Validating the Oxford Cognitive Screen-Executive Function (OCS-EF) & the Rapid Assessment of Cognitive and Emotional Regulation (RACER) as a Unified Measure of Executive Function for Adolescents in Peru

Emiko Bell

Masters of Science in Psychological Research (JRPS)

2021

8394 words

Acknowledgements

I would first like to thank my supervisor, Professor Gaia Scerif, on her invaluable supervision, advice, and support throughout my dissertation. I would also like to thank Dr Alan Sanchez, Dr Marta Favara, and Monica Lizama from Peru GRADE team – my thesis would not have been possible without their aid and support. I also thank Dr Mihaela Duta and Professor Nele Demeyere for their support with the OCS-EF, and Professor Margaret Sheridan for her support with RACER. Thank you to Dr Kirsten Rowe for her pioneering in this research.

In addition, I would like to thank Professor Amina Abubakar, Atena Barat, Professor Kathleen Khan, Professor Charles Newton, Tshegofatso Seabi, Professor Alan Stein, Professor Stephen Tollman, and Dr Ryan Wagner from the MHEF (Mental Health and Executive Function) study team. Their support and input were invaluable. I would also like to thank the fieldworkers and the participants for making this study possible.

Thank you to St. Catherine's College and the Kobe Institute for generously funding my studies.

Thank you to my friends and family for their support throughout my writing. Thank you to Amelia Jeffrey, Meg Hobson, Pierre Osselin, and Gauthier Mourre for keeping me in good spirits during the COVID-19 pandemic. Thank you to my course mates, Irena Tetkovic, Jeevun Grewal, Gwen Williams, Luke Reeves, Katie Brown, and Orestis Zavlis, for the wonderful conversations and advice. Finally, I would especially like to thank my significant other, Nathan Tudor, for his vital support.

Table of Contents

Abstract

Research and validated measures of executive function (EF) in low- to middle-income countries (LMIC) are minimal compared to high-income countries, with even less of a focus on adolescent EF. Currently, a widely validated measure of EF is unavailable in LMICs to our knowledge, limiting further research on EF and the development of potential interventions. This study aimed to fill this research gap by establishing a feasibly administered EF assessment through cultural validation. We examined the validity of the Oxford Cognitive Screen – Executive Function (OCS-EF) and the Rapid Assessment of Cognitive and Emotional Regulation (RACER) as a combined EF assessment for use among Peruvian adolescents (N = 223). We examined construct validity using confirmatory and exploratory factor analyses and compared three-factor, two-factor, and unitary EF models, and we examined criterion validity using linear mixed models to establish whether the assessment negatively predicted attention-deficit/hyperactivity disorder (ADHD) scores using the Swanson, Nolan, and Pelham Teacher and Parent Rating Scale. The construct validity analyses revealed a one-factor model fit in line with previous research among LMIC adults. However, internal reliability measures indicated that the assessment tasks did not account for the majority of the variability in the latent factor. The criterion validity findings indicated that the OCS-EF and RACER did not significantly predict ADHD subscores. We recommend adaptations to the assessment and further research in LMIC contexts to establish OCS-EF and RACER as a valid measure of EF among LMIC adolescents.

Validating the Oxford Cognitive Screen-Executive Function (OCS-EF) & the Rapid Assessment of Cognitive and Emotional Regulation (RACER) as a Unified Measure of Executive Function for Adolescents in Peru

Background

Validated assessments for measuring executive function (EF) in adolescents from low- to middle-income countries (LMIC) are minimal, limiting further research and the development of EF-focused interventions in these contexts. Establishing a robustly validated measure is crucial as multiple studies revealed that EF is vital to development – adolescent deficits in EF have been associated with lower scholastic achievement, risky sexual and substance-use behaviours, and detrimental mental health conditions such as depression, anxiety, and attention-deficit/hyperactivity disorder (ADHD; Ahmed et al., 2019; Han et al., 2016; Martel et al., 2007; Wray et al., 2020). While multiple measures were validated and utilised in high-income countries (HIC), these measures must first be adapted and culturally validated for use in LMIC to establish that they are correctly measuring EF. Validating EF assessments across LMIC establishes a standardised method of measurement that can be utilised to further examine the impact of EF on behaviour and health.

EF is a group of interrelated cognitive skills that regulate goal-directed actions, emotions, thoughts, and behaviours (Carlson et al., 2013). The specific skills that constitute EF have been disputed but are generally understood as the neurocognitive self-regulatory functions relating to the prefrontal cortex (Carlson et al., 2013). However, EF remains difficult to measure and define (Miyake & Friedman, 2012). Although EF is commonly measured using a wide array of cognitive tasks assessing skills such as spatial and verbal working memory, problem-solving, cognitive flexibility, and inhibitory control, these tasks face the issue of "task-impurity" where non-EF variance, such as motor skills and articulation speed, can affect the measurement of EF (Miyake & Friedman, 2012; Nyongesa et al., 2019).

To account for this "task-impurity" problem and better define EF, recent research utilised a latent-variable approach to identify the underlying factors of EF by using multiple cognitive tasks and statistically extracting the factors using structural equation modelling (SEM; Miyake & Friedman, 2012). Using this method, common EF factor structures throughout the lifespan were identified, ranging from a unitary structure, a two-factor structure of hot and cool EF, and three core correlated but separable factors of updating, shifting, and inhibition (Hughes et al., 2009; Karr et al., 2018; Miyake & Friedman, 2012; Shing et al., 2010; Wiebe et al., 2011; Zelazo & Carlson, 2012). The unitary structure identifies a general EF skill among the cognitive tasks, while the hot and cool factors focus on EF in motivational or affective contexts compared to neutral contexts (Carlson et al., 2013; Zelazo & Carlson, 2012). The three correlated but separate EF factors include updating, or monitoring and maintaining working memory content; shifting, or flexibly switching between mental tasks; and inhibition, or purposefully stopping oneself or overriding a dominant response (Miyake & Friedman, 2012). A complementary model with a "common EF" factor that correlates with all EF tasks with the two correlated but independent factors of updating and shifting was also identified as an alternative unified three-factor model (Miyake & Friedman, 2012).

Crucially, the current literature highlights two research gaps. First, these EF factor structures were found to vary across different age groups, but adolescent factor structures have been investigated much less than either child or adult EF (e.g. Wiebe et al., 2011; Xu et al., 2013). EF development begins and continues from early childhood to adolescence and young adulthood before stabilising or declining in later adulthood lifespan, typically with increasing factors (Ferguson et al., 2021; Huizinga et al., 2006; Xu et al., 2013). The unitary EF structure is mostly supported in early to mid-childhood, while the hot/cool structure is more evident in late childhood to adulthood (Wiebe et al., 2011; Xu et al., 2013; Zelazo & Carlson, 2012). However, the evidence for a clear adolescent EF factor structure is mixed, and the unitary, hot & cool, and the three correlated factor structure of inhibition, shifting, and updating have all been observed in the adolescent population in HIC (Ferguson et al., 2021; Hartung et al., 2020; Karr et al., 2018; Lee et al., 2013; Miyake & Friedman, 2012; Xu et al., 2013). The gradual development of EF during adolescence could explain the variability of these results since the differentiation of EF is more observed in the adult population compared to children (Xu et al., 2013). Overall, further work is needed to understand the structure of EF in adolescence.

Secondly, and in addition to the limited focus on adolescence, EF was more rarely investigated in LMICs (Ford et al., 2019; Xu et al., 2013). Evidence in LMICs is limited, with a few studies examining the EF factor structure of adults in Guatemala, the Philippines, and South Africa; and older adolescent to young adult women in South Africa (Rowe, Duta, & Demeyere et al., 2021; Wray et al., 2020). These studies reveal that EF factor structures can also vary among LMIC populations – the unitary EF model best fit the adult population from Guatemala, the Philippines, and South Africa, while the three-factor model best fit the young adult population in South Africa (Rowe, Duta, & Demeyere et al., 2021; Wray et al., 2020). Other studies have utilised EF assessments in LMIC children and adolescents but have not examined the factor structure of EF (e.g. Chen et al., 2019; Ford et al., 2019; Rowe, Pozuelo, & Nickless et al., 2021; Willoughby et al., 2019). Further investigation is required to test alternative EF factor structures across multiple LMIC adolescent populations. The implication of identifying accurate EF structures and measurement in LMICs creates opportunities for intervention and prevention that are critical in highly volatile environments.

Understanding EF in LMIC adolescents is vital since deficits in EF has been linked to behavioural and clinical symptoms and potential interventions. Research from HICs revealed that lower EF scores in early childhood significantly predicted lower scholastic achievement in adolescence (Ahmed et al., 2019). Furthermore, higher errors in EF tasks among adolescents were linked to significantly higher depressive and anxiety symptoms (Han et al., 2016). Adolescents with diagnosed ADHD exhibited significantly impaired EF performance to their undiagnosed counterparts, regardless of ADHD subtype (Krieger & Amador-Campos, 2017; Martel et al., 2007). Adolescents with oppositional defiant disorder and conduct disorder also had impaired EF performance compared to healthy controls, accounting for ADHD comorbidity (Kleine Deters et al., 2020). Moreover, EF assessments significantly predicted engagement in risky substance, sexual, and antisocial behaviours in adolescents (Pharo et al., 2011). In the case of LMIC adolescents, challenges such as poverty and malnutrition have been found to impact EF in children and adolescents – such as missing important cognitive development milestones or lower performance on working memory tasks (Chen et al., 2019; McCoy et al., 2016; Sania et al., 2019). As a factor in these behavioural and clinical symptoms, EF can be targeted for potential interventions in adolescents, both for treatment and as an outcome measure. Investigations in LMICs show promise in EF intervention, with ADHD-diagnosed children and adolescents in Brazil showing improvement in working memory and inhibitory control after undergoing a self-regulation intervention program (Menezes et al., 2015). However, a robustly validated measurement is first needed in LMICs to properly assess EF and to investigate potential interventions.

In summary, EF in adolescents from LMIC is understudied, with most studies focusing on HICs. A construct established in a HIC context cannot be assumed to automatically transfer to LMICs and should be examined to be an equivalent and meaningful construct in another environment and culture (Matsumoto, 2003). Translation alone does not guarantee the exact application of the measure in different countries and environmental and cultural factors, such as familiarity with tablet-based assessments, can play a role in affecting the adaptation of measures and assessments. The reliability and validity of a measure should be empirically examined in the target context to establish convergent and predictive validity

(Matsumoto, 2003). Recommendations for robust cultural validation include cultural adaptation, translation and back-translation, piloting, and full psychometric testing measuring reliability, convergent/divergent validity, criterion validity, dimensionality, and model fit (Matsumoto, 2003; Sousa & Rojjanasrirat, 2011). Currently, validated EF measures for LMIC adolescents are limited, making further studies of EF difficult because of the lack of established EF assessments. A validated measure of EF must first be established to target adolescent EF in LMICs as a potential method of intervention.

Current Study

The purpose of this study is to validate an easily administered and unified EF measure in an LMIC context through secondary data analysis. Primary data were collected from adolescents in a province of Lima, Peru. EF was measured using the Oxford Cognitive Screen – Executive Function (OCS-EF) and the Rapid Assessment of Cognitive and Emotional Regulation (RACER), both tablet-based EF assessment tools. The OCS-EF was originally adapted to measure EF from OCS-Plus, a clinical tool to measure subtle cognitive impairments among stroke patients (Demeyere et al., 2021). The measure was specifically adapted for use in the LMIC adolescent population and was intended to be brief, inexpensive, and easily administered by trained lay fieldworkers (Rowe, Duta, & Demeyere et al., 2021). An earlier version of the OCS-EF was utilised and validated among older adolescent girls and young adult women from South Africa and was able to identify individual differences in a non-clinical sample (Rowe, Duta, & Demeyere et al., 2021). RACER was also developed for the LMIC population, initially developed to measure EF in LMIC children with limited resources and has further been utilised in Ghana, Lebanon, Niger, and Peru (Ford et al., 2019; Hamoudi & Sheridan, 2015; Yuan et al., 2020).

We utilised OCS-EF and RACER as a unified and relatively short assessment that measures multiple EF factors and incorporates both verbal and visuospatial tasks. The two assessments administered together have comprehensively assessed EF factors and were effective EF measurement tools in a small sample of South African HIV-positive adolescents (Rowe, Pozuelo, & Nickless et al., 2021). However, the combined current version of OCS-EF and RACER was not validated for a broader population of adolescents aside from the older adolescents and women in South Africa. Further validation outside of the South African context is needed to establish OCS-EF and RACER as a widely viable EF measure in LMIC contexts. This project was conducted to establish OCS-EF and RACER as a unified, easily administrable, and robust measure of EF, available for use in wider LMIC settings.

For our current study, we examined the construct and criterion validity of OCS-EF and RACER in the adolescent sample from Peru. For construct validity, exploratory and confirmatory factor analyses were conducted to account for the understudied nature of the LMIC adolescent population while examining the factor structure of EF in line with previous literature (Rowe, Duta, & Demeyere et al., 2021). For criterion validity, we examined the association of performance on EF tasks with ADHD scores in line with previous literature (Martel et al., 2007). Specifically, we hypothesised that:

1. Miyake and Friedman's (2012) three-factor model would best fit the data from OCS-EF and RACER as opposed to the single, unitary EF model (see Figure 1).
2. Lower OCS-EF and RACER task performance would predict higher average ADHD subscores, measured through the Swanson, Nolan, and Pelham-IV Teacher and Parent Rating Scale (SNAP-IV; see Figure 2).

In addition, we conducted exploratory analyses to account for the understudied nature of the LMIC adolescent population and to clarify any potential confounding variables that may interfere with the interpretation of the validity. We preregistered these hypotheses along with the methods and analysis plan to the Open Science Framework.

Figure 1

Diagram of Hypothesis One

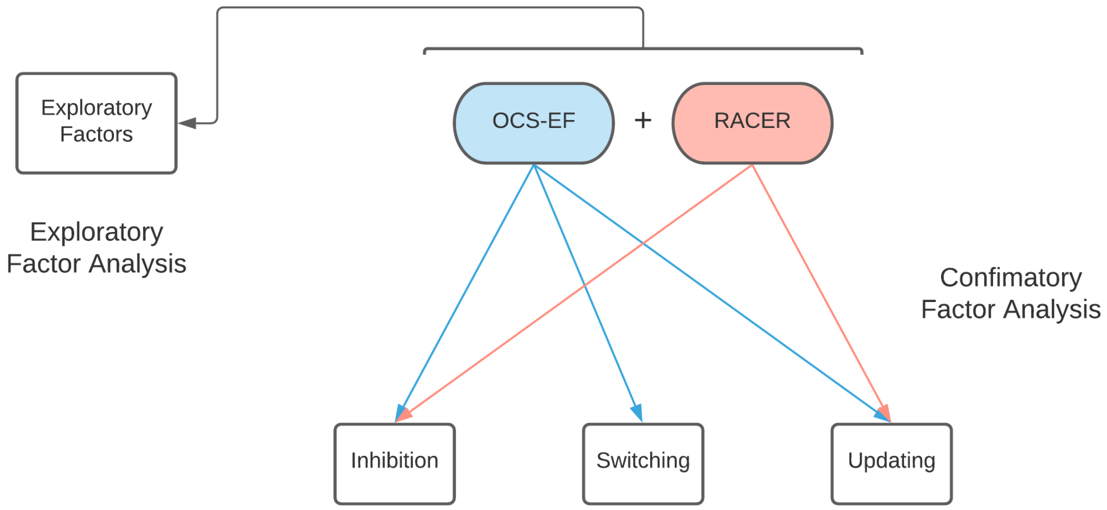
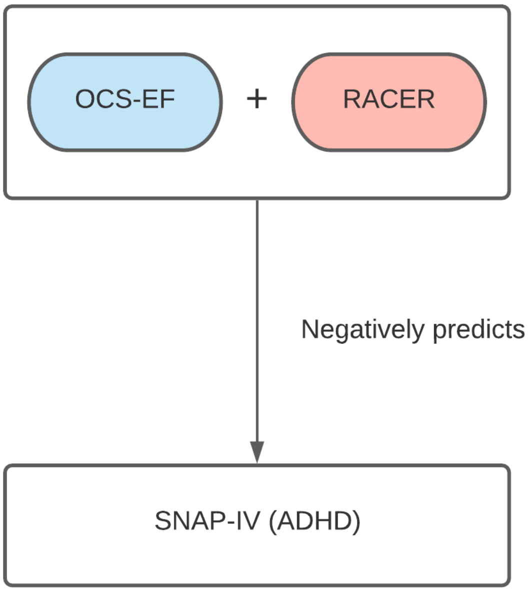


Figure 2

Diagram of Hypothesis Two



Method

Study Design

The study data were collected in 2017 across three international sites in Kenya, Peru, and South Africa in partnership with the Aga Khan University in Nairobi, Grupo de Análisis para el Desarrollo (GRADE), KEMRI Wellcome Trust, MRC/Wits-Agincourt Rural Public Health and Health Transitions Research Unit, and Young Lives as a part of the Mental Health and Executive Function (MHEF) study. The MHEF study collected data from adolescents in low- to middle-income countries listed above with the aim to examine the relationship between executive function, mental health, and socioeconomic status in each location. We only utilised the data from Peru for this study due to data transfer issues and in the interest of time. In Peru, 223 participants were recruited in the Cañete province as a representative sample of the study population. Eligibility criteria for the MHEF study included 14 to 22 year-old male and female participants residing in the province without intellectual, visual, or auditory disabilities or neurodevelopmental disorders. The participants were originally recruited for the Niños del Milenio project through the Young Lives study at the University of Oxford.

