updating, Inhibition, Shifting, and Access. Journal of Clinical and Experimental Neuropsychology, 26(7), 874–890. https://doi.org/10.1080/13803390490510680

Franzen, S., van den Berg, E., Goudsmit, M., Jurgens, C. K., Van De Wiel, L., Kalkisim, Y., ... & Papma, J. M. (2020). A systematic review of neuropsychological tests for the assessment of dementia in non-western, low-educated or illiterate populations. *Journal of the International Neuropsychological Society*, *26*(3), 331-351.

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.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex, 86, 186-204

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.

Friedman, N. P., & Robbins, T. W. (2021). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, 1-18.

Gaillard, A., Fehring, D. J., & Rossell, S. L. (2021). A systematic review and meta-analysis of behavioural sex differences in executive control. European Journal of Neuroscience, 53(2), 519–542. https://doi.org/10.1111/ejn.14946

Glisky, E. L., Alexander, G. E., Hou, M., Kawa, K., Woolverton, C. B., Zigman, E. K., ... & Ryan, L. (2021). Differences between young and older adults in unity and diversity of executive functions. *Aging, Neuropsychology, and Cognition*, *28*(6), 829-854.

Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: Evaluating gender and sex differences in executive function. *Neuropsychopharmacology*, 44(1), 86–96. https://doi.org/10.1038/s41386-018-0179-5

Gustavson, D. E., Panizzon, M. S., Elman, J. A., Franz, C. E., Reynolds, C. A., Jacobson, K. C., ... & Kremen, W. S. (2018). Stability of genetic and environmental influences on executive functions in midlife. *Psychology and Aging*, *33*(2), 219.

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science*, *18*(5), 686-702.

Kang, T. H., Paese, L. H. Z., & Felix, N. F. (2021). Late and unequal: measuring enrolments and retention in Brazilian education, 1933-2010. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History*, *39*(2), 191-218.

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

Kessels, R. P. (2019). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *The Clinical Neuropsychologist*, *33*(2), 357-368.

King, M. C. (2009). Occupational segregation by race and sex in Brazil, 1989–2001. *The review of black political economy*, *36*(2), 113-125.

Kim, H. J., Baek, M. J., & Kim, S. (2014). Alternative type of the trail making test in nonnative English-speakers: the trail making test-black & white. *PLoS One*, *9*(2), e89078.

Lavrencic, L. M., Churches, O. F., & Keage, H. A. (2018). Cognitive reserve is not associated with improved performance in all cognitive domains. *Applied Neuropsychology: Adult*, *25*(5), 473-485.

Lawson, G. M., Hook, C. J., Hackman, D. A., Farah, M. J., Griffin, J. A., Freund, L. S., & McCardle, P. (2014). Socioeconomic status and neurocognitive development: Executive function. *Executive Function in Preschool Children: Integrating Measurement, Neurodevelopment, and Translational Research. American Psychological Association*, 1-28.

Liebal, K., & Haun, D. B. M. (2018). Why Cross-Cultural Psychology Is Incomplete Without Comparative and Developmental Perspectives. Journal of Cross-Cultural Psychology, 49(5), 751–763. https://doi.org/10.1177/0022022117738085

Ma, J., Tang, L., Peng, P., Wang, T., Gui, H., & Ren, X. (2023). Shifting as an executive function separate from updating and inhibition in old age: Behavioral and genetic evidence. *Behavioural Brain Research*, *452*, 114604.

Maes J. H. R., Eling P. A. T. M., Reelick M. F., & Kessels R. P. C. (2011). Assessing executive functioning: On the validity, reliability, and sensitivity of a click/point random number generation task in healthy adults and patients with cognitive decline. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 366-378.

Maldonado, T., Orr, J. M., Goen, J. R., & Bernard, J. A. (2020). Age differences in the subcomponents of executive functioning. *The Journals of Gerontology: Series B*, *75*(6), e31-e55.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognive Psychology,* 41(1):49-100.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8-14.

Muthén, L. K., & Muthén, B. (1998). Mplus User's Guide (8th ed.). Muthén & Muthén.

de Neubourg, E., Borghans, L., Coppens, K., & Jansen, M. (2018). Explaining children's life outcomes: parental socioeconomic status, intelligence and neurocognitive factors in a dynamic life cycle model. *Child Indicators Research*, *11*, 1495-1513.

Nooyens, A. C., Wijnhoven, H. A., Schaap, L. S., Sialino, L. D., Kok, A. A., Visser, M., ... & van Oostrom, S. H. (2022). Sex differences in cognitive functioning with aging in the Netherlands. *Gerontology*, *68*(9), 999-1009.

