JOURNAL OF EDUCATIONAL SCIENCES & PSYCHOLOGY

Vol. XII (LXXIV)
No. 1/2022

113 - 118

Psychometric properties of the Comprehensive Executive Function Inventory - Self-Report in a sample of Polish adolescents

Anna Hawrot[a] *, Sara Filipiak[b]

[a]Leibniz Institute for Educational Trajectories (LIfBi), Department of Competencies, Personality, Learning Environments, Wilhelmsplatz 3, 96047 Bamberg, Germany
[b]Institute of Psychology, Maria Curie-Skłodowska University, ul. Głęboka 45 (Budynek B) 20-612 Lublin, Poland

Abstract

This study aimed at preparing a Polish self-report version of the Comprehensive Executive Function Inventory (CEFI), verifying its psychometric properties, and shortening it. The sample included 1109 adolescents aged 12 to 19 attending schools located in a city in eastern Poland. Drawing upon the theoretical background of the CEFI and results presented in its manual, we tested three competing factor structures: a unidimensional, nine-factor, and a bifactor structure to verify the inventory's construct validity. These analyses did not fully confirm any of the hypothesized models. Moreover, the results did not support the analysis of subscale scores. A shortened nine-item version had satisfactory reliability ($\alpha = .778$; $\omega = .779$), was unidimensional, and scalarly invariant across gender and age.

Keywords: Adolescent Behavior; Executive Functioning Measures; Factor Structure; Measurement Invariance

1. Introduction

Executive function is an umbrella term that refers to higher-order cognitive processes that are necessary for the cognitive control of behavior, thoughts, and emotions (Zelazo et al., 2008). Although no single definition of executive function exists, most researchers agree that it is the exertion of control over automatic responses in order to regulate intentional, autonomous, and goal-directed behaviors (Anderson, 2002; Zelazo et al., 1997). Executive function is associated with independent yet interrelated processes of inhibitory control, working memory, and set shifting (i.e., mental flexibility, Anderson, 2002; Miyake et al., 2000). It is particularly important in novel or demanding situations which require a rapid and flexible adjustment of behavior to the changing demands of the environment (Huizinga et al., 2018). It also allows the selection and successful monitoring of behaviors that facilitate the attainment of goals and problem solving (Berthelsen et al., 2017; Poon, 2018). The development of executive function in childhood and adolescence depends on the interplay between the biological maturation of the brain and external influences such as the home environment or teaching practices (Carlson, 2003; Luca et al., 2003; Selemon, 2013; Tripon, 2021).

Because executive function is manifested in many aspects of the everyday lives of adolescents (e.g., Abreu-Mendoza et al., 2018; Dekker et al., 2017; Jacob & Parkinson, 2015), the assessment of executive function is immensely important in

* Corresponding author. Tel.: +49 951 863-3453
E-mail address: anna.hawrot@lifbi.de

DOI: 10.51865/JESP.2022.1.12

public education (Gioia et al., 2002). Academic success is mostly dependent on students' ability to plan their time, organize and prioritize information, distinguish main ideas from details, and reflect on their work (Meltzer & Krishnan, 2007). Numerous learning processes are closely linked with executive function, for example, explicit and implicit learning, flexible use of learning strategies, or metacognitive knowledge about tasks (Eslinger, 1996). Executive function is also key for social and emotional functioning (e.g., Holmes et al., 2016; Séguin & Zelazo, 2005) and efficient interpersonal problem solving (Mata et al., 2017). Students with executive function deficits often engage in off-task behavior, need frequent prompts and reminders, and have unstable peer relationships (Otero et al., 2014). As a consequence, school-based interventions fostering executive function have been suggested and implemented (Dawson, 2014; Kavanaugh et al., 2019; Otero et al., 2014). However, implementing an intervention requires prior knowledge that an intervention is needed. Therefore, the diagnosis of students' executive function is a necessary prerequisite.

In Poland, measures of executive functions in adolescents are scarce. The three available scales have important limitations that reduce their feasibility in the school setting. The first scale, which is part of a Polish adaptation of the IDS-2 Scales of Intelligence and Development (Jaworowska et al., 2019), requires on average 30 minutes to complete and offers only individual assessment in children and youth aged 5 to 20. Its cost is high, which precludes wider use in publicly funded institutions. The second scale, the PU1 (Borkowska et al., 2015), is designed for younger students with a very narrow age range (10-12 years). It requires the use of a computer and additional aids, which makes it impractical when a quick group assessment is needed. The third assessment tool, the Dimensional Card Sorting Task (Test Sortowania Kart dla Dzieci, Jabłoński et al., 2018), measures inhibitory control and mental flexibility in children aged 3-11. It can be administered individually only.

To fill in this gap, this study sought to prepare a Polish-language measure of executive function in adolescence which could be suitable use in school settings. Such a measure should be time- and cost-effective, offer both individual and group assessment over the whole adolescence period and serve as a basis for tailored interventions. However, since sometimes a detailed diagnosis is not feasible or only screening is needed, we also sought to provide a short version of this measure. Such a brief version would also be useful in research on executive function and the evaluation of intervention programs.