Ethical approval for the original study in Peru was granted by the Comité Institucional de Bioetica (CIB) VÍA LIBRE in Lima, and approval was also granted for this current study through the Oxford Tropical Research Ethics Committee (533-21). The participants provided informed consent for participation in the study. For minors, the caregivers provided consent and the participants also provided assent.

Sample Characteristics

The sample's demographic characteristics can be found in Table 1.

Table 1

Demographic Characteristics of the Sample

Characteristics			N
Sex	Male (%)	49.77	110
	Female (%)	50.23	111
	NA (%)	0.45	1
Age	Mean (*SD*)	17.53 (2.55)	
	Range	14-22	
Language	Spanish (%)	99.55	221
	Aymara (%)	0.45	1

Note. N = 223.

Materials & Procedure

Demographic data, cognitive assessments, and the SNAP-IV questionnaire were administered via interview by trained fieldworkers at the participants' homes. The participants completed each OCS-EF and RACER task using an Android tablet, either using a stylus or their finger. Local researchers translated all measures into Spanish, back-translated them, and verified them as culturally appropriate (e.g. ensuring the fruits and vegetables represented in the Selection task were common and familiar in the administered location). We assessed each task completed under OCS-EF and RACER using a single indicator score, which we determined by previous research conducted among LMIC women, adolescents, and children (Ford et al., 2019; Rowe, Duta, & Demeyere, 2021a; Rowe, Pozuelo, & Nickless et al., 2021). We listed the screen captures for each task in Appendix A.

OCS-EF

The current version of the OCS-EF was adapted from the OCS-Plus and was developed with eight tasks of Digit Recall, Spatial Span, Go/No-go, Trails, Rule Finding, Selection, the Iowa Gambling Task, and Figure Copy. The Iowa Gambling and Figure Copy tasks were omitted from the assessment administration due to participant fatigue.

For each task, fieldworkers assessed whether the participants had specific conditions during testing. The options were between "no issue", "incomplete", "interruptions", "technical problems", "fatigue", "visual problems", "motor problems", and "other". If any tasks were skipped, reasons for skipping were coded as "no time", "not relevant", "declined", "fatigue", "no understanding", "technical problems", "already assessed", or "other". We included data from the analyses only if "no issue" was reported in both conditions.

Digit Recall. This task was developed to assess verbal working memory. The fieldworkers instructed the participants to listen to a sequence of numbers read out loud with 1 second between each number. After the sequence was read, the fieldworkers then asked the participants to repeat the sequence of digits in the order that they were read. The fieldworkers entered the numbers onto the tablet after the sequence was recalled by the participant. The number of digits increased from two to nine with each successful trial, and the task terminated after three incorrect attempts at a certain length. In the second section, fieldworkers instructed the participants to listen to the sequence of numbers in the same manner, then repeat the numbers in reverse order after the numbers were no longer displayed. Identical to the first section, the sequence of digits increased with each successful attempt and terminated after three incorrect attempts. We used the backward span of the task to assess EF as it requires manipulating the memorised content, measuring working memory.

Spatial Span. This task was developed to measure visuospatial working memory. The fieldworkers instructed the participants to look at a set of blocks on the tablet screen, which played an animation with successively highlighted blocks in yellow. The fieldworkers instructed the participants to memorise the pattern in the order the blocks were highlighted, and then to tap on the blocks in the same sequence with their finger. The number of highlighted blocks in a sequence increased with participant performance (two to nine) or terminated after three incorrect attempts at a specific length. For the second testing section, the fieldworkers instead instructed the participants to tap the blocks in the reverse order that they were highlighted. The highlighted blocks increased with successful attempts and terminated after three unsuccessful attempts. We used the number of blocks correctly identified in the reverse order to test EF as a measure of working memory.

Go/No-go. This task was developed to measure inhibitory control and was adapted from an emotional Go/No-go task with faces and instead used shapes to account for wider cultural applicability (Rowe, Duta, & Demeyere et al., 2021). The fieldworkers instructed the participants to look at a series of shapes (circles and triangles) in varying levels of grey that appeared for 500ms then disappeared with a blank screen between each shape appearing for

1500ms. They then instructed the participants to tap a grey bar on the bottom of the screen with a finger only when a circle (go condition) appears and not when other shapes appear on the screen (no-go condition). The fieldworkers administered one short practice session with six trials to acclimate the participants to the task. The main assessment section consisted of 36 trials, and a reminder was displayed to tap the grey bar when circles appeared between the practice session and the assessment session. We assessed EF using d-prime, a signal sensitivity score that accounts for both go accuracy and no-go inaccuracy and can be expressed using the following formula: d' = z(accuracy on go trials) – z(commission error rate). z indicates the normalisation of the accuracy and commission error rate.

Trails. This task was developed to measure shifting. The fieldworkers instructed the participants to use a finger to draw lines to connect the shapes displayed on the screen. The first baseline section instructed the participants to connect dark grey circles in the order of small to large, with an increasing number of circles and distractor squares of varying sizes. In the second baseline section, the participants were instructed to connect light grey squares from large to small with an increasing number of squares and distractor circles. The third testing section instructed participants to alternate between squares and circles, with the squares in the order from large to small and the circles in the order from small to large with an increase in the number of circles and squares after each trial (i.e. largest square to smallest circle, second largest square to second smallest circle, … etc.). We assessed inhibitory control by comparing the difference in the accuracy proportion scores between the baseline conditions and the switching condition to measure accuracy decrease in the switching condition.

Rule Finding. This task was developed to measure shifting. The fieldworkers instructed the participants to observe three columns of squares and triangles on the screen with triangles on the centre column. Each square-triangle-square row repeated eight times in a grid with one shape all coloured dark grey. The fieldworkers instructed the participants to look at the red dot that appeared on a shape and to guess where the dot will appear next. They also informed the participants that the red dot will move according to a certain rule and that the previous position of the dot will be highlighted to help predict its next position. Using this information, the participants tapped on the shape where they predicted that the red dot will appear. The fieldworkers also informed the participants that the rules will change without notice. The task utilised five different rules, with four focused on the position of the shape while the last rule focused on both the position and colour of the shape. The red dot switched positions with a total of 51 moves, out of which 45 moves were assessed. We used the accuracy proportion score, or the number of correct guesses divided by the total number of moves, to measure shifting.

Selection. This task was developed to measure visuospatial working memory. The fieldworkers instructed the participants to tap on all the fruits within a series of fruit and vegetable drawings on the screen by tapping on them with a finger. Local researchers confirmed that the fruits and vegetables in this task are appropriate for use in Kenya, Peru, and South Africa. In the first visible trial, a box appeared and remained around the fruits that the participants selected on the screen. A practice session was administered with drawings of an apple, carrot, bell pepper, banana, cabbage, and pear. The testing session then displayed 10 drawings of each of these fruits and vegetables. In the second invisible trial, the distribution of the fruits and vegetables changed, and a box only appeared briefly then disappeared when the fruits were selected on the screen. A practice session was administered, then the testing session was conducted. The invisible condition accuracy (dividing the correctly selected fruits by the total number of fruits) was used to measure updating, with the repeated selection of the same fruit penalised.

RACER

RACER contained two tasks, Spatial Delayed Match to Sample and the Simon Task. These tasks were completed using finger input.

Spatial Delayed Match to Sample. This task was developed to measure visuospatial working memory. The participants were shown a white screen on the tablet with either one (low-load) to three (high-load) black dots appearing on one of nine sections each on an invisible grid. The fieldworkers instructed the participants to memorise the dot locations while they were on the screen for 2000ms, then a blank screen appeared for either 100ms (short condition) or 3000ms (long condition). After the blank screen, the fieldworkers asked them to tap on the locations that the dots appeared previously. Accuracy was measured by distance from the tapped location to the centre of the nearest dot, by pixels. 39 trials were conducted, with 11 test trials and two practice trials for each load condition. We measured updating using dot location recall accuracy difference, which was calculated by subtracting the mean accuracy (pixel distance) for high-load conditions from the mean accuracy from low-load conditions.

Simon Task. This task was developed to measure inhibitory control. A solid pink or striped yellow and black dot appeared on the right or left side of the tablet screen with a small cross on the opposite side of the dot. The fieldworkers instructed the participants to touch the centre of the dot if the dots were pink (congruent) and to touch the cross if the dot was yellow and black (incongruent). 30 congruent and 30 incongruent trials were conducted, with seven practice trials. We calculated the mean inhibitory control score by subtracting the mean incongruent trial reaction time from the mean congruent trial reaction time. We also used accuracy in addition to the reaction time difference score, however, we observed extreme ceiling effects and retained the inhibitory control score as our composite measure.

SNAP-IV

The SNAP-IV Teacher and Parent Rating Scale is a 26-item ADHD screening tool and was administered in interview format by fieldworkers to the primary caregiver of the participant to assess three domains of inattention, hyperactivity/impulsivity, and opposition/defiance (Swanson et al., 2001). Although not yet validated in Peru, the SNAP-IV has been validated in children and adolescents from LMICs such as Brazil and Argentina (Costa et al., 2019; Grañana et al., 2011).

For the purposes of criterion validity, we used all subscales to establish OCS-EF and RACER's association to ADHD. The primary caregivers were asked whether they have seen a specific behaviour in the participant in the last two weeks and answered on a four-point Likert scale: "Not at all", "Just a little", "Quite a bit", or "Very much", with values ranging from 0 to 3. The fieldworkers repeated each statement a maximum of two times if the primary caregiver had difficulty answering, and items were marked as "Not known," "Refused to answer," or "NA" if an answer could not be obtained. We calculated the means for each subscale to evaluate the severity of ADHD symptomology in accordance with the original scoring guidelines (Swanson et al., 2001). The translated SNAP-IV measure is listed in Appendix B.

Results

We conducted all data analyses in R 4.0.2 (R Core Team, 2020) with the tidyverse (Wickham et al., 2019), lavaan (Rosseel, 2012), mice (Van Buuren & Groothuis-Oudshoorn, 2011), psych (Revelle, 2021), and lme4 (Bates, Mächler, & Bolker et al., 2015) packages.

Missing Sample Data

Of the 223 recruited for the study, we removed one participant for not taking the OCS-EF assessment. Furthermore, 72 participants (32.23%) had invalid or missing data in one or more of the OCS-EF and RACER tasks due to technology issues, interruptions, and other problems (see Table 2). All tasks had valid data greater than 84%. We implemented multivariate imputation by chained equations, a type of Markov chain Monte Carlo method, per recommended imputation guidelines since less than 20% of the data for each variable was missing at random (e.g. Graham, 2009; Graham et al., 2013; Yoo, 2009). We used the mice R package to impute the missing data (Van Buuren & Groothuis-Oudshoorn, 2011). We then analysed the dataset for multivariate outliers using Mahalanobis distance and removed five participants for significant Chi squares, indicating an extreme distance from the central mean ($p < .001$, $df = 8$). We utilised the data from 217 participants for analyses.

Table 2

Invalid Data Characteristics

Task	Issue	% of Task Data	n
Digit Recall	Interruptions	8.1	18
	Other	0.45	1
	Total	8.56	19
Spatial Span	Interruptions	6.76	15
	Total	6.76	15
Go/No-go	Interruptions	3.15	7
	Visual Problems	0.45	1
	Other	0.45	1
	NA	0.45	1
	Total	4.5	10
Trails	Interruptions	3.6	8
	Technical Problems	0.45	1
	Other	1.35	3
	NA	0.9	2
	Total	6.31	14
Rule-Finding	Interruptions	2.7	6
	Fatigue	0.45	1
	NA	0.45	1
	Total	3.6	8
Selection	Interruptions	3.6	8
	Other	0.45	1
	NA	0.45	1
	Total	4.5	10
Spatial Delayed Match to Sample	Missing Tablet Data	15.32	34
	Total	15.32	34
Simon Task	Missing Tablet Data	15.77	35
	Total	15.77	35

Analysis Samples & Task Performance

Factor Analyses Sample

We utilised the complete data from both the adolescent and young adult samples for our factor analyses in the interest of statistical power. Factor analyses can require a large sample size for sufficient power and participant numbers from anywhere between 40-450 are recommended for analyses depending on indicators, the magnitude of factor loadings, and factor correlations (Wolf et al., 2013). We chose to maximise the number of participants for the analyses since the OCS-EF and RACER contained eight total tasks with small correlations. However, to accurately represent the adolescent population in accordance to age specifications presented by the World Health Organization (2020), we conducted a multivariate analysis of variance (MANOVA) to examine any significant differences between the adolescent group (ages 14-19, *M*(SD) = 16.39(1.83), n = 163) and the young adult group (Ages 20-22, *M*(SD) = 20.91(0.81), n = 53) in cognitive task performance. Backward digit span, backward spatial span, d-prime, normalised switch accuracy cost, rule accuracy, invisible selection accuracy, spatial accuracy difference, and inhibitory control scores were entered as dependent variables. The resulting Pillai's Trace was significant (0.07, $F(8, 207) = 2.01$, $p = 0.047$), indicating that the adolescent group significantly differed in cognitive task performance compared to the young adult group. However, follow-up Welch two-sample *t*-tests and Wilcoxon Rank Sum tests (for tasks with non-normal distributions) with a Bonferroni-corrected alpha for multiple comparisons ($a = 0.006$) indicated that none of the tasks significantly differed between the two groups (see Table 3).

Table 3

Welch t-tests and Wilcoxon Rank Sum tests between the Adolescent Group and the Young Adult Group

Task	Indicator	Test	*df*	*t*/ *W*	*p*	95% CI
Digit Recall	Backward Span	Welch	86.43	0.10	0.92	-0.28, 0.30
Spatial Span	Backward Span	Welch	75.99	2.67	0.009	0.16, 1.12
Go/No-Go	D-Prime	Wilcoxon Rank Sum		3833	0.18	
Trails	Normalised Switch Accuracy Cost	Wilcoxon Rank Sum		5045.5	0.07	
Rule Finding	Accuracy	Welch	90.30	0.66	0.51	-0.04, 0.07
Selection	Accuracy	Wilcoxon Rank Sum		3964.5	0.37	
Spatial Delayed Match to Sample	Accuracy Difference	Welch	81.06	-0.11	0.92	-20.61, 18.52
Simon Task	Inhibitory Control Score	Welch	79.57	0.25	0.81	-19.34, 24.81

Note. Bonferroni-corrected *p*-value ($a' = \frac{a}{n}$) = 0.006.

Linear Mixed Model Sample

To establish criterion validity with the ADHD subscores, we used the participant data of those who underwent SNAP-IV (ages 14-17, n =110). Since the sample ages are in line with adolescent age specifications, we did not conduct age-group comparisons.

Task Performance

The cognitive task and SNAP-IV performance data for the complete sample, the adolescent sample, and the young adult sample are listed in Tables 4, 5, and 6. The Go/No-Go (d-prime) and Selection (accuracy) task indicators showed extreme negative skew with a ceiling effect. Furthermore, the SNAP-IV hyperactivity/impulsivity subscales saw a floor effect with a slight positive skew. We applied data transformation methods for remediation, however, the distributions retained the significant skew. We used the non-transformed data with robust estimators and nonparametric analyses to account for the nonnormal distributions.