Oomens, W., Maes, J. H., Hasselman, F., & Egger, J. I. (2015). A time series approach to random number generation: using recurrence quantification analysis to capture executive behavior. *Frontiers in Human Neuroscience*, *9*, 319.

Reynolds, C. A., & Finkel, D. (2015). A meta-analysis of heritability of cognitive aging: minding the "missing heritability" gap. *Neuropsychology review*, *25*, 97-112.

Rodríguez-Nieto, G., Seer, C., Sidlauskaite, J., Vleugels, L., Van Roy, A., Hardwick, R., & Swinnen, S. (2022). Inhibition, Shifting and Updating: Inter and intra-domain commonalities and differences from an executive functions activation likelihood estimation meta-analysis. *Neuroimage*, *264*, 119665.

Roldán-Tapia, L., García, J., Cánovas, R., & León, I. (2012). Cognitive reserve, age, and their relation to attentional and executive functions. *Applied Neuropsychology: Adult*, *19*(1), 2-8.

Satorra, A., & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75, 243-248.

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, 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

Schweizer, K., Reiß, S., and Troche, S. (2019). Does the Effect of a Time Limit for Testing Impair Structural Investigations by Means of Confirmatory Factor Models? Educat. Psychol. Measurement 79, 40–64. doi: 10.1177/0013164418770824

Segura, I. A., Cogo-Moreira, H., Nouri, A., Miranda, M. C., & Pompeia, S. (2023). Cross-country (Brazil and Iran) invariance of fractionation of executive functions in early adolescence. *Journal of Cognition and Development*, 1-21.

Staff, R. T., Chapko, D., Hogan, M. J., & Whalley, L. J. (2016). Life course socioeconomic status and the decline in information processing speed in late life. *Social Science & Medicine*, *151*, 130-138.Tonet Maciel, F. (2021). Occupational segregation and wage differentials by gender and race in Brazil: evidence from a quantile decomposition. *International Journal of Manpower*, *42*(5), 810-841.

Tonet Maciel, F. (2021), "Occupational segregation and wage differentials by gender and race in Brazil: evidence from a quantile decomposition", International Journal of Manpower, Vol. 42 No. 5, pp. 810-841. https://doi.org/10.1108/IJM-06-2019-0277

Towse, J.N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behav. Res. Meth. Instrum. Comp.* 30(4):583-591.

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. Behavior Research Methods, 49(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. Psychology and Aging, 25(2), 343–355. https://doi.org/10.1037/a0017729

Veríssimo, L., Costa, M., Miranda, F., Pontes, C., & Castro, I. (2021). The Importance of Phonological Awareness in Learning Disabilities' Prevention: Perspectives of Pre-School and Primary Teachers. Frontiers in Education, 6, 750328. https://doi.org/10.3389/feduc.2021.750328

Wray, C., Kowalski, A., Mpondo, F., Ochaeta, L., Belleza, D., DiGirolamo, A., ... & Cohorts. (2020). Executive functions form a single construct and are associated with schooling: Evidence from three low-and middle-income countries. *PloS one*, *15*(11), e0242936.

Yangüez, M., Bediou, B., Chanal, J., & Bavelier, D. (2023). In search of better practice in executive functions assessment: Methodological issues and potential solutions. Psychological Review. https://doi.org/10.1037/rev0000434

Zanini, G. A., 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.

Table 1: Demographics and descriptive statistics of scores per executive function variable (indicator) according to each executive domain (shifting, inhibition and updating).

Variables	N (1253)		Mean (±SD)	Confidence Interval (±95%)	Minimum	Maximum
Demographics	**Females**	**Males**				
Age (years)	745	499	37.76 (13.10)	36.90/38.63	18	88
Schooling (years)	745	499	7.61 (4.71)	10.83/3.60	0	23
Random No. Generation task						
Inhibition indexes						
Turning Point Index (TPI)	521	356	81.68 (13.71)	80.77/82.59	40.62	123.44
Runs	508	354	1.23 (0.52)	1.19/1.26	0.08	3.34
Adjacency	520	358	29.30 (11.62)	28.53/30.07	2.04	63.92
Updating indexes						
Coupon	518	352	16.88 (4.68)	16.57/17.19	0	36.5
Redundancy	511	351	1.37 (1.01)	1.30/1.44	0.06	6.79
Mean Repetition gap	520	358	2.42 (1.46)	2.32/2.51	0	7
Shifting measures						
Trail Making task						
Accurate shifts (no.)	544	341	24.09 (1.18)	24.01/24.17	16	25
Speed (s)	545	340	123.26 (53.30)	119.74/126.77	44	296
Rate Correct Score (no./s)	543	340	0.23 (0.10)	0.23/0.24	0.05	0.57
Plus-Minus task						
Accurate shifts (no.)	485	371	27.55 (2.84)	27.36/27.75	12	30
Speed (s)	485	369	100.63 (40.36)	97.89/103.37	19	320
Rate Correct Score (no./s)	463	365	0.32 (0.12)	0.31/0.32	0.07	0.74

N.B.: the three indicators used in the statistical model to measure inhibition and updating were indexes from the Random Number (no.) Generation task; the two indicators used to measure shifting were accurate shifts between numbers and letters in the Trail Making Task part B and accurate shifts between adding and subtracting three to two-digit numbers in the Plus-Minus tasks, both of which were divided by the time in s taken to complete the tasks (Rate Correct Scores).