1.1. Comprehensive Executive Function Inventory

A measurement tool that allows us to meet the goals of this study is the Comprehensive Executive Function Inventory (CEFI, Naglieri & Goldstein, 2013). The CEFI is a relatively new measure of executive function in children and adolescents. Unlike manifold performance measures (e.g., Towers of London or Hanoi, Wisconsin Card Sorting Task), which require individual assessment, the CEFI is a questionnaire and allows convenient use in both individual and group settings. The CEFI includes nine subscales which reflect various aspects of executive function, and its content is based on current theory and research as well as on the Authors' clinical and research experience (Naglieri & Goldstein, 2013). The CEFI includes three forms to be filled in by different informants: a self-report, a report by parents, and a report by educators/teachers. Its reviews have been favorable (Climie et al., 2014; Fenwick & McCrimmon, 2015), and the inventory has been used in clinical practice (Sotelo-Dynega, 2017; Stevanovic et al., 2018) and research (Cascia & Barr, 2017; Hickey & Flynn, 2019). Moreover, it can be used as a basis for tailored interventions (Naglieri & Goldstein, 2013).

Although all three perspectives provide useful and complementary information, this study focuses on the self-report form, which is designed to be completed by individuals aged 12 to 18 and can be administered both in individual and group settings. This form fills in the gap in measurement tools available in Poland to the greatest extent. Although the use of self-reports poses various challenges, research indicates their usefulness (Credé et al., 2012). This includes especially their application in a group setting and when quick and low-cost assessment is needed (Credé et al., 2012; Ebesutani et al., 2012). Self-reports are also useful for identification of individually perceived emotional and behavioral problems at an early stage (Levitt et al., 2007).

Moreover, students are the main clients of schools and the counseling centers with which both establishments work directly, which allows more frequent contact and easier access. Parents are not as easy to reach because of their multiple family and work obligations. The teacher's report also does not seem as suitable as the self-report. If multiple students required screening, one teacher would need to fill in the inventory multiple times, causing a significant burden and potentially lowering the quality of the diagnosis due to fatigue or comparison of the students being rated.

1.2. Properties of the Self-Report Form

The CEFI’s self-report validation study was performed on the normative sample of 700 U.S. adolescents aged 12-18 which was representative with respect to gender, race/ethnicity and parental educational level. Additionally, a sample of 327 students with a clinical diagnosis was included (Naglieri & Goldstein, 2013). The study have found that the inventory’s internal consistency in the general sample (measured by Cronbach’s α) equaled .97 for the full scale, and ranged from .78 (Self-Monitoring) to .86 (Attention) for the subscale scores (Naglieri & Goldstein, 2013). Exploratory factor analyses presented in the CEFI’s manual, run on the first half of the normative sample, included all diagnostic items. Based on an analysis of a scree plot, Kaiser criterion, the ratio of the first and second eigenvalues, and the interpretability of the results, its Authors concluded that the CEFI was unidimensional. Exploratory analyses run on the second half of the sample, which included parcels formed from subscale scores, led to the same conclusion.

The criterion validity of the total score is supported by the finding that adolescents from the clinical sample, that is diagnosed with attention deficit hyperactivity disorder (ADHD), autism-spectrum disorder (ASD), and mood disorder, when separately compared to adolescents from the general population, scored significantly lower on the inventory. This is expected because impairment in executive function is an element of these disorders (e.g., Gilbert et al., 2008; Martel et al., 2007; Toplak et al., 2009). The CEFI self-report is correlated with the Behavior Rating Inventory, another measure of executive function, with coefficients in an ADHD sample, a mixed clinical sample, and the general population (calculated separately) ranging from .63 to .68 (Naglieri & Goldstein, 2013).

Information on the validity of the subscales (Naglieri & Goldstein, 2013) is limited to evidence that the subscale scores differ between the ADHD, mood disorder, and ASD samples (compared separately) and the general population. They also correlate with the Behavior Rating Inventory, which supports their criterion validity. However, the pattern of correlations between the subscales of the two inventories raises questions; for example, the CEFI’s Emotion Regulation measure should correlate most strongly with the analogous subscale in the Behavior Rating Inventory, which is not the case.

The CEFI also has some limitations which the current study seeks to overcome. First, the available information about the inventory’s psychometric properties and validity is limited to what is published in its manual. As test validation is rather a process than a one-time task, it is key to collect further data on the inventory’s properties and independently verify findings with different samples and populations.

Moreover, several aspects of the inventory give rise to questions. First, despite the finding that the inventory is unidimensional, its Authors suggest that the subscale scores may be calculated and meaningfully interpreted (Naglieri & Goldstein, 2013). Evidence on the construct validity of the subscale scores is lacking. Therefore, it still needs to be verified if the CEFI’s subscales can be distinguished and treated as separate but correlated facets of executive function, or if only the total score should be used.

In addition, although the Authors of the CEFI compared its scores between age, gender, race, and nationality groups as well as between clinical groups and the general population as a mean of testing construct and criterion validity, they did not verify the CEFI's measurement invariance, which is considered to be a necessary condition for meaningful group comparisons (e.g., Vandenberg & Lance, 2000). They only compared the size of factor loading in various gender, race/ethnicity, age and clinical groups by running exploratory factor analyses in each subgroup separately and calculating a coefficient of congruence. However, no analysis of item thresholds or intercepts was performed.

Finally, the CEFI consists of 100 items. Completing the inventory is not particularly time-consuming (according to the manual, it takes about 15 minutes), which makes it useful in individual diagnosis. However, its length reduces its practical utility in screening and research and may negatively affect data quality or response rates, especially if it is paired with other measures (Bogen, 1996; Galesic & Bosnjak, 2009). Shortening the inventory would decrease the amount of time necessary to complete it, making the inventory more suitable for screening purposes or for usage when there are time constraints limiting the utility of the long version. However, it is key to retain the theoretical model and assure adequate construct coverage, including adequate representation of various aspects of the measured characteristic (Smith et al., 2000).