Table 4

Complete Sample Task Performance

Task	Indicator	N	Mean (SD)	Range (min, max)	Distribution Shape
Digit Recall	Backward Span	217	4.16 (0.91)	2, 7	Normal
Spatial Span	Backward Span	217	5.71 (1.4)	0, 8	Normal
Go/No-Go	D-Prime	217	3.37 (0.68)	-0.33, 3.77	Negatively Skewed
Trails	Normalised Switch Accuracy Cost	217	0.51 (0.1)	0, 1	Normal
Rule Finding	Accuracy	217	0.47 (0.17)	0.02, 0.89	Normal
Selection	Accuracy	217	0.85 (0.17)	-0.8, 1	Negatively Skewed
Spatial Delayed Match to Sample	Accuracy Difference	217	-233.94 (58.60)	-418.13, -88.18	Normal
Simon Task	Inhibitory Control Score	217	6.06 (65.26)	-206.41, 196.58	Normal
SNAP-IV	Inattention Subscore	110	0.95 (0.57)	0, 2.44	Normal
	Hyperactivity/ Impulsivity	110	0.7 (0.53)	0, 2.44	Positively Skewed
	Opposition/ Defiance	110	0.88 (0.66)	0, 2.75	Normal

Note. N = 217.

Table 5

Adolescent Sample Task Performance

Task	Indicator	N	Mean (SD)	Range (min, max)	Distribution Shape
Digit Recall	Backward Span	163	4.17 (0.9)	2, 6	Normal
Spatial Span	Backward Span	163	5.87 (1.3)	2, 8	Normal
Go/No-Go	D-Prime	163	3.35 (0.68)	-0.33, 3.77	Negatively Skewed
Trails	Normalised Switch Accuracy Cost	163	0.52 (0.1)	0.29, 1	Normal
Rule Finding	Accuracy	163	0.47 (0.17)	0.02, 0.89	Normal
Selection	Accuracy	163	0.84 (0.19)	-0.8, 1	Negatively Skewed
Spatial Delayed Match to Sample	Accuracy Difference	163	-234.12 (57.08)	-418.13, -88.18	Normal
Simon Task	Inhibitory Control Score	163	6.68 (63.25)	-166.68, 196.58	Normal
SNAP-IV	Inattention Subscore	110	0.95 (0.57)	0, 2.44	Normal
	Hyperactivity/ Impulsivity	110	0.7 (0.53)	0, 2.44	Positively Skewed
	Opposition/ Defiance	110	0.88 (0.66)	0, 2.75	Normal

Note. Complete adolescent sample (ages 14-19) is n = 163, while the SNAP-IV sample (ages 14-17) is n = 110.

Table 6

Young Adult Sample Task Performance

Task	Indicator	N	Mean (SD)	Range (min, max)	Distribution Shape
Digit Recall	Backward Span	53	4.15 (0.93)	2, 7	Normal
Spatial Span	Backward Span	53	5.23 (1.59)	0, 8	Normal
Go/No-Go	D-Prime	53	3.43 (0.67)	0.31, 3.77	Negatively Skewed
Trails	Normalised Switch Accuracy Cost	53	0.49 (0.11)	0, 0.68	Negatively Skewed
Rule Finding	Accuracy	53	0.46 (0.17)	0.13, 0.87	Normal
Selection	Accuracy	53	0.87 (0.12)	0.3, 1	Negatively Skewed
Spatial Delayed Match to Sample	Accuracy Difference	53	-233.07 (63.71)	-367.13, -103.12	Normal
Simon Task	Inhibitory Control Score	53	3.94 (72.24)	-206.41, 135.73	Normal

Note. Young adult sample (ages 20-22) is n = 53.

Construct Validity via Confirmatory Factor Analyses

Confirmatory factor analyses were conducted using the lavaan package in R (Rosseel, 2012). The following models were compared according to our hypotheses:

1. Miyake and Friedman's (2012) correlated three-factor model of updating, shifting, and inhibition:
    i. Updating: Digit Recall, Spatial Span, and Selection tasks from OCS-EF and the Spatial Delayed Match to Sample task from RACER
    ii. Shifting: Trails and Rule-finding tasks from OCS-EF
    iii. Inhibitory control: Go/no-go task from OCS-EF and the Simon task from RACER
2. Unitary EF with all OCS-EF and RACER tasks loading onto a single factor.

We used maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic to account for non-normally distributed data (Satorra & Bentler, 1994). Latent factors were correlated, their variances were fixed to one, and all parameters were standardised. The parameter estimates of the three-factor and one-factor models are found in Table 7 (three-factor model) and Table 8 (one-factor model). The model, standardised factor loadings, and residuals are listed in Figure 3 (three-factor model) and Figure 4 (one-factor model).

Global goodness-of-fit indices are listed in Table 9. The robust root mean square error of approximation (RMSEA), robust comparative fit index (CFI), Akaike information criterion (AIC), and sample-size adjusted Bayesian information criterion (BIC) all indicated better fit in the one-factor model. In addition, the three-factor model had a significant Chi-squared test between the identified model and the saturated model ($\chi^2(17) = 30.85$, $p = .02$), indicating a significant difference between the observed and expected covariances and a worse fit. We retained the one-factor model and calculated the composite reliability to examine the internal consistency of OCS-EF and RACER. We calculated the variance accounted for by the cognitive tasks compared to the total variance (Geldhof et al., 2014). The results indicated that the cognitive tasks explained 25% of the total variance in the unitary EF factor (ω = 0.50), which was below the acceptable value for internal consistency (ω > .60).

Table 7

Three-Factor Model Parameter Estimates

Loading	Coefficient	SE	z	p
IN - Go/No-Go	0.06	0.15	0.29	0.77
IN - Simon Task	-0.70	145.13	-0.32	0.75
SH - Trails	0.15	0.01	1.10	0.27
SH - Rule Finding	0.34	0.05	1.25	0.21
UP - Digit Recall	0.35	0.08	4.10	<.001
UP - Spatial Span	0.69	0.14	7.06	<.001
UP - Selection	0.31	0.01	5.20	<.001
UP - Dot Task	0.36	5.43	3.91	<.001
Cov(IN, SH)	0.13	0.50	0.26	0.79
Cov(IN, UP)	-0.13	0.42	-0.31	0.76
Cov(SH, UP)	1.48	1.16	1.28	0.20

Note. IN = inhibition, SH = shifting, UP = updating. Cov = covariance. Dot Task = Spatial Delayed Match to Sample.

Table 8

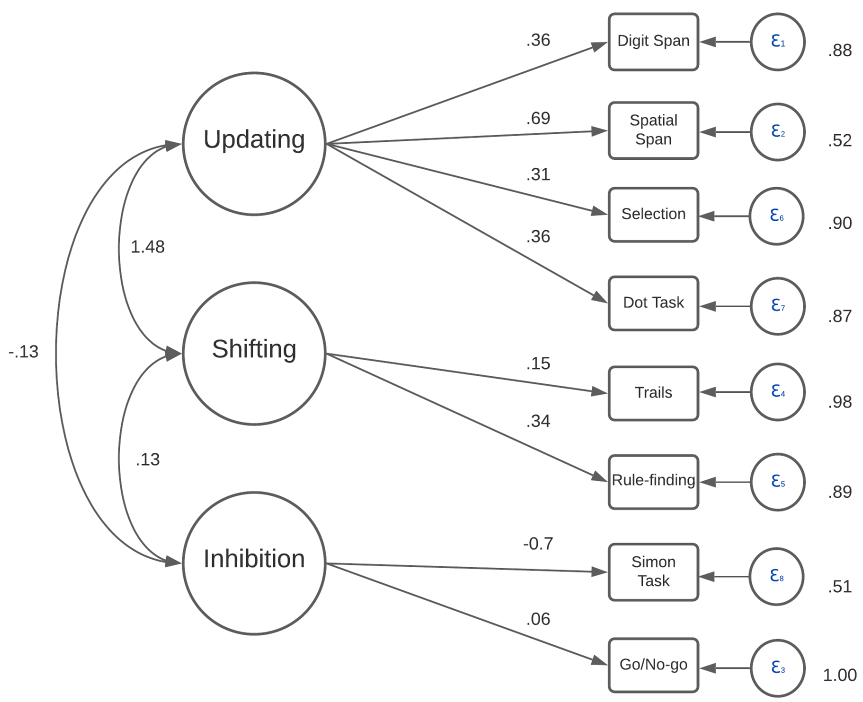
One-Factor Model Parameter Estimates

Loading	Coefficient	SE	z	p
EF - Go/No-Go	0.13	0.07	1.34	0.18
EF - Simon Task	0.06	6.15	0.67	0.50
EF - Trails	0.19	0.01	2.33	0.02
EF - Rule Finding	0.49	0.01	6.05	<.001
EF - Digit Recall	0.36	0.08	4.20	<.001
EF - Spatial Span	0.69	0.14	7.14	<.001
EF - Selection	0.32	0.01	5.24	<.001
EF - Dot Task	0.36	5.47	3.88	<.001

Note. EF = executive function. Dot Task = Spatial Delayed Match to Sample.

Figure 3

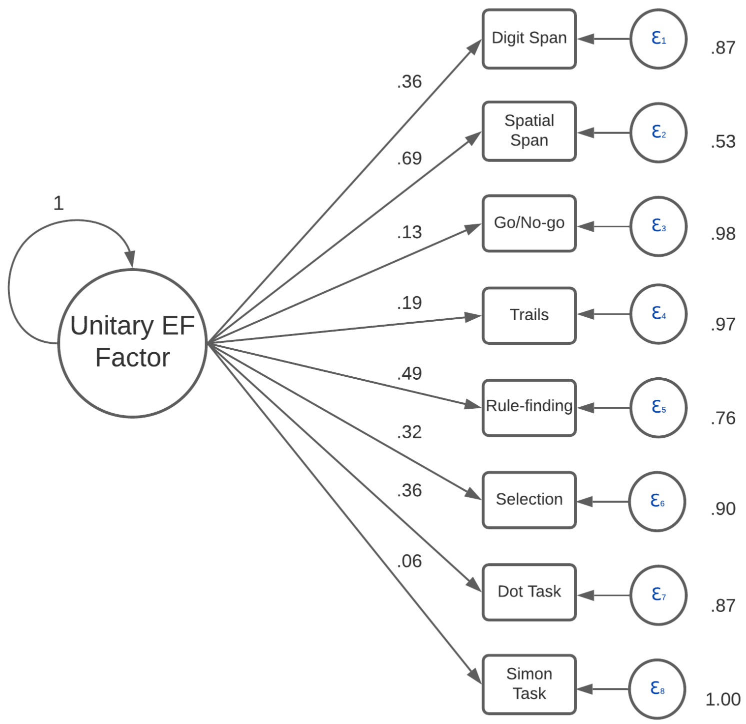
Three-Factor Model Diagram



Note. Dot Task = Spatial Delayed Match to Sample.

Figure 4

One-Factor Model Diagram



Note. Dot Task = Spatial Delayed Match to Sample.

Table 9

Global Goodness-of-fit Indices

Model	χ^2	*df*	*p*	RMSEA	CFI	AIC	BIC
Three-factor	30.85	17	0.02	0.06	0.83	5873.37	5877.38
One-factor	31.78	20	0.05	0.05	0.86	5867.56	5870.93

Note. χ^2 = robust χ^2 test. *df* and *p*-value are for the χ^2 test. RMSEA = robust root mean square error of approximation, CFI = robust comparative fit index, AIC = Akaike information criterion, BIC = sample-size adjusted Bayesian information criterion.

Criterion Validity via Linear Mixed Models

We selected linear mixed model analyses to predict SNAP-IV scores while accounting for the random effects introduced through extraneous variables. These models are extensions of linear regression but allow the specification of random effects, or non-manipulated conditions, aside from a non-specified error term by calculating the intraclass correlation (Field, Miles, & Field, 2012). Furthermore, the random factors can be included in one model without multiple analyses, which prevents the inflation of potential Type I error resulting from multiple regression analyses (Field, Miles, & Field, 2012). The model can account for differences in regression slopes across the categorical variables, the assumption of independence does not need to be met, and normal data distributions are not required, provided that the residuals are normally distributed (Field, Miles, & Field, 2012).

We conducted the linear mixed model analyses in R with the lme4 package (Bates, Kliegl, & Vasishth et al., 2015) to examine the association between the OCS-EF and RACER tasks and the SNAP-IV subscales of inattention, hyperactivity/impulsivity, and opposition/defiance. We used the task composite scores as predictors over unitary-EF factor loadings from the CFA due to the low composite reliability. Backward digit span (Digit Recall), backward spatial span (Spatial Span), d-prime (Go/No-go), switch accuracy cost (Trails), rule accuracy (Rule Finding), selection accuracy (Selection), spatial accuracy (Spatial Delayed Match to Sample), and inhibitory control score (Simon Task) were entered as fixed effects to predict inattention, hyperactivity/impulsivity, and opposition/defiance, respectively. Age was accounted for as a fixed effect, and sex was accounted for as a random effect both by-intercept and slope on all eight tasks. We selected this maximal model to account for the effect of age and sex on the predictability of the SNAP-IV subscores while controlling for Type I errors (Barr et al., 2013). Preliminary correlation coefficients for all tasks and subscores are listed in Table 10.

Table 10

Correlation Coefficients for OCS-EF & RACER Tasks, & SNAP-IV Subscores

	Digit Recall	Spatial Span	Go/No-go	Trails	Rule Finding	Selection	Dot Task	Simon Task	SNAP Inattention	SNAP Hyperactivity	SNAP Opposition
Digit Recall	1.00	0.34	0.12	0.22	0.02	0.21	0.15	0.09	-0.08	-0.06	-0.02
Spatial Span	0.34	1.00	0.23	0.09	0.34	0.11	0.25	0.17	-0.20	-0.14	-0.13
Go/No-go	0.12	0.23	1.00	0.11	0.17	0.09	0.00	-0.12	-0.25	-0.22	-0.06
Trails	0.22	0.09	0.11	1.00	0.10	0.04	0.02	0.10	-0.06	-0.02	0.12
Rule Finding	0.02	0.34	0.17	0.10	1.00	0.10	0.21	-0.01	-0.12	-0.02	0.00
Selection	0.21	0.11	0.09	0.04	0.10	1.00	0.28	-0.03	0.00	0.12	0.06
Dot Task	0.15	0.25	0.00	0.02	0.21	0.28	1.00	0.18	-0.05	0.13	-0.03
Simon Task	0.09	0.17	-0.12	0.10	-0.01	-0.03	0.18	1.00	0.08	0.12	0.06
SNAP Inattention	-0.08	-0.20	-0.25	-0.06	-0.12	0.00	-0.05	0.08	1.00	0.69	0.57
SNAP Hyperactivity	-0.06	-0.14	-0.22	-0.02	-0.02	0.12	0.13	0.12	0.69	1.00	0.60
SNAP Opposition	-0.02	-0.13	-0.06	0.12	0.00	0.06	-0.03	0.06	0.57	0.60	1.00

Note. Dot Task = Spatial Delayed Match to Sample.

Inattention Subscore

The inattention subscore model did not converge due to overparametrisation and we fitted a zero-correlation model as the first step of model remediation (Bates, Kliegl, & Vasishth et al., 2015). However, a follow-up primary components analysis (PCA) indicated that the model was still overparametrised, with the random slopes of the tasks aside from the Go/No-go and Trails task accounting for less than .001% of the model variance. Reducing the model by removing the extremely low-variance parameters increases the power of the model while still accounting for Type I errors (Bates, Kliegl, & Vasishth et al., 2015). We removed the random slopes for the minimal variance and fit the mixed model once more. The second PCA for this fitted model indicated that the Go/No-go and Trails random slopes accounted for 0% of the variance, and they were removed. The final model included the random intercepts for sex and the fixed factors for all eight tasks and age. The model reported a better fit compared to the maximal model (maximal BIC: 271.14; final model BIC: 229.65). None of the OCS-EF and RACER tasks, accounting for sex, significantly predicted the SNAP-IV inattention subscores. However, the Go/No-go Task showed the largest beta coefficient close to significance (β (*SE*) = -0.11 (0.06), t (99) = -1.95, p = .054), indicating that for one standard deviation increase in the task score, the inattention subscores decreased by -0.11 SD. The standardised Beta coefficients, standard errors, and *t*-test results can be found in Table 11. The standardised Beta coefficients are interpreted as regression coefficients, indicating that for one standard deviation increase in the predictor variable, the outcome variable changes by the coefficient.

Table 11

SNAP-IV Inattention Subscale Linear Mixed Model Results

Variable	β	*SE*	*df*	*t*	*p*
Intercept	0.95	0.06	0.79	16.02	0.07
Go/No-go	-0.11	0.06	99.00	-1.95	0.05
Spatial Span	-0.09	0.06	98.87	-1.46	0.15
Digit Recall	-0.01	0.06	98.71	-0.19	0.85
Selection	0.04	0.06	98.43	0.64	0.52
Dot Task	-0.02	0.06	98.81	-0.33	0.74
Rule Finding	-0.02	0.06	86.65	-0.26	0.79
Trails	-0.01	0.06	98.99	-0.22	0.83
Simon Task	0.05	0.06	98.53	0.97	0.33
Age	0.02	0.06	98.89	0.43	0.67

Note. β = standardised beta coefficient. *SE* = standard error. Dot Task = Spatial Delayed Match to Sample.