Table 2: Model fit for the tested Confirmatory Factor Analyses model configurations.

Model	χ^2 (df)	CFI	TLI	RMSEA	90% C.I.	SRMR
One factor	251.45 (18)*	0.81	0.70	0.102	[0.091, 0.113]	0.101
Two factors (updating merged with switching)	133.98 (18)*	0.90	0.85	0.072	[0.061, 0.083]	0.089
Two factors (updating merged with inhibition)	82.58 (14)*	0.95	0.90	0.058	[0.046, 0.070]	0.045
Two factors (inhibition merged with switching)	**54.92 (17)***	**0.97**	**0.95**	**0.042**	**[0.030, 0.055]**	**0.038**
Three factors	**42.57 (16)***	**0.98**	**0.96**	**0.036**	**[0.023, 0.050]**	**0.032**
Bifactor			No convergence			
Bifactor-(S-1)			No convergence			

Note: *p < 0.05. Model in bold presented acceptable fit (for fit cutoff scores, see text).

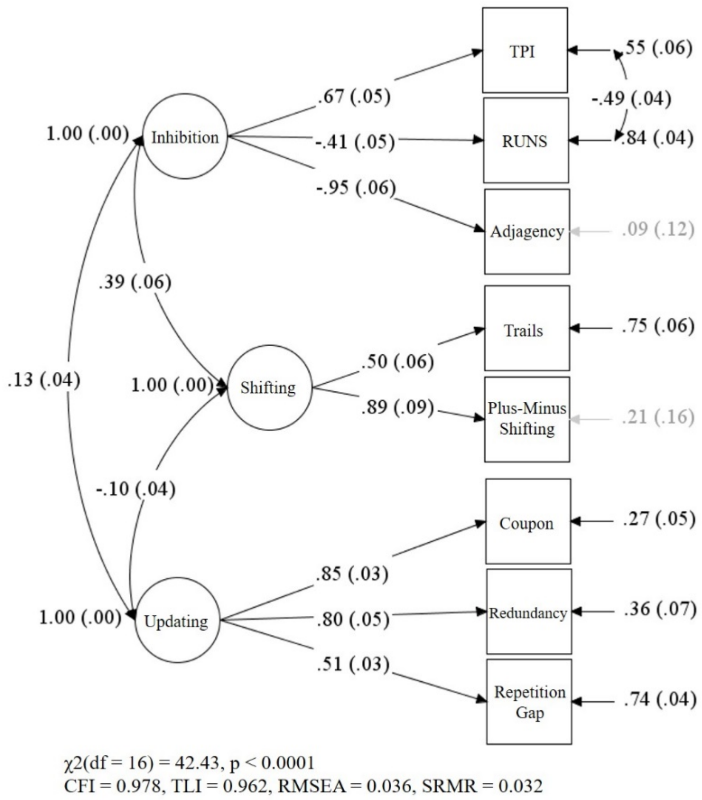

Figure 1: factor structure of the three-correlated executive functions factor model solution with its fit indices.

N.B.: circles represent latent factors, squares represent the indicators (raw scores) per executive domain. Curved, two-headed arrows linking latent factors indicate correlations, expressed as values on the arrows; values on the straight arrows pointed from the latent factors to the indicators are factor loadings and values on the right, next to the arrows pointing to each indicator are residuals. The curved, two-headed arrow linking residuals are a modification index (MI). Estimates are completely standardized. Non-significant values are represented in grey. EF indicators were: TPI=turning point index of the Random Number Generation task (RNG), Trails=rate correct score of the trail making test part B, Plus-Minus shifting=rate correct scores of shifting cost; repetition gap=mean repetition gap index of the RNG; the other indicators are indexes of the RNG. For fit indices, cut-off scores and details on the indicators, see text.