1.3. The present study

This study seeks to fill the gaps in the literature by pursuing the following goals. First, it aims at preparing a Polish adaptation of the CEFI self-report form. The adaptation, by offering cost-effective and flexible use in an individual and group setting, could serve as a convenient measure of executive function in schools. We also sought to verify the psychometric properties of the Polish-language version with a particular focus on internal structure, invariance across gender and age, and on evidence supporting the use of either total or subscale scores. Second, the study intends to shorten the Polish-language version in order to improve its utility for screening and all other situations when the assessment with the full version is not feasible.

We expected the Polish full version to have reliability similar in value to the original version (approximately .95). However, the reliability of a shortened version was expected to be lower. We planned to reduce the amount of time necessary to complete the inventory by two-thirds, to under 5 minutes, which translated proportionately to 27 diagnostic items (three items per subscale). The expected reliability of such a version, calculated using the Spearman-Brown formula (Nunnally & Bernstein, 1994), was .81, which we deemed satisfactory in relation to potential uses (Nunnally & Bernstein, 1994).

Drawing upon the theoretical background of the CEFI and empirical results presented in its manual, we identified three potential and competing factorial models of the CEFI: a unidimensional model, a correlated trait model, and a bifactor model. In a unidimensional model all items load on one and only one factor. This model was expected because analyses presented in the CEFI manual provided evidence on the inventory's unidimensionality. We sought to verify whether the result could be replicated in our sample.

The second hypothesized model, a correlated trait solution, is a model with nine correlated factors representing subscales in which each item loads on only one factor (e.g., Reise et al., 2010). As the CEFI includes nine subscales taping various aspects of executive function, this model reflects the intended internal structure of the inventory. If such a model fits the data, it supports the calculation and interpretation of subscale scores. However, it also suggests that shortening the CEFI, as potentially resulting in decreased reliability and reduced coverage of factor domains (Smith et al., 2000), may be difficult because the number of items in each subscale is already limited (7 to 12 items).

Third, we hypothesized a bifactor model. In the bifactor model, each item loads both on (a) one general factor, accounting for variance shared by all items, and on (b) one of several specific facets, accounting for variance shared by a subset of items that is over and above the general factor. All the factors are uncorrelated (e.g., Reise et al., 2010). This model is appropriate when a construct is unidimensional but complex, and therefore items in a scale measuring that construct tap a wide variety of behaviors indicative of the construct. As a consequence, items representing similar behaviors tend to cluster together, providing some evidence for multidimensionality even if the inventory is essentially unidimensional (e.g., Reise et al., 2010). We hypothesized that the bifactor model may be a potential factorial solution because the original version includes subscales taping different types of behavior (e.g., planning, emotion regulation) in which processes of executive function are expressed. However, exploratory factor analyses in the inventory's manual provided some evidence for unidimensionality (Naglieri & Goldstein, 2013). All expected factorial solutions are presented in a simplified way in Figure 1 (Models 1-3).

Irrespective of the internal structure of the CEFI, we expected the long and shortened Polish-language versions to be invariant across gender and age (early and late adolescence).

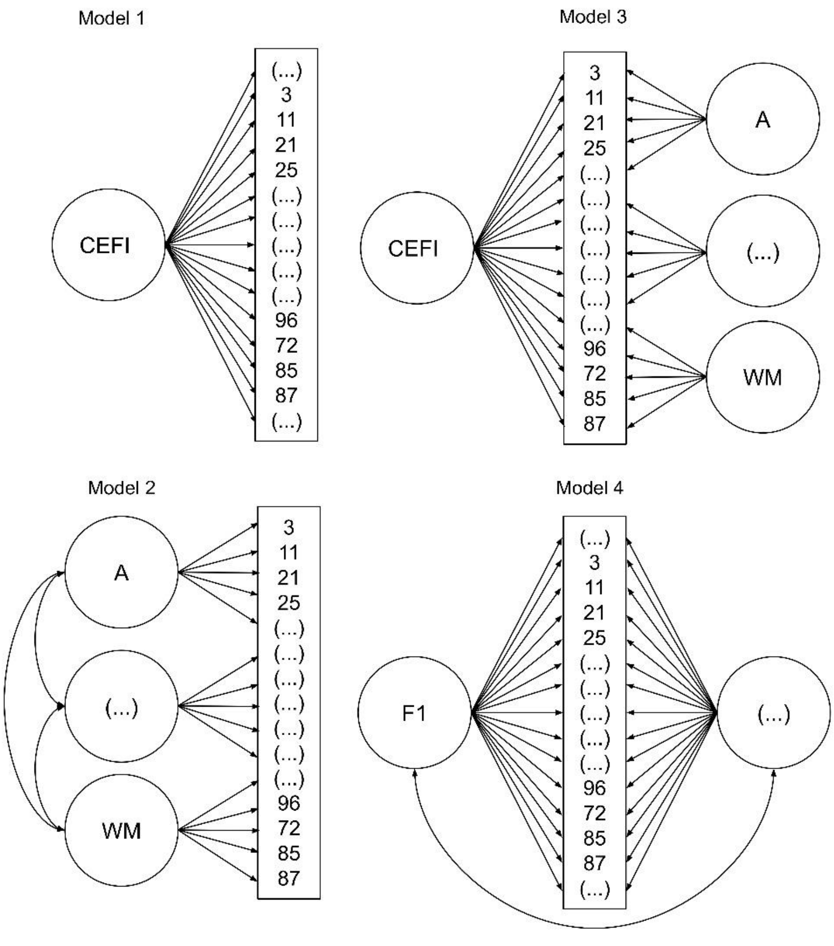

Fig. 1. Hypothesized Models of Factorial Structure of the Comprehensive Executive Function Inventory, Self-Report Form. Note. Model 1 represents a unidimensional structure in confirmatory factor analysis. Model 2 represents a multidimensional structure (correlated trait model). Model 3 represents a bifactor structure. Model 4 represents an oblique multidimensional structure in exploratory factor analysis. F = Factor, A = Attention; WM = Working Memory.