Hyperactivity/Impulsivity Subscore

The hyperactivity/impulsivity subscore maximal and zero-correlation models also did not converge due to overparametrisation. After a PCA, we removed the parameters accounting for near-zero variance and retained the following by-sex random slopes: Selection, Go/No-go, Rule Finding, and Trails. The model fit still indicated overparametrisation, and we retained the by-sex intercept after the follow-up PCA indicated that it accounted for 100% of the random effect variance. The final model included the random intercepts for sex and the fixed factors for all eight tasks and age and reported a better model fit compared to the maximal model (maximal BIC: 422.14; final model BIC: 236.07). None of the OCS-EF and RACER tasks, accounting for sex, significantly predicted the SNAP-IV hyperactivity/impulsivity subscores. The standardised Beta coefficients, standard errors, and *t*-test results can be found in Table 12.

Table 12

SNAP-IV Hyperactivity/Impulsivity Subscale Linear Mixed Model Results

Variable	β	*SE*	*df*	*t*	*p*
Intercept	0.70	0.05	99	13.98	< 0.001
Go/No-go	-0.10	0.05	99	-1.81	0.07
Spatial Span	-0.08	0.06	99	-1.42	0.16
Digit Recall	-0.02	0.06	99	-0.43	0.67
Selection	0.07	0.05	99	1.37	0.17
Dot Task	0.06	0.06	99	1.08	0.28
Rule Finding	0.02	0.06	99	0.31	0.76
Trails	0.00	0.05	99	0.06	0.95
Simon Task	0.06	0.05	99	1.19	0.24
Age	0.04	0.05	99	0.69	0.49

Note. β = standardised beta coefficient. *SE* = standard error. Dot Task = Spatial Delayed Match to Sample.

Opposition/Defiance Subscore

Finally, we fitted the opposition/defiance subscore maximal model. The model, along with its zero-correlation counterpart, did not converge. We retained the Rule Finding and Trails by-sex random slopes after a PCA. We further removed the parameters accounting for zero variance and retained the by-sex intercept after the follow-up PCA. The final model included the random intercepts for sex and the fixed factors for all eight tasks and age and reported a better model fit compared to the maximal model (maximal BIC: 484.78; final model BIC: 291.40). Similar to the other two subscores, none of the OCS-EF and RACER tasks, accounting for sex, significantly predicted the SNAP-IV opposition/defiance subscores. The standardised Beta coefficients, standard errors, and *t*-test results can be found in Table 13.

In summary, the linear mixed models revealed that the OCS-EF and RACER tasks did not significantly predict inattention, hyperactivity/impulsivity, or opposition/defiance. Furthermore, age did not significantly predict SNAP-IV scores and the by-sex random intercepts accounted for less than .04 % of the random variance in all three SNAP-IV subscores (inattention: .04%; hyperactivity/impulsivity: < .001%; opposition/defiance: < .001%) indicating that the SNAP-IV score variance between male and female participants was minimal.

Table 13

SNAP-IV Opposition/Defiance Subscale Linear Mixed Model Results

Variable	β	*SE*	*df*	*t*	*p*
Intercept	0.88	0.06	99	13.69	<0.001
Go/No-go	-0.05	0.07	99	-0.66	0.51
Spatial Span	-0.10	0.08	99	-1.35	0.18
Digit Recall	-0.02	0.07	99	-0.25	0.80
Selection	0.07	0.07	99	1.00	0.32
Dot Task	-0.03	0.07	99	-0.45	0.65
Rule Finding	0.05	0.07	99	0.66	0.51
Trails	0.08	0.07	99	1.26	0.21
Simon Task	0.06	0.07	99	0.85	0.40
Age	0.10	0.07	99	1.43	0.16

Note. β = standardised beta coefficient. *SE* = standard error. Dot Task = Spatial Delayed Match to Sample.

Exploratory Analyses

Confirmatory Factor Analyses with Adolescent-only Sample

We conducted CFA model comparisons with the adolescent-only sample to examine whether the model fit aligns with our findings in the complete sample. We conducted CFA with the three-factor model and the one-factor model, also with maximum likelihood estimation and a Satorra-Bentler scaled test statistic. As expected, the lower participant count in the adolescent sample underpowered the three-factor model analysis and the model did not converge. However, we found that the one-factor model did converge with similar factor loadings. The model, standardised factor loadings, and residuals are listed in Appendix C.

Exploratory Factor Analyses

We conducted planned exploratory factor analyses to examine the factor structure of OCS-EF and RACER according to the dataset and to complement the results in the CFA. We tested the three-factor and two-factor exploratory analyses using the lavaan package in R (Rosseel, 2012). We utilised maximum likelihood estimation with robust standard errors, a Satorra-Bentler scaled test statistic, and Oblimin rotation for correlated latent factors. All parameters were standardised, latent factors were correlated, and their variances were fixed to one. The three-factor model was not identified. The two-factor model was identified and reported a better fit compared the CFA one-factor model on all fit indices ($\chi^2(13) = 12.31, p = .50$; robust RMSEA = 0; robust CFI =1; AIC = 5860.73; sample-size adjusted BIC = 5865.58). The parameter estimates for each factor, the standard errors, *z*-scores, and *p*-values are listed in Table 14, and the model, standardised factor loadings, and residuals are listed in Figure 5. Composite reliability was calculated for factor 1 ($\omega = 0.45$) and factor 2 ($\omega = 0.76$), with the Simon Task, Trails, Rule Finding, Spatial Span, Selection, and Spatial Delayed Match to Sample accounting for 20.25% of the factor 1 variance, and Go/No-Go and Digit Recall accounting for 57.76% of the factor 2 variance. Tasks for factor 2 showed acceptable reliability, however, tasks for factor 1 were below the acceptable threshold ($\omega >$ .60).

Table 14

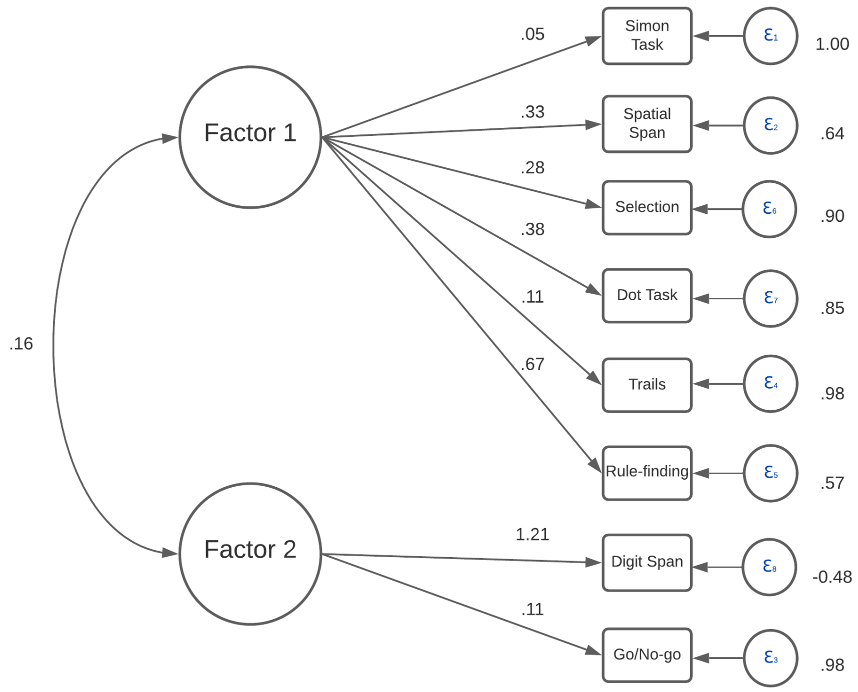
EFA Two-Factor Model Parameter Estimates

Loading	Coefficient	SE	z	p
Factor 1 - Simon Task	0.05	6.05	0.49	0.63
Factor 1 - Trails	0.11	0.01	0.90	0.37
Factor 1 - Rule Finding	0.67	0.02	4.74	<.001
Factor 1 - Spatial Span	0.33	0.19	4.00	<.001
Factor 1 - Selection	0.28	0.02	3.21	.001
Factor 1 - Dot Task	0.38	5.78	3.87	<.001
Factor 2 - Go/No-Go	0.11	0.08	0.91	0.36
Factor 2 - Digit Recall	1.21	0.91	1.21	0.23

Note. Dot Task = Spatial Delayed Match to Sample.

Figure 5

EFA Two-Factor Model Diagram



Note. Dot Task = Spatial Delayed Match to Sample.

Discussion

Findings

The current study aimed to validate OCS-EF and RACER as an easily administrable, unified, and valid measure of EF in Peruvian adolescents. However, the results were mixed with some evidence of administration feasibility and construct validity. OCS-EF and RACER combined showed feasibility in administration in LMIC contexts, with fieldworkers reporting that over 90% of participant conditions raised “no issue” within a home-based assessment environment for all tasks. Furthermore, the observed variability in Digit Recall, Spatial Span, Trails, Rule Finding, Simon Task, and Spatial Delayed Match to Sample tasks showed that OCS-EF and RACER can identify individual performance differences among non-clinically diagnosed participants, though many of the tasks were originally designed for use in a clinical population (Demeyere et al., 2021). The construct validity results were complex with some convergence with previous LMIC literature, while issues were raised for further research and adaptation.

Examining construct validity through CFA model comparisons revealed that the unitary EF model best fit the data, which was in line with a previous finding in LMIC adults in Guatemala, the Philippines, and South Africa (Wray et al., 2020). However, the unitary model fit differed from our hypothesis and the three-factor model fit previously found in OCS-EF tasks among adolescent and young adult women in South Africa (Rowe, Duta, & Demeyere et al., 2021). Furthermore, additional findings questioned the construct validity of OCS-EF and RACER. The calculated composite reliability for the CFA of the one-factor model, which fit the data best, explained only 25% of the variance in unitary EF. This indicates that the OCS-EF and RACER tasks do not explain a large proportion of the latent factor variance.

We further conducted EFA to complement the confirmatory approach with the optimal data-driven model. The results revealed that the cognitive tasks loaded onto a two-factor model with six tasks loading onto one factor and the Digit Span task and the Go/No-Go task loading onto another factor. However, composite reliability once again showed that the six tasks explained a low proportion of the total variance in the factor (20.25%). Furthermore, although the second factor's two loadings explained a sufficient percentage of the total variance (57.76%), the Digit Span task showed an abnormally high standardised loading (1.21) which indicated that the single task primarily determined the identification of the factor.

Criterion validity through linear mixed effect modelling also raised potential issues with OCS-EF and RACER. The results indicated that none of the cognitive tasks significantly predicted the SNAP-IV subscales, and the correlations between the tasks and the subscales confirmed that all the tasks had small correlation coefficients (< 0.25). Furthermore, the Selection and Simon tasks correlated positively with all three SNAP-IV subscales, while the Spatial Delayed Match to Sample task correlated positively with the hyperactivity/impulsivity subscale. These findings reveal that the OCS-EF and RACER tasks did not predict ADHD scores as previously found, and some tasks predicted the opposite association from what has been seen in EF tasks (Krieger & Amador-Campos, 2017; Martel et al., 2007).

These issues with construct and criterion validity can be due to multiple reasons. Firstly, the Go/No-go and Selection tasks both reported ceiling effects with extreme skew and little variability among the majority of the participants. Although robust estimators and tests were used to account for the skew, the lack of variability can affect correlations that determine both factor loadings and the linear mixed models (Arnau et al., 2012; Kowalski, 1972). Secondly, the relationship between ADHD and EF in LMICs have not yet been observed to our knowledge, and the cognitive tasks may not predict ADHD as expected in HIC contexts. Studies from HICs have shown negative associations between EF assessments and ADHD scores, however, predictors from HIC do not always have the same effects in LMIC contexts (Haft & Hoeft, 2017). Lastly, the Simon Task composite score had the opposite effect of what was expected. The Simon Task composite score compared incongruent (inhibition) conditions to congruent (neutral) conditions, with the expectation that the participants will perform better on congruent conditions compared to the incongruent conditions. However, the participants performed better, on average, on the incongruent conditions. This effect explains the positive correlation of the Simon Task on the SNAP-IV subscores. In conclusion, OCS-EF and RACER in their current condition do not effectively measure EF – the extreme skew combined with unexpected performance in the data contributed to a poor fit in the factor analyses and the linear mixed models, and ADHD measures may not be predicted by EF in LMICs.

Limitations

Several limitations were found in this study. One limitation was imposed by secondary data analysis – since the data were collected before this study, the participants, procedure, and assessments were determined and could not be retroactively changed. This limited the number of adolescent participants with sufficient power for the factor analyses. Another limitation was the adaptation of OCS-EF and RACER for administration feasibility across LMICs. The Iowa Gambling and Figure Copy tasks were removed due to participant fatigue and the Go/No-Go task was adapted to replace facial expressions with shapes for transferability across multiple LMIC countries. However, these adaptations limited the exploration of hot/cool EF factors in addition to the unitary and three-factor models. An additional limitation was available, validated external measures for criterion validity. Although SNAP-IV has been previously validated in LMICs, the association between ADHD and EF is yet to be examined in that context (Costa et al., 2019; Grañana et al., 2011). Criterion validity is difficult to examine without establishing the association within LMIC contexts. Finally, the task order effect and the length of the cognitive tasks may have affected the performance of the participants.

Recommendations & Future Research

Based on these results, we recommend adaptations of the current OCS-EF and RACER tasks for use in LMIC adolescents. Firstly, we recommend replacing, removing, or increasing the difficulty in some of the tasks. The eight administered tasks currently do not represent the "hot" EF factors in the hot/cool EF model and mostly measures the updating factor in the three-factor model (50% of total tasks). In addition, the Iowa Gambling and Figure Copy tasks were removed for participant fatigue, and the current assessment should not be expanded further. Furthermore, the current combined measure incorporates tasks normally utilised in children or clinical populations, which may have contributed to the ceiling effects in the adolescent sample. While the majority of the tasks reported sufficient variability and normal distributions, the adolescent participants performed at ceiling for the Go/No-go and Selection tasks and performed better in the inhibition condition for the Simon task. To better examine the variability in this adolescent sample while not adding to the total number of administered tasks, the Iowa Gambling task could replace the Simon or the Go/No-Go task to account for both "hot" and inhibition factors. The Selection task could be removed as there are three other tasks also accounting for the updating factor. Additionally, both the Go/No-go and Selections tasks could have more distractor items to raise the difficulty of these tasks (e.g. adding different shapes and more vegetables).

We also recommend further validation in other LMIC contexts. While our study found mixed results in this sample, previous validation in South African adolescent and young-adult women found that OCS-EF is feasible and valid to use (Rowe, Duta, & Demeyere et al., 2021). Furthermore, OCS-EF and RACER combined have also been used in South African HIV-positive adolescents (Rowe, Pozuelo, & Nickless et al., 2021). The OCS-EF and RACER may not currently be widely applicable in all LMIC contexts but may be appropriate for use in some countries as opposed to others.

In conclusion, we found mixed results for the construct and criterion validity of the OCS-EF and RACER among adolescents in Peru. While the measure is currently feasibly administered, issues such as non-conforming tasks were identified with potential adaptations for the improvement of the unified measure. Currently, widely validated EF measures for adolescents in LMICs are lacking, preventing further exploration into the role of EF and the development of potential interventions for this age group and context. Additional research examining the validity of EF measures are necessary for an accurate understanding of EF in LMICs. Adaptation and further implementation of the OCS-EF and RACER in other countries are recommended to establish a valid measure of EF among LMIC adolescents for future research and EF intervention development.