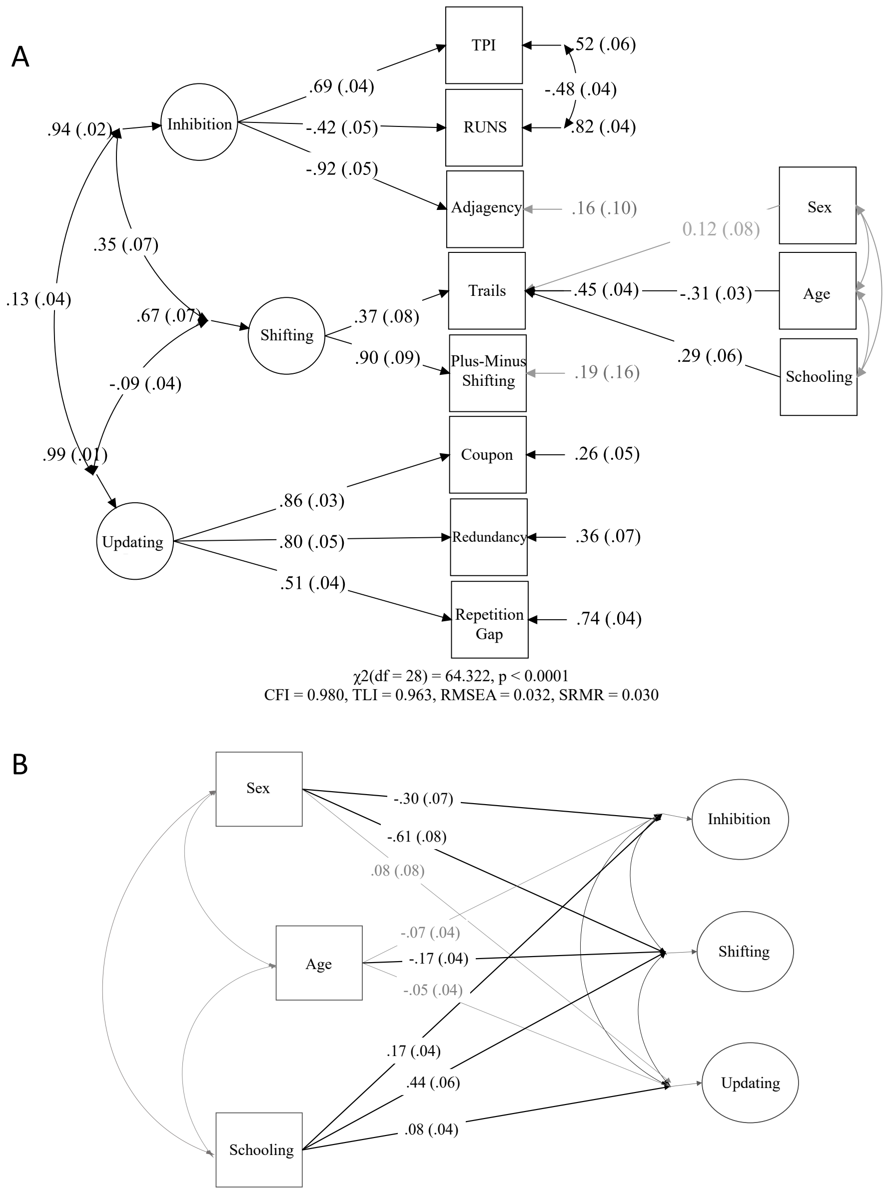

Figure 2: Invariance testing across age, sex and schooling and fit indices for the three-correlated latent factor model of executive functions, showing partial invariance. The model is presented in two parts to facilitate viewing. A (details the factor loadings of the three factors of executive functions on the tasks and the factors inter-correlations). To the left, the diagram shows the direct effects of the covariates on the indicator that presented Differential Item Functioning (i.e., Trails). B (diagram details the population

heterogeneity part of the model, showing the effects of the three covariates on the latent factors of executive functions).

N.B: this Multiple Indicators, Multiple Causes (MIMIC) model shows the effect of the covariates (squares to the left: age=age in years; schooling=years of schooling) on the three EF latent factors (ovals) of the three-correlated factor solution. The diagram A also displays factor loadings (values on the linear arrows) of executive tasks (squares in the middle) on the executive latent variables. 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 on A and covariances on B (left). Linear arrows on the diagram B indicates βs of the regressions. Estimates are completely standardized, except for the effects of sex on the factors and Trail Making task indicator, where partially standardized estimates (StdY) are shown. Fit indices are presented in the middle of the diagram (for fit indices cutoff scores, see text). Faded grey lines indicate non-significant effects ($p < 0.05$). EF indicators were: TPI=turning point index of the Random Number Generation task (RNG). Trails=rate correct score of the trail making test part B; Plus-Minus shifting=rate correct scores of shifting cost; repetition gap=mean repetition gap index of the RNG; the other indicators are indexes of the RNG. For fit indices, cut-off scores and details on the indicators, see text.