2. Methodology

2.1. Translation procedure

We prepared a Polish-language version of the inventory using a back-translation procedure. Although it has some drawbacks, this method remains in use (e.g., Sălăgean, Stan & Crișan, 2021) as it is very popular and strongly recommended in cross-cultural research (Brislin & Freimanis, 2001). After obtaining the publisher's permission, the items were translated into Polish by two translators working independently, with a particular focus on the theoretical content of the respective subscales and equivalence of meaning. Next, the translators compared their translations, discussed discrepancies, and reconciled them to produce a final forward translation. Next, a third translator, who did not have access to the original version of the inventory, translated the inventory back into English. The back translation was verified by the publisher, and all the discrepancies were resolved in discussions between the translators and the publisher. To assure the high quality of the translation, all three translators held at least M.A. degrees in both psychology and English studies (one of the forward translators was a psychology professor), were fluent in Polish and English, and had familiarized themselves with the theoretical background of the inventory and research on executive function before starting the procedure (Ercikan & Lyons-Thomas, 2013). In the next step, the final translation was tested in order to verify its psychometric properties and prepare a shortened version.

2.2. Participants

A total of 1109 (53.56% female) adolescents took part in the study. Their age ranged between 12 and 19 (mean age = 14.5, SD = 1.16); however 84% of them were aged between 13 and 15. The students attended eight public schools located in a city in eastern Poland; they were included in the sample only if they had agreed to participate and their parents had provided informed consent. The inventory was administered by a trained psychologist during a single lesson in the schools. Students were informed about the scientific purpose of the study, the confidentiality of the information provided, voluntary participation, and the possibility of withdrawing from the study at any stage.

2.3. Instrument

The CEFI (Naglieri & Goldstein, 2013) consists of 100 items, 90 of which are of a diagnostic nature. The diagnostic items refer to behaviors that are indicative of executive function and are divided into nine subscales: Attention (12 items), Emotion Regulation (9 items), Flexibility (7 items), Inhibitory Control (10 items), Initiation (10 items), Organization (10 items), Planning (11 items), Self-Monitoring (10 items), and Working Memory (11 items). Each item has a form of a short question referring to a behavior indicative of executive function. The test persons are asked to specify the frequency of a given behavior for each item using six response options: never, rarely, sometimes, often, very often, and always. The ten non-diagnostic items are used to assess potential bias in the results and are not analyzed in the present study. Since the copyright holder of the inventory does not allow to disclose the items, example items cannot be provided.

2.4. Data analysis

The first step of the analyses involved checking the descriptive statistics and reliabilities of the total and subscale scores. Next, we verified the inventory's internal structure by testing the three competing factorial solutions: the unidimensional model, the nine-factor confirmatory factor analysis (CFA) model, and the bifactor model.

We chose to perform an EFA to test the unidimensional model because it was the basis for concluding the inventory's unidimensionality in the CEFI manual. The model is depicted in Figure 1 (Model 4). Since unidimensional EFA and unidimensional CFA models are equivalent, we did not test a one-factor CFA model separately.

All measurement models used the MLR estimator and were run in Mplus 8.2 (Muthén & Muthén, 1998-2017), whereas the remaining calculations were performed in Stata 15.1. The scales of CFA and bifactor factors were set by fixing one factor loading to unity. Because the subscale scores are supposed to represent various areas in which executive function is expressed, we expected them to correlate and therefore used an oblique rotation in the EFA. We chose Oblimin to ensure consistency with the analyses presented in the CEFI's manual. The following commonly used criteria served as a basis for determining the number of factors in the EFA (e.g., Hayton et al., 2004): parallel analysis with 500 simulated datasets, the scree test, Kaiser criterion, the ratio of the first and the second eigenvalues (Zwick & Velicer, 1986), and the interpretability of the results. Although parallel analysis is recommended over the widely criticized scree test, Kaiser criterion, and the ratio of the first and the second eigenvalues (e.g., Hayton et al., 2004), we included the latter three criteria because they served as selection criteria in analyses presented in the manual (Naglieri & Goldstein, 2013).

To test measurement invariance (configural/general structure, metric/loading, and scalar/intercept), we used a multigroup CFA framework with a default Mplus specification (Muthén & Muthén, 1998-2017).

Model fit was assessed with four commonly used fit indices: the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the standardized root mean squared residual (SRMR). Although we assumed that CFI and TLI values not lower than .95, a RMSEA not higher than .06, and a SRMR not higher than .08 indicated a good fit (Hu & Bentler, 1999), we were aware that CFI and TLI might be downward biased in complex models (e.g., Cheung & Rensvold, 2002). However, both indices perform reasonably well in large samples ($N >= 500$) even if as many as 120 items are included (Shi et al., 2019).