References

Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology, 111*(3), 446–458. https://doi.org/10.1037/edu0000296

Arnau, J., Bono, R., Blanca, M. J., & Bendayan, R. (2012). Using the linear mixed model to analyze nonnormal data distributions in longitudinal designs. *Behavior Research Methods*, *44*(4), 1224–1238. https://doi.org/10.3758/s13428-012-0196-y

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Retrieved from http://arxiv.org/abs/1506.04967

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

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). *Executive function.* In P. D. Zelazo (Ed.), *Oxford library of psychology. The Oxford handbook of developmental psychology (Vol. 1): Body and mind* (p. 706–743). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199958450.013.0025

Chen, A., Panter-Brick, C., Hadfield, K., Dajani, R., Hamoudi, A., & Sheridan, M. (2019). Minds under siege: Cognitive signatures of poverty and trauma in refugee and non-refugee adolescents. *Child Development*, *90*(6), 1856–1865. https://doi.org/10.1111/cdev.13320

Costa, D. S., de Paula, J. J., Malloy-Diniz, L. F., Romano-Silva, M. A., & Miranda, D. M. (2019). Parent SNAP-IV rating of attention-deficit/hyperactivity disorder: Accuracy in a clinical sample of ADHD, validity, and reliability in a Brazilian sample. *Jornal de Pediatria*, *95*(6), 736–743. https://doi.org/10.1016/j.jped.2018.06.014

Demeyere, N., Haupt, M., Webb, S.S., Strobel, L., Milosevich, E.T., Moore, M.J., Wright, H., Finke, K., & Duta., M. D. (2021). Introducing the tablet-based Oxford Cognitive Screen-Plus (OCS-Plus) as an assessment tool for subtle cognitive impairments. *Scientific Reports, 11*(1), 1-14. https://doi.org/10.1038/s41598-021-87287-8

Ferguson, J., Brunsdon, V., & Bradford, E. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports, 11*(1), 1-17. https://doi.org/10.1038/s41598-020-80866-1

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R.* Los Angeles, CA: Sage.

Ford, C. B., Kim, H. Y., Brown, L., Aber, J. L., & Sheridan, M. A. (2019). A cognitive assessment tool designed for data collection in the field in low- and middle-income countries. *Research in Comparative and International Education, 14*(1), 141–157. https://doi.org/10.1177/1745499919829217

Geldhof, J., Preacher, K., & Zyphur, M. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19* (1), 72-91. https://doi.org/10.1037/a0032138

Graham, J.W. (2009). Missing data analysis : Making it work in the real world. *Annual Review of Psychology, 60*(1), 549-576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W., Cumsille, P. E., & Shevock, A. E. (2013). Methods for handling missing data. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), *Handbook of psychology: Research methods in psychology* (pp. 109–141).

Grañana, N., Richaudeau, A., Gorriti, C. R., O'Flaherty, M., Scotti, M. E., Sixto, L., Allegri, R., & Fejerman, N. (2011). Assessment of attention deficit hyperactivity: SNAP-IV scale adapted to Argentina. *Revista Panamericana De Salud Pública, 29*(5), 344-349.

Haft, S. L., & Hoeft, F. (2017). 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

Hamoudi, A. & Sheridan, M. (2015). Unpacking the black box of cognitive ability: A novel tool for assessment in a population based survey. Unpublished manuscript.

Han, G., Helm, J., Iucha, C., Zahn-Waxler, C., Hastings, P., & Klimes-Dougan, B. (2016). Are executive functioning deficits concurrently and predictively associated with depressive and anxiety symptoms in adolescents? *Journal of Clinical Child and Adolescent Psychology, 45*(1), 44-58. https://doi.org/10.1080/15374416.2015.1041592

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

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*(1), 20-36. https://doi.org/10.1080/87565640903325691

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., Areshenkoff, C., Rast, P., Hofer, S., Iverson, G., & Garcia-Barrera, M. (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

Kleine Deters, R., Naaijen, J., Rosa, M., Aggensteiner, P. M., Banaschewski, T., Saam, M. C., Schulze, U. M. E., Sethi, A., Craig, M. C., Sagar-Ouriaghli, I., Santosh, P., Castro-Fornieles, J., Penzol, M. J., Arango, C., Werhahn, J. E., Brandeis, D., Franke, B., Glennon, J., Buitelaar, J. K., … & Dietrich, A. (2020). Executive functioning and emotion recognition in youth with oppositional defiant disorder and/or conduct disorder. *The World Journal of Biological Psychiatry*, *21*(7), 539–551. https://doi.org/10.1080/15622975.2020.1747114

Kowalski, C. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *21*(1), 1-12. https://doi.org/10.2307/2346598

Krieger, V., & Amador-Campos, J. A. (2018). Assessment of executive function in ADHD adolescents: Contribution of performance tests and rating scales. *Child Neuropsychology*, *24*(8), 1063–1087. https://doi.org/10.1080/09297049.2017.1386781

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

Martel, M., Nikolas, M., & Nigg, J. T. (2007). Executive function in adolescents with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(11), 1437–1444. https://doi.org/10.1097/chi.0b013e31814cf953

Matsumoto, D. (2003). Cross-cultural research. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology*. Oxford: Blackwell. https://doi.org/10.1002/9780470756973

McCoy, D., Peet, E., Ezzati, M., Danaei, G., Black, M., Sudfeld, C., Fawzi, W., & Fink, G. (2016). Early childhood developmental status in low- and middle-income countries: National, regional, and global prevalence estimates using predictive modeling. *PLoS Medicine, 13*(6), E1002034. https://doi.org/10.1371/journal.pmed.1002034

Menezes, A., Dias, M. N., Trevisan, B. T., Carreiro, L. R. R., & Seabra, A. G. (2015). Intervention for executive functions in attention deficit and hyperactivity disorder. *Arquivos De Neuro-psiquiatria, 73*(3), 227-236. https://doi.org/10.1590/0004-282X20140225

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. https://doi.org/10.1177/0963721411429458

Nyongesa, M. K., Ssewanyana, D., Mutua, A. M., Chongwo, E., Scerif, G., Newton, C. R. J. C., & Abubakar, A. (2019). Assessing executive function in adolescence: A scoping review of existing measures and their psychometric robustness. *Frontiers in Psychology*, *10*, 311. https://doi.org/10.3389/fpsyg.2019.00311

Pharo, H. , Sim, C., Graham, M., Gross, J., & Hayne, H. (2011). Risky business: Executive function, personality, and reckless behavior during adolescence and emerging adulthood. *Behavioral Neuroscience, 125* (6), 970-978. https://doi.org/10.1037/a0025768.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Revelle, W. (2021). *psych: Procedures for psychological, psychometric, and personality research*. https://cran.r-project.org/web/packages/psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rowe, K., Pozuelo, J. R., Nickless, A., Nkosi, A. B., dos Santos, A., Wagner, R. G., Kahn, K., Tollman, S., Scerif, G., & Stein, A. (2021). The Adolescent HIV Executive Function and Drumming (AHEAD) study, a feasibility trial of a group drumming intervention amongst adolescents with HIV. Under review.

Rowe, K., Duta, M., Demeyere, N., Wagner, R.G., Pettifor, A., Kahn, K., Tollman, S., Scerif, G. and Stein, A. (2021). Validation of Oxford Cognitive Screen: Executive Function (OCS-EF), a tablet-based executive function assessment tool amongst adolescent females in rural South Africa. *International Journal of Psychology*. https://doi.org/10.1002/ijop.12764

Sania, A., Sudfeld, C. R., Danaei, G., Fink, G., Mccoy, D. C., Zhu, Z., Fawzi, M. C. S., Akman, M., Arifeen, S. E., Barros, A. J. D., Bellinger, D., Black, M. M., Bogale, A., Braun, J. M., van den Broek, N., Carrara, V., Duazo, P., Duggan, C., Fernald, L. C. H., … & Fawzi, W. (2019). Early life risk factors of motor, cognitive and language development: A pooled analysis of studies from low/middle-income countries. *BMJ Open, 9*(10), E026449. https://doi.org/10.1136/bmjopen-2018-026449

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology, 35*(6), 679-697. https://doi.org/10.1080/87565641.2010.508546

Sousa, V. D., & Rojjanasrirat, W.. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice, 17*(2), 268-274. https://doi.org/10.1111/j.1365-2753.2010.01434.x

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., CONNERS, C. K., Abikoff, H. B., Clevenger, W., Davies, M., Elliott, G. R., Greenhill, L. L., Hechtman, L., Hoza, B., Jensen, P. S., March, J. S., Newcorn, J. H., Owens, E. B., Pelham, W. E., Schiller, E., Severe, J. B., … Wu, M. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(2), 168–179. https://doi.org/10.1097/00004583-200102000-00011

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436-452. https://doi.org/10.1016/j.jecp.2010.08.008

Willoughby, M. T., Piper, B., Kwayumba, D., & McCune, M. (2019). Measuring executive function skills in young children in Kenya. *Child Neuropsychology*, *25*(4), 425–444. https://doi.org/10.1080/09297049.2018.1486395

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237

World Health Organization. (2020). *Adolescent Mental Health.* https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

Wray, C., Kowalski, A., Mpondo, F., Ochaeta, L., Belleza, D., DiGirolamo, A., Waford, R., Richter, L., Lee, N., Scerif, G., Stein, A. D., & Stein, A. (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. https://doi.org/10.1371/journal.pone.0242936

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

Yoo, J. E. (2009). The effect of auxiliary variables and multiple imputation on parameter estimation in confirmatory factor analysis. *Educational and Psychological Measurement*, *69*(6), 929–947. https://doi.org/10.1177/0013164409332225

Yuan, H., Ocansey, M., Oaks, B., Sheridan, M., Okronipa, H., Hamoudi, A., Kumordzie, S., Adu-Afarwuah, S., & Prado, E. (2020). Feasibility of using tablet-based cognitive assessments in a large randomized trial in Ghana. *Current Developments in Nutrition*, *4*(Suppl 2), 1110. https://doi.org/10.1093/cdn/nzaa054_182

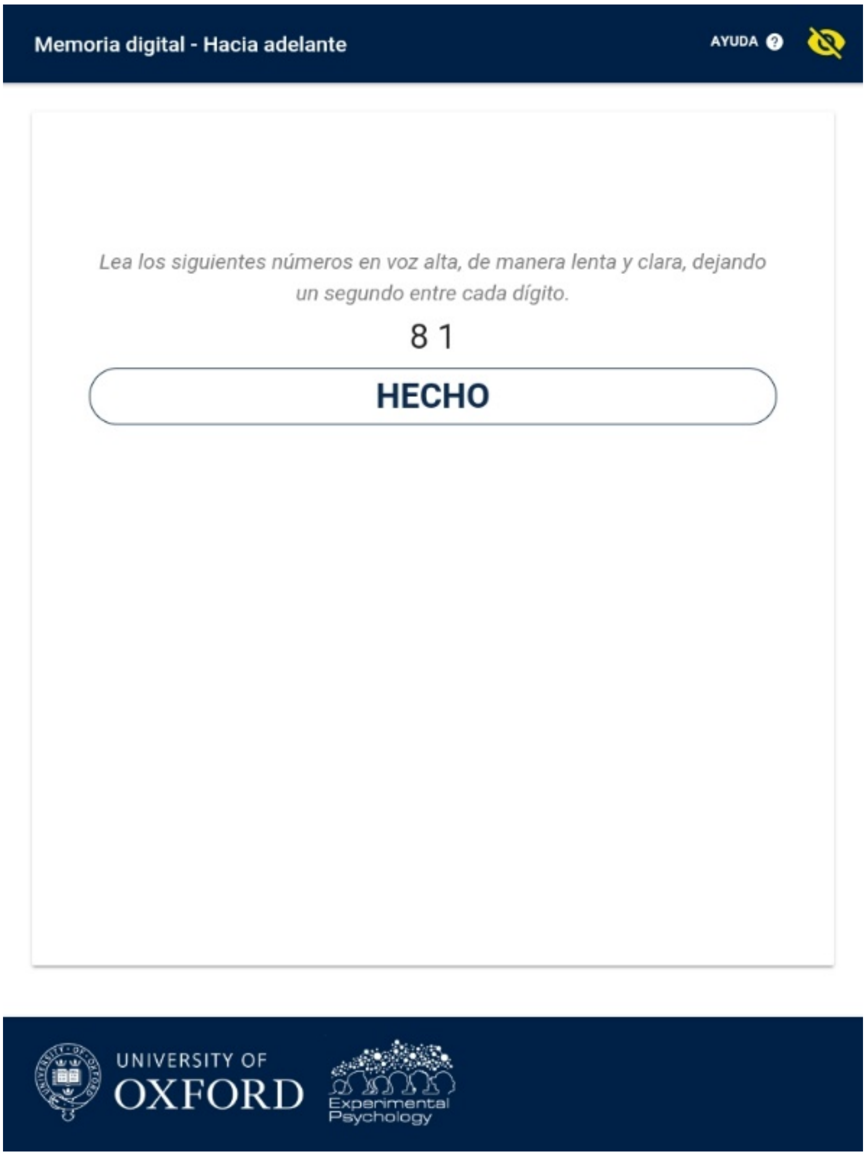
Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354-360. https://doi.org/10.1111/j.1750-8606.2012.00246.

Appendix A

OCS-EF & RACER Task Figures

Figure I

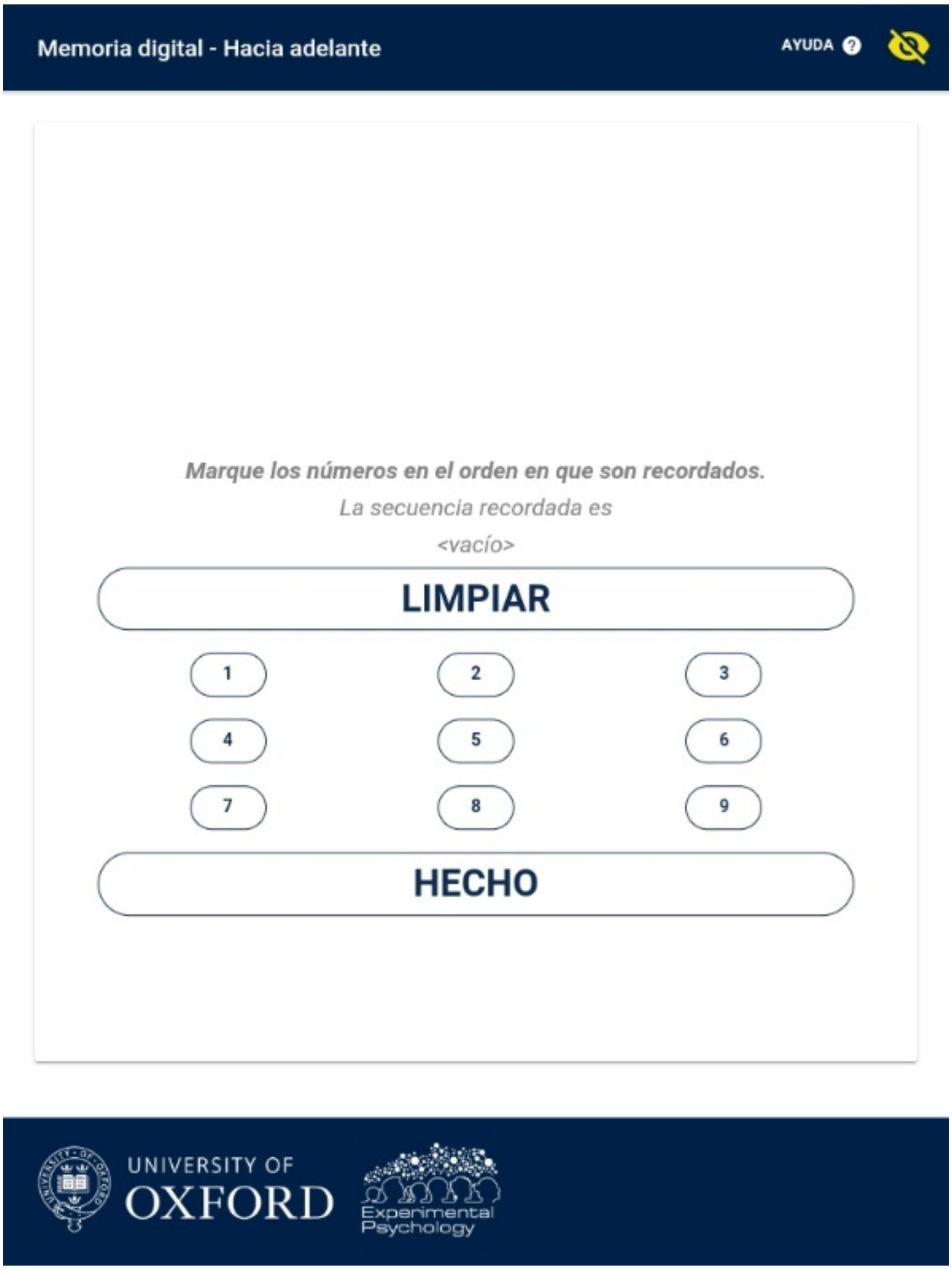
OCS-EF Digit Recall Task Screen A



Note. The instructions ask the fieldworker to read the displayed values to the participant.