2.5. Shortening procedure

To fulfill the second goal of the study, that is to prepare a shortened version of the inventory that could be used in screening and research, we followed general guidelines suggested by Nunnally and Bernstein (Nunnally & Bernstein, 1994). Meanwhile, we sought to retain the theoretical model and ensure an adequate coverage of factor domains. To this end, we decided to retain an equal number of items from each subscale and used two types of information while selecting items: information on their relevance for executive function delivered by experts in the field and information on items' statistical properties. Similar procedures are recommended in the literature (Smith et al., 2000) and have been used in research (e.g., Kruyen et al., 2013; Lockwood et al., 2017).

In the first step of the shortening procedure, we asked two content experts (a psychology professor with extensive experience in child and adolescent counseling, and a senior researcher, both with a publication record in the field of executive function) to independently indicate which items in each subscale had the highest relevance for the construct. They could also comment on the items' fit, clarity, and assignment to subscales. The experts were requested to select up to five items most indicative for a given aspect of executive function in each subscale. This allowed us to determine which items were key and should be considered for inclusion in the shortened version. Meanwhile, we refrained from setting a minimum number of items in each subscale the experts should select to avoid prompting them to indicate items they did not find crucial for valid measurement (Hoyt, 2000).

Second, we ran an item-level analysis and evaluated all diagnostic items using the following criteria (Lockwood et al., 2017; Nunnally & Bernstein, 1994): (a) the item-rest correlation with the total and its own subscale score, with values not lower than .3 providing support for the convergent validity of an item; (b) the item-rest correlation of a given item with its own subscale higher than the correlation of this item with other subscale scores, with a difference of .1 supporting the discriminant validity of an item; (c) an item mean close to the center of the range of possible scores and a relatively high variance of an item.

Next we combined information from both sources and narrowed down the pool of items considered for inclusion in the shortened version to those indicated by both experts as key. Then, we further reduced the pool to those items that fulfilled our selection criteria to the greatest extent and selected best performing items. In the last step, the experts evaluated the selected and discarded items in order to check if the selected items adequately represented executive function and no key items were dropped.

3. Results

Table 1 presents the descriptive statistics for the inventory's full and subscale scores. The internal consistency of the full scale as measured by Cronbach's α was very high (.95). The reliabilities of the subscale scores were visibly lower and ranged between .65 and .79. Additional information on the distributions of the total score and subscale scores is available in the online supplement.

Table 1. Descriptive statistics and reliabilities

Variable	Mean	SD	Skewness	Kurtosis	Min	Max	α
Age	14.502	1.159	0.766	0.572	12.00	19.00	
CEFI Total Score	3.866	0.526	0.226	-0.047	2.400	5.811	.947
Attention	3.660	0.682	0.097	0.091	1.167	6.000	.790
Emotion Regulation	3.791	0.687	0.063	0.133	1.444	6.000	.646
Flexibility	3.716	0.693	0.181	0.092	1.571	5.714	.654
Inhibitory Control	4.059	0.656	-0.084	-0.002	1.900	6.000	.703
Initiation	3.896	0.595	0.121	0.166	1.900	6.000	.675
Organization	3.841	0.665	-0.029	-0.287	2.000	5.700	.709
Planning	3.889	0.639	0.067	-0.096	1.909	6.000	.714
Self-Monitoring	3.857	0.629	-0.057	0.337	1.200	5.700	.634
Working Memory	4.054	0.656	0.104	-0.133	2.000	6.000	.762

Note. N = 1109; α = Cronbach's alpha.

3.1. Internal structure

We verified the CEFI's internal structure by testing three competing factorial solutions. The analyses did not fully confirm any of the tested solutions. First, the correlated trait model ran into estimation problems as the latent variable covariance matrix was not positive definite. The inspection of the latent variable correlation matrix revealed that collinearity between latent factors was a probable cause of the estimation problems. A total of 5 correlations exceeded 1 (which was over the allowed range for this parameter) and 5 further correlations were .95 or higher. Overall, 23 out of 36 pairwise correlation equaled .8 or more and 14 equaled .9 or more. This suggested that the model was misspecified and some of the factors might not be distinguishable. The bifactor model did not fit the data (X^2 (3825) = 12846.7, $p < .001$, RMSEA = .046, CFI = .658, TLI = .642, SRMR = .067). The analysis of modification indices revealed multiple correlations between subscale factors and multiple residual correlations between items with similar wording and/ or content.

3.1.1. Exploratory factor analyses

Second, the EFA did not provide a fully conclusive answer. The parallel analysis for both the average and 95th percentile criteria pointed to the nine-factor solution, since the eigenvalues for factors 1 to 9 were higher than the ones present in random data. However, the difference in eigenvalues in present and random data was small for factors 3 to 9, which suggested that in the present analysis, determining the number of factors based on the criterion of differences in eigenvalues might not be appropriate. The eigenvalue of the third factor in real data was higher by less than 1 than in random data. The difference for the ninth factor was less than 0.1. For comparison, the difference for the first and second factor equaled 15.7-15.8 and 4.2-4.4. Moreover, the nine extracted factors did not correspond to the expected ones and were difficult to interpret. Difficulties in interpretation were particularly pronounced in the case of factors 4-9 because few items loaded on them saliently. Details on the parallel analysis are available the online supplement (Figure 3s).

The scree test pointed to a two- or three-factor solution, whereas the Kaiser criterion – to a 21-factor one. Since the Kaiser criterion clearly overestimated the number of factors, we focused on the scree test and analyzed the two- and three-factor solutions with respect to their interpretability.

The three-factor solution was difficult to interpret due to multiple items loading on more than one factor. Nevertheless, the first factor showed a tendency to comprise positively worded items on meeting deadlines, completing tasks and staying focused, the second factor showed a tendency to comprise reverse worded items, whereas the third factor showed a tendency to comprise positively worded items on acting flexibly and adjusting behavior to changing circumstances. However, the pattern was not very clear due to multiple cross-loadings.

In the two-factor solution, positively worded items (items which agreeing with indicated good executive function) loaded on the first factor, whereas reverse worded items (items which agreeing with indicated poor executive function) loaded on the second factor. Further analysis of the two-factor model revealed that salient factor loadings were low to moderate; they ranged between 0.26 and 0.61 in the first factor and between 0.29 and 0.63 in the second factor. Meanwhile, only 38% and 35% of salient factor loadings, respectively, equaled 0.5 or more, which suggested multiple items were weakly related to the latent trait.

Model fit for the two- and three-factor EFA solutions was below satisfactory values (the two-factor model: X^2 (3826) = 11730.5, $p < .001$, RMSEA = .043, CFI = .700, TLI = .686, SRMR = .046; the three-factor model: X^2 (3738) = 10631.7, $p < .001$, RMSEA = .041, CFI = .739, TLI = .720, SRMR = .042). Each additional factor included in the EFA model resulted in a significantly better fit than that of the previous model. The difference in fit was particularly pronounced between the one- and two-factor models (ΔX^2 (89) = 3565.31, $p < .001$), which additionally pointed to the two-factor solution. The ratio of the first and second eigenvalues was 2.98, suggesting that the CEFI was not unidimensional. Moreover, the one-factor EFA model also had a poor fit: X^2 (3915) = 15597, $p < .001$, RMSEA = .053, CFI = .557, TLI = .547, SRMR = .072. Overall, although the EFA did not provide a conclusive answer on the number of factors, it provided further evidence against the CEFI's unidimensionality.

3.1.2. Additional analyses

Since the two-factor EFA solution suggested that the CEFI might be essentially unidimensional, with positively and reverse worded items tending to cluster, we estimated an additional bifactor model that included one general factor and two method factors representing positive and reverse wording. Each item loaded on the general factor and one of the method factors; all factors were orthogonal. The model fit was again unsatisfactory, X^2 (3826) = 11416.4, $p < .001$, RMSEA = .042, CFI = .712, TLI = .699, SRMR = .049. The analysis of modification indices revealed multiple residual correlations between items having similar wording and/or content (irrespective if they belonged to the original subscales or not). Seven such correlations equaled .3, whereas 32 equaled .2 or more.

To summarize, the results, although not fully conclusive, suggested the CEFI was two-dimensional, with two factors comprising positively and negatively worded items. Such factors are commonly found in scales comprising a mixture of the two types of items, but they rarely are of substantive nature and rather reflect a response style or are an artifact (e.g., Horan et al., 2003; Lindwall et al., 2012). As a consequence, we cautiously concluded that the two-dimensional structure was not substantive, which in turn suggested essential unidimensionality of the inventory. However, since the bifactor model that tested essential unidimensionality did not fit to the data and multiple residual correlations were present, the CEFI's unidimensional structure might be confounded by the inclusion of multiple items having similar item content and wording. Moreover, relatively weak relationship of multiple items to the latent trait might contribute to the poor fit of both the additional bifactor and the two-factor EFA model. The scree plot, additional details regarding the EFA and EFA model comparisons are available in the online supplement.

3.2. Shortened version of the inventory

Although none of the competing models fitted the data, we decided to shorten the inventory in accordance with the second goal of the study. We followed the procedure described in the "Shortening Procedure" section. The item selection process sought to maximize validity of the shortened version of the inventory while assuring the quality of the items.

In order to select 27 best performing items to be included in the shortened version of the inventory, we combined the expert ratings and the results of the item-level analysis. However, none of the items selected by the experts from the Flexibility subscale had sufficient properties in the item-level analysis. As a consequence, we decided to allow in the pool those items from the Flexibility subscale that had good properties but were indicated by only one expert. We did that to assure the construct not be overly narrowed during the item selection process. Since we assumed equal representation of items from each subscale in the shortened version (which also aimed at retaining items covering a broader range of behaviors in which executive function manifests itself), the whole selection process resulted in a pool of items of a size that allowed the inclusion of only one item from each subscale in the shortened version. Finally, the experts evaluated the selected and discarded items and confirmed that the items adequately represented executive function.

In the next step, the nine selected items were included in a unidimensional CFA model. This allowed us to verify the internal structure of the shortened inventory. The model, which comprised nine indicators representing all original subscales, had a good fit (X^2 (27) = 55.04, p < .001, RMSEA = .031, CFI = .980, TLI = .973, SRMR = .023) and showed scalar invariance across gender (scalar model tested against configural model: ΔX^2 (16) = 15.18, p = .512, n_{female} = 594, n_{male} = 515). Statistical testing indicated that the model was metric but did not exhibit scalar invariance across age when early adolescents (age 12-14, n = 613) were compared to late adolescents (age 15-19, n = 496; scalar model tested against metric model: ΔX^2 (16) = 19.59, p = .012) in the sample. However, ΔX^2 was relatively small; the test is overconservative in large samples, and changes in fit indices (RMSEA, CFI, and SRMR) did not exceed the criteria for recognizing non-invariance (Chen, 2007). Therefore, we concluded that there was enough evidence to assume scalar invariance across the two age groups. Detailed information on invariance testing is available in the online supplement.