Figure II

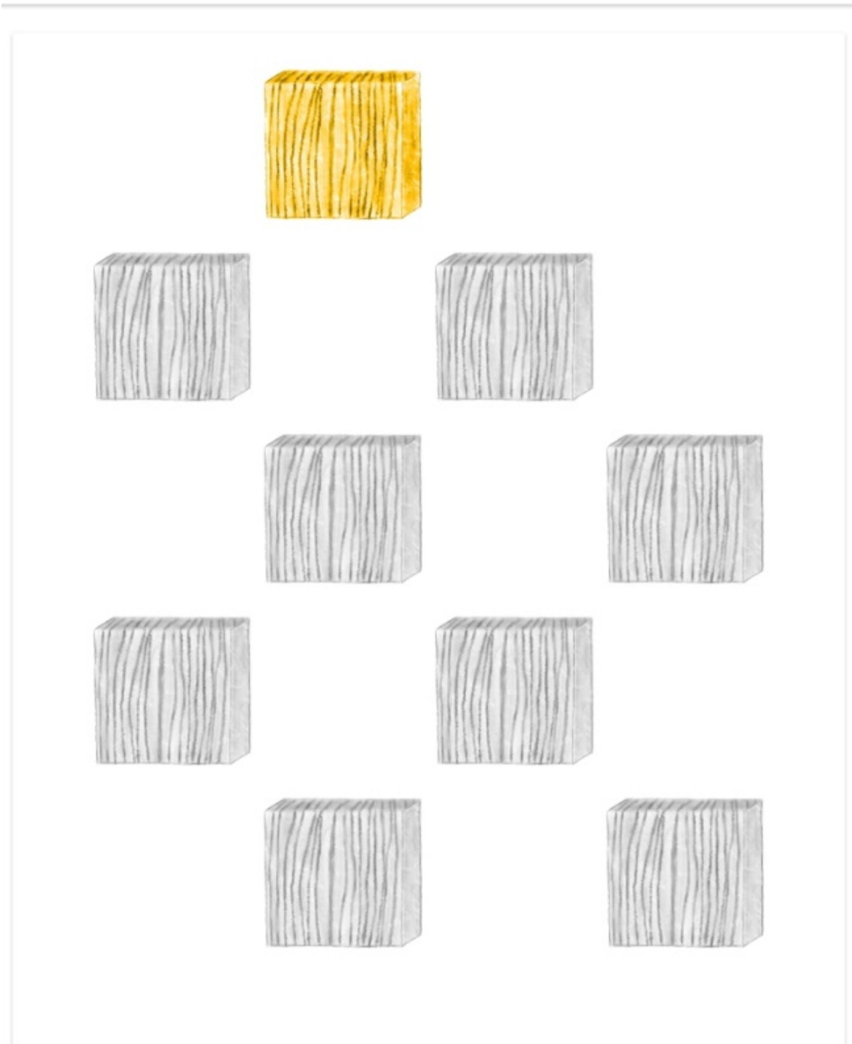
OCS-EF Digit Recall Task Screen B



Note. The numbers read in the previous step are entered on this screen.

Figure III

OCS-EF Spatial Span Task Screen



Note. The blocks are highlighted in succession, in a pattern.

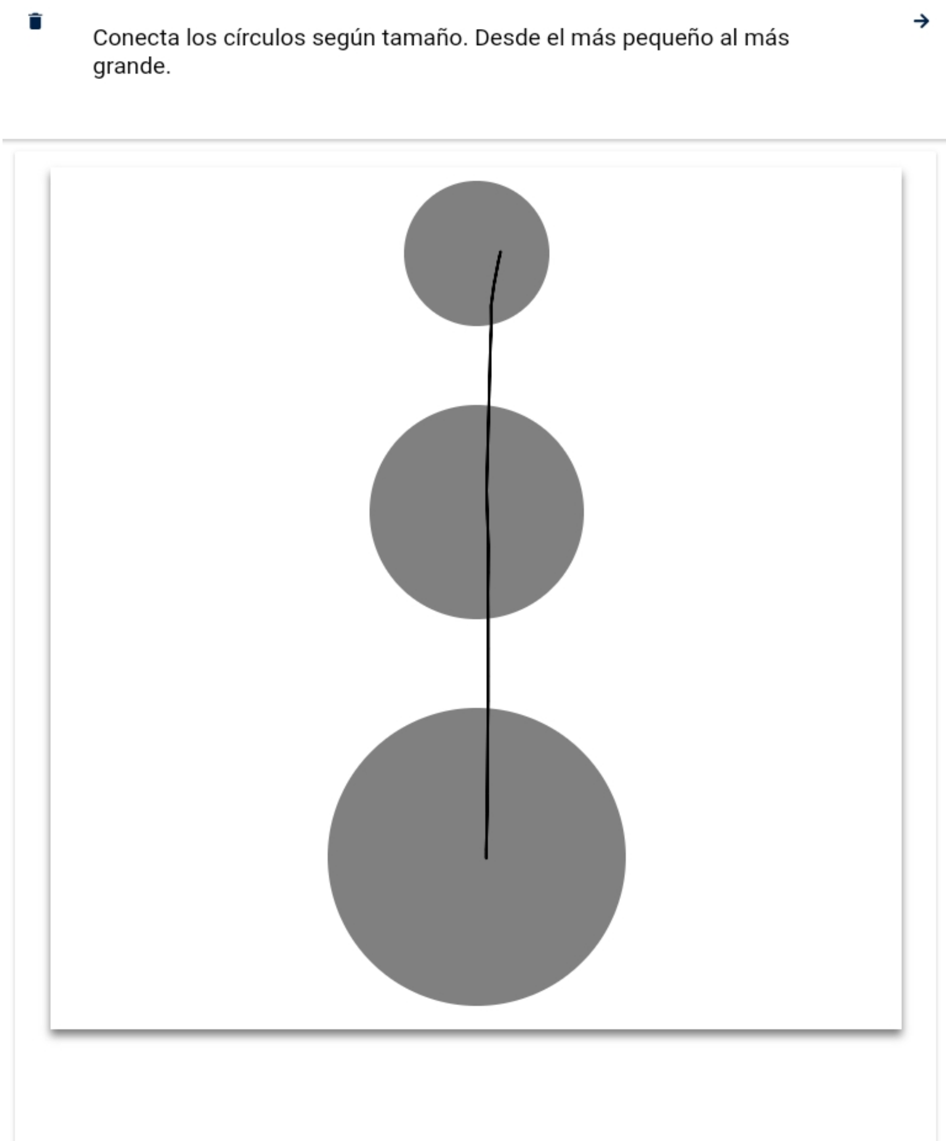
Figure IV

OCS-EF Go/No-Go Task Screen

Note. The grey circles and triangles appear in the white space above the grey bar. The participants were instructed to press the grey bar when a circle appeared.

Figure V

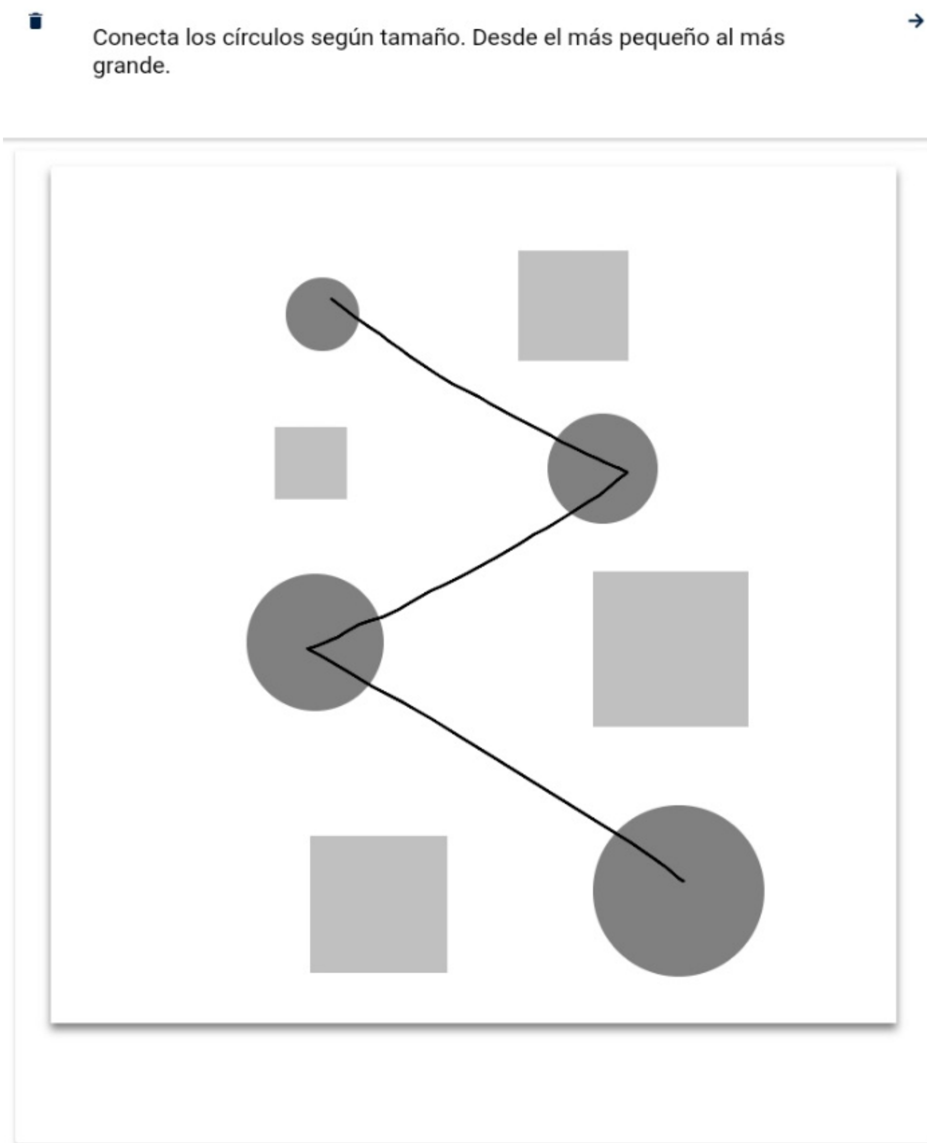
OCS-EF Trails Task Screen A



Note. In the circle-only condition, the participants were asked to draw a line from the smallest shape to the largest.

Figure VI

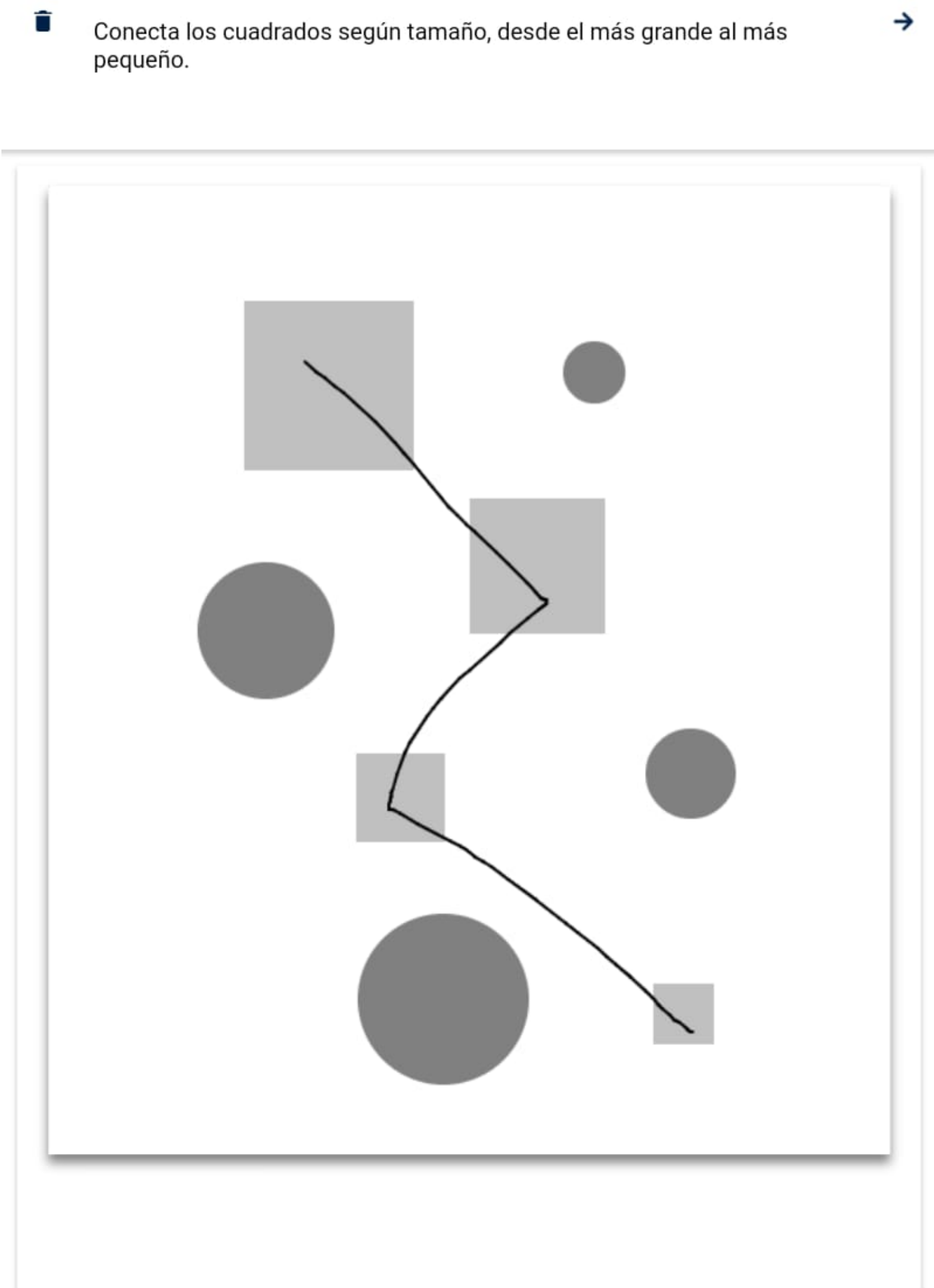
OCS-EF Trails Task Screen B



Note. In the circle-only condition, the participants were asked to draw a line from the smallest shape to the largest. The squares acted as distractor shapes.

Figure VII

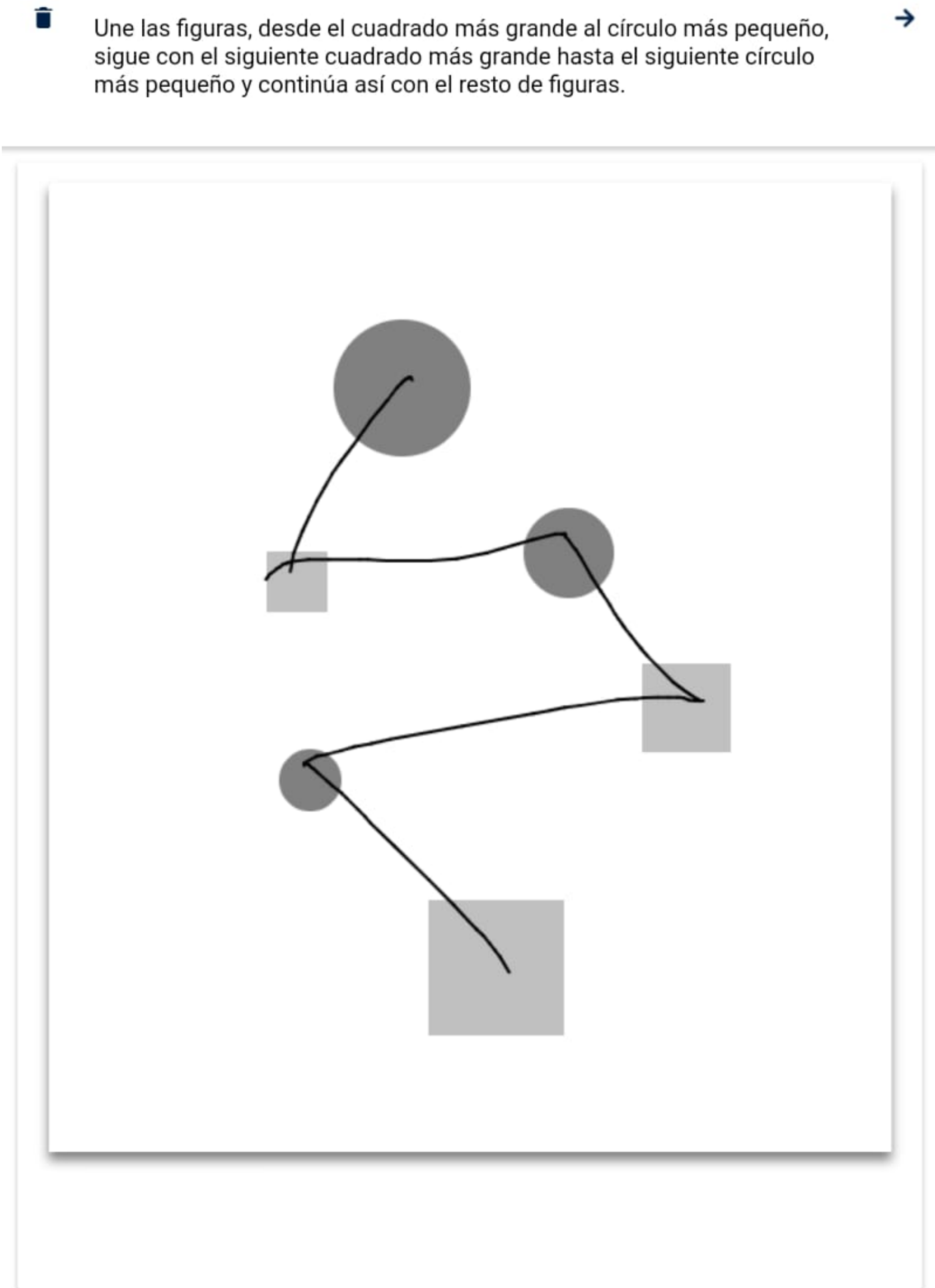
OCS-EF Trails Task Screen C



Note. In the square-only condition, the participants were asked to draw a line from the largest shape to the smallest. The circles acted as distractor shapes.