Table 2. Standardized factor loadings in the CFA model for the shortened 9-item version of the CEFI

Item	Original Subscale	Factor Loading	Item-Rest Correlation	Average Inter-Item Correlation
CEFI_21	Attention	.689	.584	.266
CEFI_12	Emotion Regulation	.433	.387	.298
CEFI_45	Flexibility	.500	.435	.289
CEFI_32	Inhibitory Control	.489	.437	.290
CEFI_39	Initiation	.570	.490	.280
CEFI_76	Organization	.532	.464	.285
CEFI_22	Planning	.565	.486	.282
CEFI_37	Self-Monitoring	.541	.477	.283

CEFI_8	Working Memory	.483	.423	.291

Note. All loadings are statistically significant at p < .001.

The item loadings, which ranged from .433–.689, were satisfactory. An item-level analysis revealed that item-rest correlations ranged between .387 and .584, which indicated that the items had adequate discriminatory power (Nunnally & Bernstein, 1994). Moreover, inter-item correlations varied between .266 and .291, which indicated that the items were reasonably homogenous but not redundant (Nunnally & Bernstein, 1994). Higher values would suggest that the items are overly similar and most probably represent only part of the construct (Smith et al., 2000). Details are presented in Table 2. The inventory's reliability, as measured by Cronbach's α and Omega (Zinbarg et al., 2005), equaled .778 and .779, respectively. Additionally, the raw scores of the original and short versions of the CEFI correlated strongly (.871), indicating that the short version provided an adequate representation of the overall score. However, the correlation is probably overestimated because it is based on a single administration of the inventory (Smith et al., 2000).

4. Discussions

The goal of this study was to prepare a Polish adaptation of the CEFI self-report, examine its psychometric properties, and shorten it. The analyses did not confirm unidimensionality of the Polish version of the inventory. We neither fully confirmed its unidimensional structure, nor positively verified two other competing factorial solutions (multidimensional and bifactor structure). However, the results suggested a two-factor solution representing positively and reverse worded items. Moreover, a shortened nine-item version of the inventory covering all content areas included items with good properties, had satisfactory reliability, proved to be unidimensional and showed scalar invariance across gender and age. This shortened version of the CEFI self-report, if further verified in research, could be used as a participant-friendly tool for the initial assessment of executive function among adolescents for school and research purposes.

To our knowledge, this is the first study that evaluates the internal structure of the CEFI independently from the inventory's Authors. The results differ in several ways from those reported in the CEFI manual (Naglieri & Goldstein, 2013). First, the analysis presented in the manual indicates that the 90-item CEFI has a unidimensional structure, which was not fully confirmed by the present analysis, which showed that positively and reverse-worded items tended to form separate factors. We also confirmed neither the nine-factor nor the bifactor structure of the inventory. This might be due to the inadequate translation of the items; however, the CEFI translation used in this study met the standards commonly set out in scientific literature (Ercikan & Lyons-Thomas, 2013; Tyupa, 2011) and the requirements imposed by the copyright holder, who also approved the final translation. The CEFI items are short and simple, yet some of them refer to experiences and affective states which occur while the individual is performing or ceasing to perform actions. Past research has found it difficult to retain the meaning of terms that name emotions, even when using a back-translation procedure (Barger et al., 2010), because the semantic extension of terms related to emotion varies between languages based on connotations or culture-specific experiences encompassed by the term (Panayiotou, 2004).

The analysis of the item content provided another suggestion as to why the nine-factor and the bifactor models did not hold. It revealed that items with similar content appear in different subscales. For example, three similar items describe reactions in waiting situations. Two of these items (49, 74) are assigned to Inhibitory Control, whereas the third (64) is assigned to Emotion Regulation. In addition, two items (17, 88) refer to changing or maintaining a behavior despite it being ineffective, and these items appear in two different subscales: Self-Monitoring and Flexibility. In some instances, items are included in a subscale that they do not fully represent. For example, all items but one in the Initiation subscale refer to initiating various activities and actions. The remaining item (78) refers to completing activities and actions. Two additional items related to completing tasks and activities (70, 98) are included in the Inhibitory Control subscale.

On the one hand, doubts as to why a given item is assigned to a given subscale may indicate that the construct is indeed unidimensional and therefore difficult to divide into distinct subscales. This contradicts other studies indicating its multidimensionality (Ahmed et al., 2019; Anderson, 2002). However, items in questionnaires usually refer to various everyday behaviors indicative of executive function (e.g., resolving educational, logical, or social problems). Such behaviors involve several executive processes which occur simultaneously (Huizinga et al., 2018; Poon, 2018). Meanwhile, performance measures such as mazes or Towers (of London, Hanoi) allow the assessment of separate components of executive function more precisely, although their ecological validity is limited. In other words, it may be difficult to classify complex everyday behaviors, for example social problem solving or prioritization of tasks, as requiring one specific process. As a result, it is also difficult to assign a questionnaire item to a single subscale or to confirm the multidimensional structure of a questionnaire measuring executive function. If that is true, the validity subscale scores that cover different processes may be questionable unless strong support for their validity is found. On the other hand, these doubts may also indicate that the CEFI's theoretical background requires further development, with a special focus on clarifying whether its single processes can be observed separately in behavior.

However, this study findings support the unidimensionality of the shortened version of the CEFI, which corresponds with the results for the full CEFI presented in the manual (Naglieri & Goldstein, 2013). These findings suggest that executive function, when measured with items referring to complex behaviors, cannot be divided into a set of distinguishable subdimensions. This indirectly undermines the use of subscale scores.

The shortened version of the CEFI, although containing 9 instead of 27 items, is promising in terms of psychometric properties found in this study. However, it is intended to be used mostly for screening and research purposes and its reliability and validity should be further investigated. Moreover, this version is not designed to serve diagnostic purposes. Meanwhile, this study did not confirm that the Polish version of the inventory could be used for diagnosis. As a consequence, we acknowledge that the aims of this study were only partially achieved.