Figure VIII

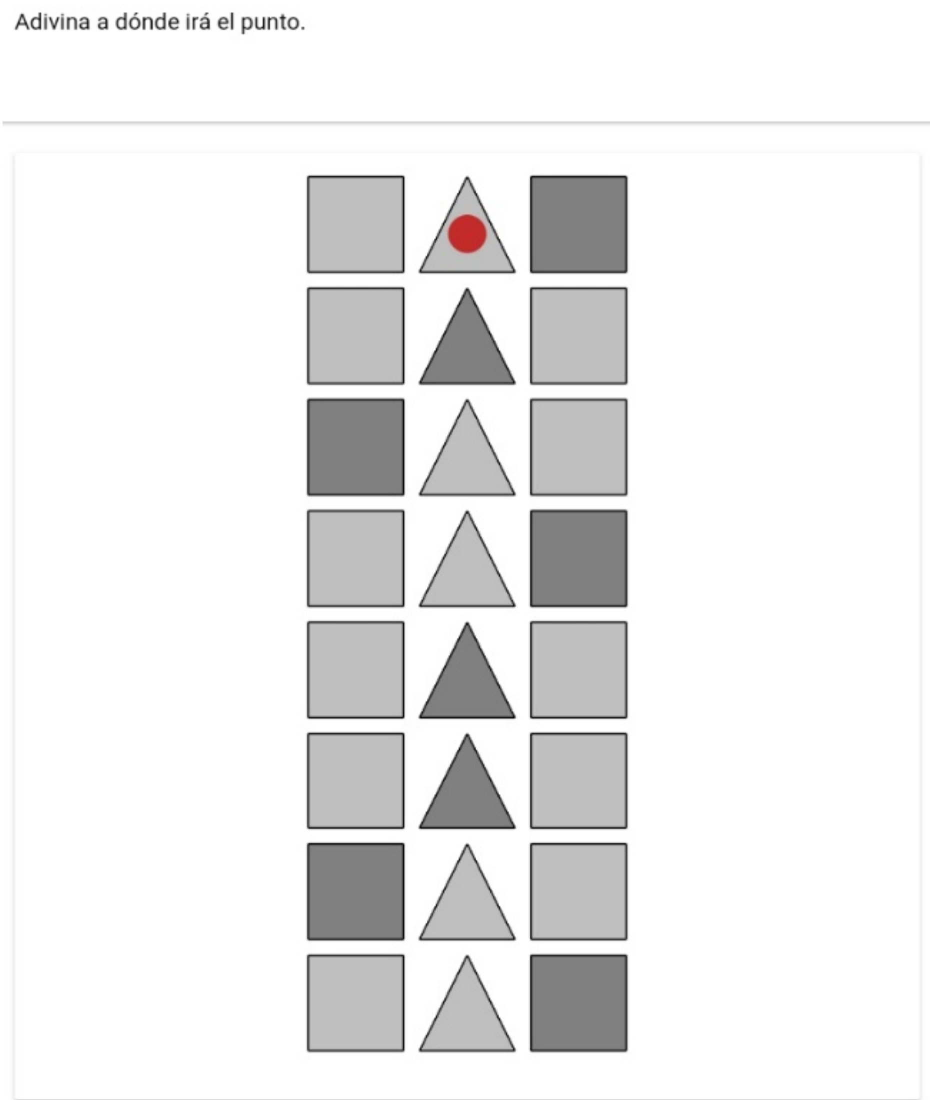
OCS-EF Task Screen D



Note. In the switch condition, the participants were asked to draw a line from the largest square to the smallest circle, the second largest square to the second smallest circle, etc.

Figure IX

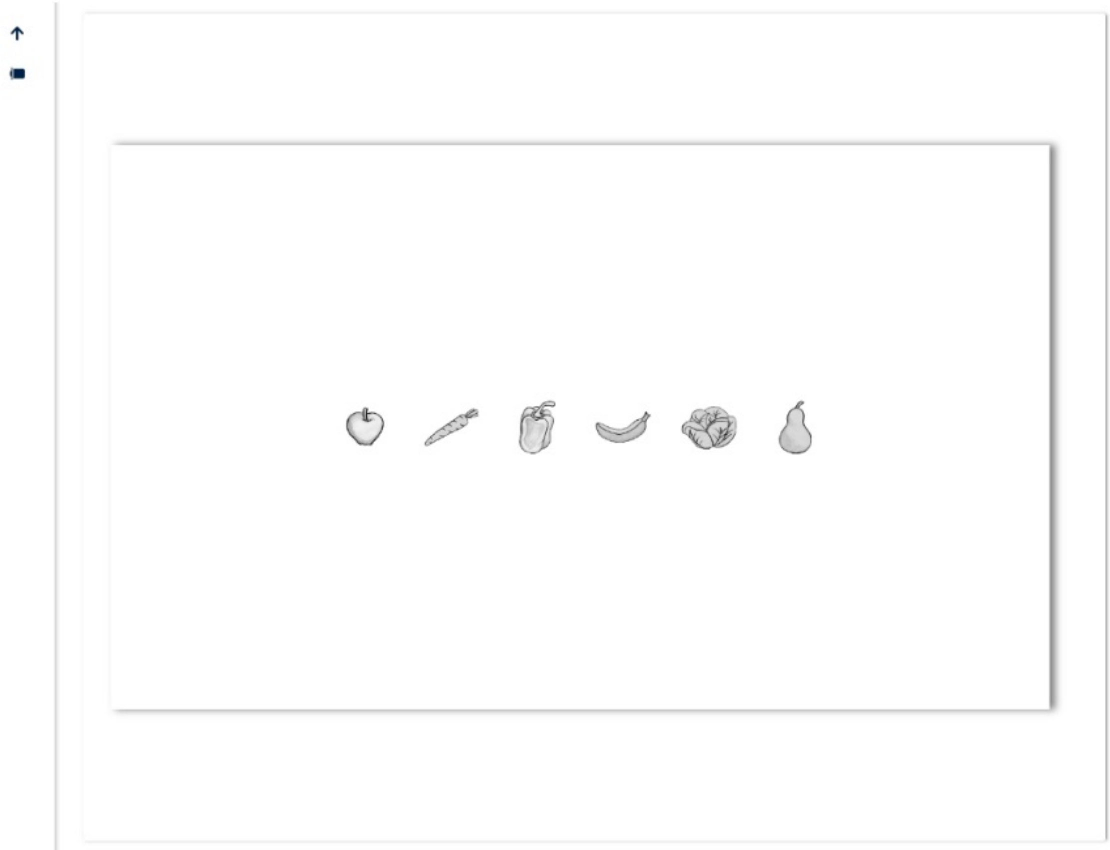
OCS-EF Rule Finding Task Screen



Note. The red dot represents the current location. The participants are asked to predict where the dot will appear next.

Figure X

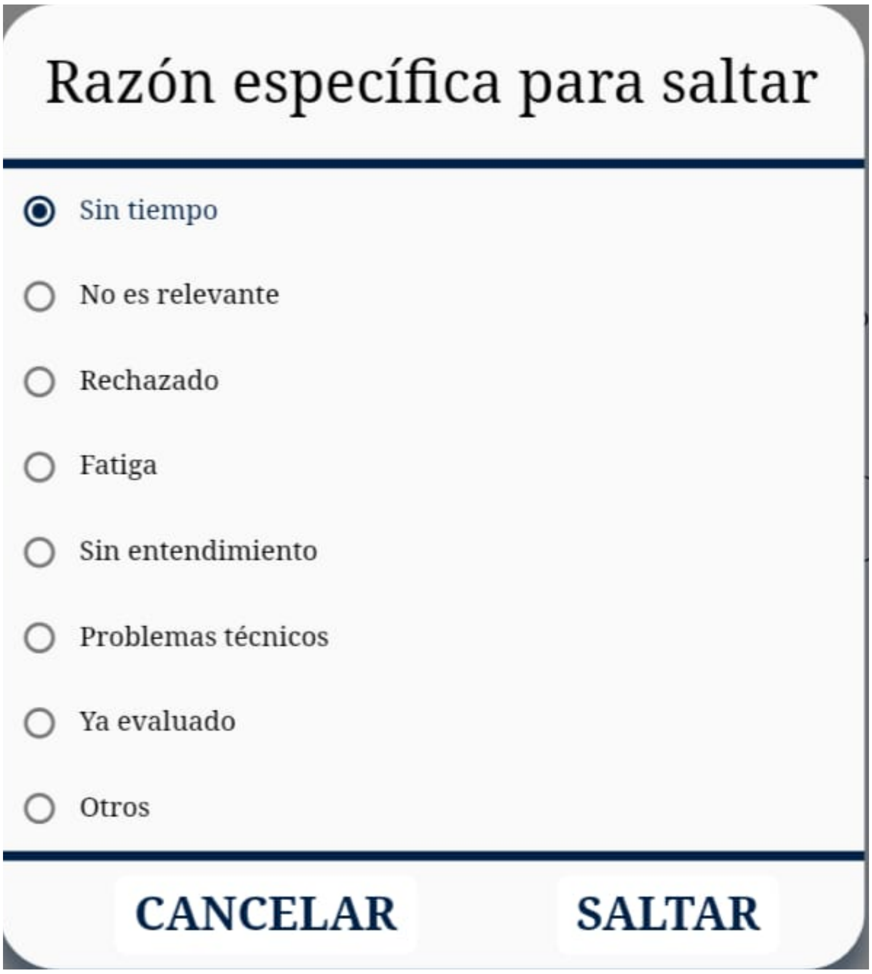
OCS-EF Selection Task Screen



Note. The above fruits and vegetables were shown in the practice section.

Figure XI

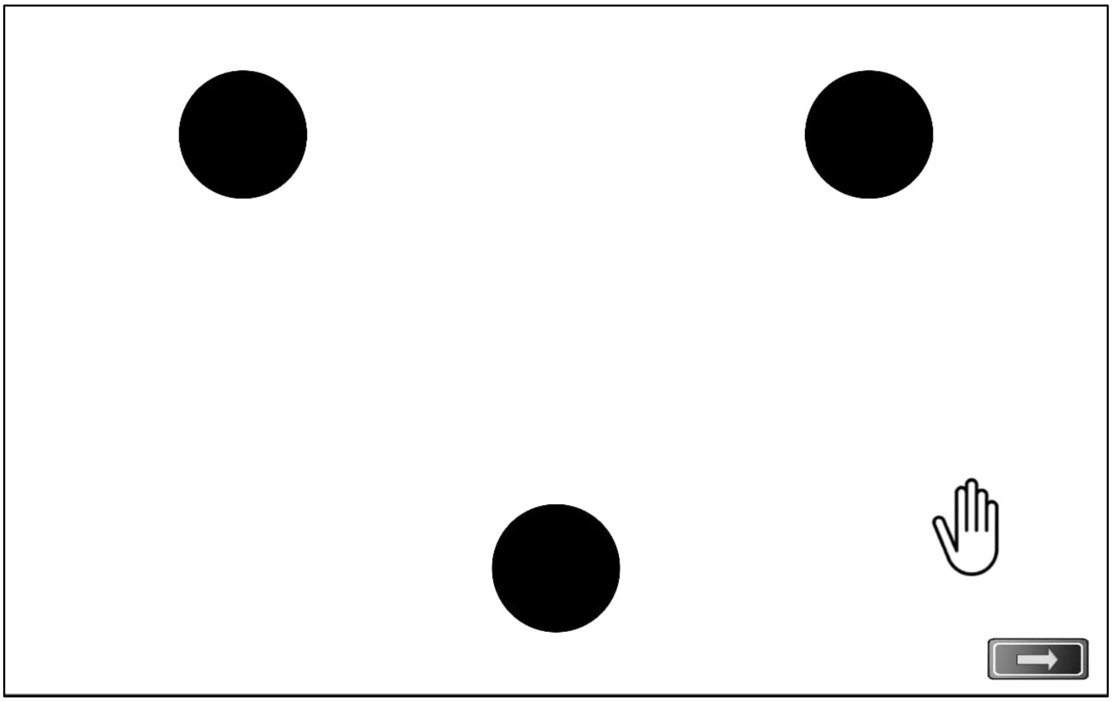
OCS-EF Condition Selection Screen



Note. Fieldworkers were asked to select the condition of the participants.

Figure XII

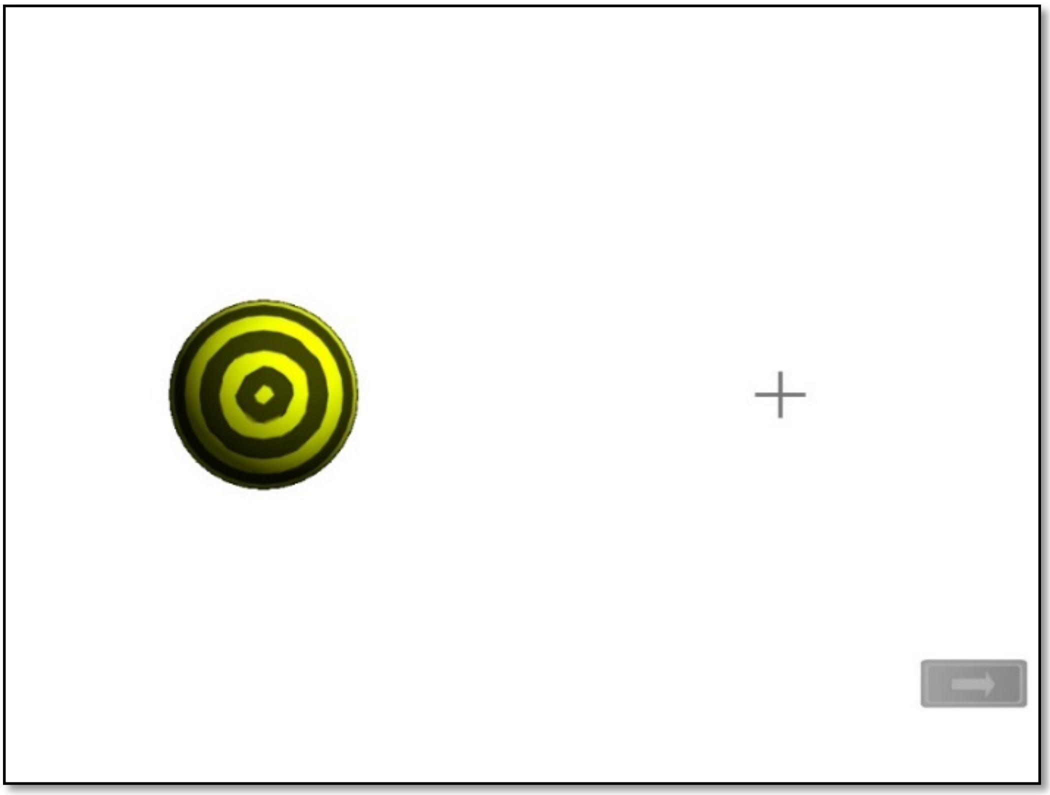
RACER Spatial Delayed Match to Sample Task Screen



Note. Participants were asked to memorise the location of these dots (high-load condition) and to tap on the screen on their locations once the screen turned black.

Figure XIII

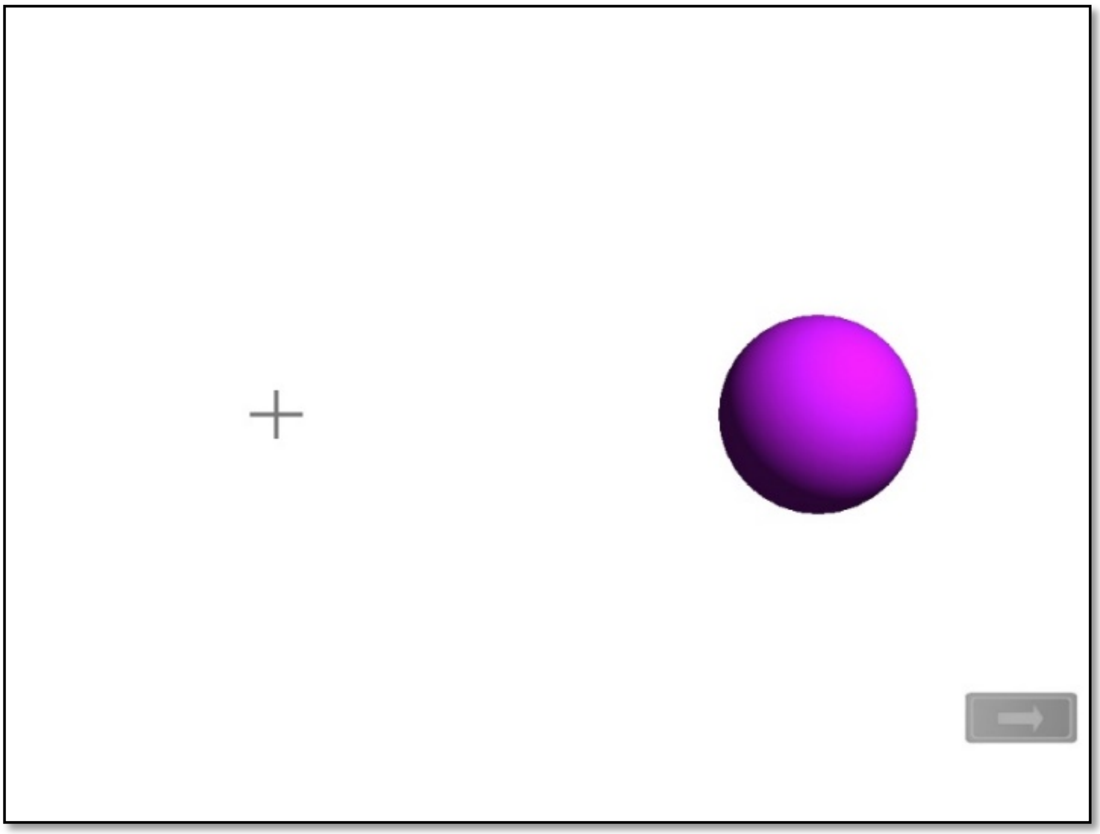
RACER Simon Task Screen A



Note. This is the same side condition. The yellow and black dot is meant to be pressed in this condition.

Figure XIV

RACER Simon Task Screen B



Note. This is the opposite side condition. The cross opposite of the pink dot is meant to be pressed in this condition.

Appendix B

SNAP-IV Questionnaire

SECCION 3 – PRUEBA SNAP - IV

Q.0	ID de la persona que responde ésta Sección. (Identifique a quien responde ésta Sección utilizando el ID de la Lista de Miembros del Hogar).	[__ __]	IDSec3

DIGA: Ahora le voy a pedir su opinión acerca de varias cosas de ***[Nombre del joven]***. Quiero que me diga la respuesta que más describa la conducta de ***[Nombre del joven]*** en **las últimas dos semanas.**

ENCUESTADOR: Muestre la cartilla # 09, con 4 opciones (desde De ningún modo, Solo un poco, Bastante y Mucho).

DIGA: Aquí tiene una cartilla (lámina) con cuatro opciones que van desde ***De ningún modo*** a ***Mucho***. Le voy a leer unas frases y quiero que me respondas cuál de ellas describe mejor la conducta de ***[Nombre del joven]*** en **las últimas dos semanas**. Para ello, señala en esta cartilla (lámina) la opción que representa su respuesta.

[Nombre del responsable], ¿Tiene alguna duda sobre cómo responder?

ENCUESTADOR: Si el responsable no comprendió como responder, repita las instrucciones hasta máximo 3 veces. De lo contrario, finalice la entrevista.

Marca (X) en el recuadro que representa su respuesta

	ENCUESTADOR: No lea como una pregunta. ***Puede repetir la frase, máximo 2 veces.** **DIGA: En las últimas dos semanas…**	**01=De ningún modo**	**02=Solo un poco**	**03= Bastante**	**04= Mucho**	**77=NS**	**79=NqC**	**88=NA**
01	A menudo no presta mucha atención a los detalles o comete errores por descuido en el trabajo o las tareas escolares	☐	☐	☐	☐	☐	☐	☐
02	A menudo tiene dificultades para mantener la atención en tareas o actividades de juego.	☐	☐	☐	☐	☐	☐	☐
03	A menudo no parece escuchar cuando se le habla directamente.	☐	☐	☐	☐	☐	☐	☐
04	A menudo no sigue las instrucciones y no termina el trabajo escolar, las tareas o los deberes.	☐	☐	☐	☐	☐	☐	☐
05	A menudo tiene dificultades organizando tareas y actividades.	☐	☐	☐	☐	☐	☐	☐
06	A menudo evita, disgusta o se involucra de mala gana en tareas que requieren un esfuerzo mental sostenido.	☐	☐	☐	☐	☐	☐	☐
07	A menudo pierde cosas necesarias para las tareas o las actividades (por ejemplo, juguetes, tareas escolares, lápices o libros).	☐	☐	☐	☐	☐	☐	☐
08	Frecuentemente se distrae por cualquier ruido o cosas que ocurren a su alrededor.	☐	☐	☐	☐	☐	☐	☐
09	A menudo es olvidadizo en las actividades diarias.	☐	☐	☐	☐	☐	☐	☐
10	A menudo se agita con las manos o los pies o se retuerce en el asiento.	☐	☐	☐	☐	☐	☐	☐
11	A menudo deja el asiento en el aula o en otras situaciones en las que se espera que permanezca sentado.	☐	☐	☐	☐	☐	☐	☐
12	A menudo corre o salta excesivamente en situaciones en las que es inapropiado.	☐	☐	☐	☐	☐	☐	☐

	ENCUESTADOR: No lea como una pregunta. *Puede repetir la frase, máximo 2 veces. **DIGA: En las últimas dos semanas…**	**01=De ningún modo**	**02=Solo un poco**	**03= Bastante**	**04=Mucho**	**77=NS**	**79=NqC**	**88=NA**
13	A menudo tiene dificultad para jugar o participar en actividades de ocio tranquilamente.	☐	☐	☐	☐	☐	☐	☐
14	A menudo está "en movimiento" o actúa como si estuviera "impulsado por un motor".	☐	☐	☐	☐	☐	☐	☐
15	A menudo habla en exceso.	☐	☐	☐	☐	☐	☐	☐
16	A menudo deja escapar respuestas antes de que se hayan completado las preguntas.	☐	☐	☐	☐	☐	☐	☐
17	A menudo tiene dificultades para esperar su turno.	☐	☐	☐	☐	☐	☐	☐
18	A menudo interrumpe o se entromete con otros (por ejemplo, se mete en conversaciones o juegos).	☐	☐	☐	☐	☐	☐	☐
19	A menudo pierde el temperamento.	☐	☐	☐	☐	☐	☐	☐
20	A menudo discute con los adultos.	☐	☐	☐	☐	☐	☐	☐
21	A menudo desafía activamente o rechaza las solicitudes o reglas de los adultos.	☐	☐	☐	☐	☐	☐	☐
22	A menudo hace deliberadamente cosas que molestan a otras personas.	☐	☐	☐	☐	☐	☐	☐
23	A menudo culpa a otros por sus errores o malas conductas.	☐	☐	☐	☐	☐	☐	☐
24	A menudo es delicado o se molesta fácilmente por los demás.	☐	☐	☐	☐	☐	☐	☐
25	A menudo está enojado y resentido.	☐	☐	☐	☐	☐	☐	☐
26	A menudo es rencoroso o vengativo.	☐	☐	☐	☐	☐	☐	☐

SECTION 3 – SNAP–IV TEST

Q.0	**ID of respondent for this section** (identify the respondent for this section using the ID from the household roster)	[__ __]	IDSec3

SAY: Now, I am going to ask your opinion about many things about ***[YL Child Name].*** I want you to tell me the answer that most describes ***[YL Child Name]*** behavior in the **last two weeks.**

FIELDWORKER: Show chart # 09, with 4 options (from Not at all, Just a little, Quite a bit, and Very much).

SAY: Here is a chart (slide) with four options ranging from ***Not at all*** to ***Very much.*** I am going to read you some sentences and I want you to answer me which of them best describes the behavior of ***[YL Child Name]*** in the **last two weeks**. To do this, indicate in this chart (sheet) the option that represents your answer.

[Name of the caregiver], Do you have any questions about how to respond?

FIELDWORKER: If the caregiver did not understand how to respond, repeat the instructions up to 3 times. Otherwise, end the interview.

Mark (X) in the box representing your answer

	FIELDWORKER: Don't read as a question. *** You can repeat the phrase, maximum 2 times.** **SAY: In the last two weeks …**	**01= Not At All**	**02= Just A Little**	**03= Quite A Bit**	**04= Very Much**	**77=NK**	**79=RTA**	**88=NA**
01	Often fails to give close attention to details or makes careless mistakes in schoolwork or tasks	☐	☐	☐	☐	☐	☐	☐
02	Often has difficulty sustaining attention in tasks or play activities	☐	☐	☐	☐	☐	☐	☐
03	Often does not seem to listen when spoken to directly	☐	☐	☐	☐	☐	☐	☐
04	Often does not follow through on instructions and fails to finish schoolwork, chores, or duties	☐	☐	☐	☐	☐	☐	☐
05	Often has difficulty organizing tasks and activities	☐	☐	☐	☐	☐	☐	☐
06	Often avoids, dislikes, or reluctantly engages in tasks requiring sustained mental effort	☐	☐	☐	☐	☐	☐	☐
07	Often loses things necessary for activities (e.g., toys, school assignments, pencils, or books)	☐	☐	☐	☐	☐	☐	☐
08	Often is distracted by extraneous stimuli	☐	☐	☐	☐	☐	☐	☐
09	Often is forgetful in daily activities	☐	☐	☐	☐	☐	☐	☐
10	Often fidgets with hands or feet or squirms in seat	☐	☐	☐	☐	☐	☐	☐
11	Often leaves seat in classroom or in other situations in which remaining seated is expected	☐	☐	☐	☐	☐	☐	☐
12	Often runs about or climbs excessively in situations in which it is inappropriate	☐	☐	☐	☐	☐	☐	☐
13	Often has difficulty playing or engaging in leisure activities quietly	☐	☐	☐	☐	☐	☐	☐

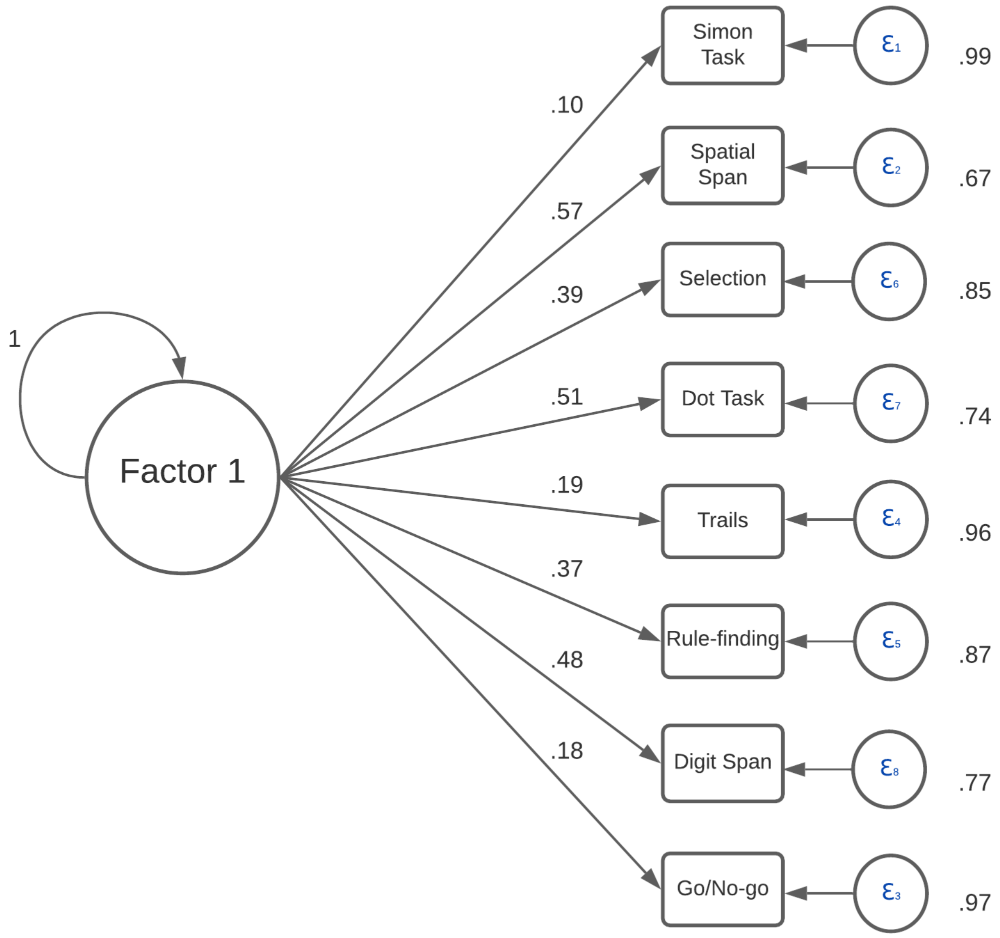
	FIELDWORKER: Don't read as a question. *** You can repeat the phrase, maximum 2 times.** **SAY: In the last two weeks ...**	**01= Not At All**	**02= Just A Little**	**03= Quite A Bit**	**04= Very Much**	**77=NK**	**79=RTA**	**88=NA**
14	Often is "on the go" or often acts as if "driven by a motor"	☐	☐	☐	☐	☐	☐	☐
15	Often talks excessively	☐	☐	☐	☐	☐	☐	☐
16	Often blurts out answers before questions have been completed	☐	☐	☐	☐	☐	☐	☐
17	Often has difficulty awaiting turn	☐	☐	☐	☐	☐	☐	☐
18	Often interrupts or intrudes on others (e.g. butts into conversations/ games)	☐	☐	☐	☐	☐	☐	☐
19	Often loses temper	☐	☐	☐	☐	☐	☐	☐
20	Often argues with adults	☐	☐	☐	☐	☐	☐	☐
21	Often actively defies or refuses adult requests or rules	☐	☐	☐	☐	☐	☐	☐
22	Often deliberately does things that annoy other people	☐	☐	☐	☐	☐	☐	☐
23	Often blames others for his or her mistakes or misbehaviour	☐	☐	☐	☐	☐	☐	☐
24	Often touchy or easily annoyed by others	☐	☐	☐	☐	☐	☐	☐
25	Often is angry and resentful	☐	☐	☐	☐	☐	☐	☐
26	Often is spiteful or vindictive	☐	☐	☐	☐	☐	☐	☐

Appendix C

Adolescent-Only One-Factor CFA Results

Figure XV

One-factor CFA Model



Note. The factor loadings are standardised. Dot Task = Spatial Delayed Match to Sample.

Figure XVI

One-Factor Model Parameter Estimates

Loading	Coefficient	SE	z	p
EF - Go/No-Go	0.18	0.08	1.52	0.13
EF - Simon Task	0.10	5.66	1.06	0.29
EF - Trails	0.19	0.19	0.01	2.18
EF - Rule Finding	0.37	0.02	3.86	<.001
EF - Digit Recall	0.48	0.09	5.15	<.001
EF - Spatial Span	0.57	0.12	6.01	<.001
EF - Selection	0.39	0.02	4.16	<.001
EF - Dot Task	0.51	6.35	4.57	<.001

Note. Dot Task = Spatial Delayed Match to Sample. The factor loading coefficients are standardised.