4.1. Limitations & future research directions

This study has several limitations. We could not verify the equivalence of the adaptation using a bilingual group design (e.g., Sireci & Berberoglu, 2000) because of the very limited share of appropriately aged bilingual students in the general population. However, the items are short, simply worded, and do not contain any idiomatic expressions, which lowers the probability of non-equivalence of meaning.

Second, although the sample included students aged 12 to 19, the majority was aged 13 to 15. Future research should ensure better representation of students in late adolescence to provide stronger evidence on the shortened inventory's properties in the whole target group. The CEFI licensing agreement does not allow modifications of the inventory; therefore, this verification process would require special legal arrangements with the copyright holder. Future research could also include the parent and teacher versions of the inventory, which were not included in this study.

Another potential goal for future inquiry is to gather additional information about the inventory's validity. Relevant information could include data from clinical samples (e.g., adolescents with conduct disorders, ADHD, or mental health disabilities) or from experimental or behavioral measures.

The present findings indicate a need for further clarification of the theoretical model of executive function. Lack of agreement related to the dimensionality of the construct, its core aspects, dimensions, or expression in behavior must be resolved to avoid hindrances in future research.

4.2. Summary

Executive function is linked to school achievement and social functioning in adolescence (e.g., Meltzer & Krishnan, 2007; Séguin & Zelazo, 2005) and its assessment is considered key in school systems (Gioia et al., 2002). However, school counselors and psychologists in Poland face a lack of cost-effective measures of executive function that would allow convenient use in individual and group settings. To fill in this gap, this study sought to prepare a Polish self-report version of the Comprehensive Executive Function Inventory and verify its psychometric properties. Moreover, it also aimed at shortening the inventory to make it more useful for screening and in situations when a full assessment is not feasible.

The analyses did not confirm any of the three hypothesized models of the Polish version of the CEFI. Moreover, since the nine-factor structure was not confirmed, the results suggest that calculating and interpreting CEFI subscale scores should be performed with caution. Evidence on the validity of subscale scores presented in the CEFI manual (Naglieri & Goldstein, 2013) is limited and our results did not provide support for their validity. We believe that using the total score is more justified as long as there is only sparse evidence that the subscale scores reflect distinguishable processes that are, at least to some extent, independent.

Meanwhile, the results suggest that the shortened version, which had satisfactory reliability, was unidimensional and showed scalar invariance across gender and age, may be a promising measurement tool. This version, if approved by the copyright holder and proved to be reliable and valid in other samples, could be used for screening purposes in the school context. Such screening may be useful because poorer executive function may not be initially recognized. It is often not until symptoms cause significant interference in academic and social functioning when a diagnosis and professional support is implemented (Ebesutani et al., 2012). Meanwhile, screening and offering early intervention to students interested in at-school support may help prevent later difficulties.

Moreover, the shortened version of the CEFI may be useful for researchers who would like to measure executive function and, at the same time, administer other scales, since a smaller number of items places less of a burden on respondents.

Although the study did not confirm the high quality of the Polish-language version of the inventory in its current form, it provides information that can be used both to improve its quality and constitutes a step in the preparation of a short respondent- and research-friendly version of the inventory.

Note: supplementary materials:
https://osf.io/sqzdh/?view_only=74a9e1a03cbc4a8b8c48e483d9b9a122

References

1. Abreu-Mendoza, R. A., Chamorro, Y., Garcia-Barrera, M. A., & Matute, E. (2018). The contributions of executive functions to mathematical learning difficulties and mathematical talent during adolescence. *PLOS ONE*, 13(12), e0209267. https://doi.org/10.1371/journal.pone.0209267
2. 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.
3. Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724.
4. Barger, B., Nabi, R., & Hong, L. Y. (2010). Standard back-translation procedures may not capture proper emotion concepts: A case study of Chinese disgust terms. *Emotion*, 10(5), 703–711. https://doi.org/10.1037/a0021453.
5. Berthelsen, D., Hayes, N., White, S. L. J., & Williams, K. E. (2017). Executive function in adolescence: Associations with child and family risk factors and self-regulation in early childhood. *Frontiers in Psychology*, 8, 903. https://doi.org/10.3389/fpsyg.2017.00903.

6. Bogen, K. (1996). The effect of questionnaire length on response rates: A review of the literature. *Proceedings of the Section on Survey Research Methods*, 1020–1025.
7. Borkowska, A., Sajewicz-Radtke, U., Lipowska, M., & Kalka, D. (2015). *Bateria diagnozy funkcji poznawczych PU1*. Pracownia Testów Psychologicznych i Pedagogicznych SEBG.
8. Brislin, R., & Freimanis, C. (2001). Back-translation: A tool for cross-cultural research. In C. Sin-wai & D. Pollard (Eds.), *An Encyclopaedia of Translation: Chinese-English, English-Chinese* (pp. 22–40). Chinese University Press.
9. Carlson, S. M. (2003). Executive function in context: Development, measurement, theory, and experience. *Monographs of the Society for Research in Child Development*, 68(3), 138–151. https://doi.org/10.1111/j.1540-5834.2003.06803012.x.
10. Cascia, J., & Barr, J. J. (2017). Associations among vocabulary, executive function skills and empathy in individuals with autism spectrum disorder. *Journal of Applied Research in Intellectual Disabilities*, 30(4), 627–637. https://doi.org/10.1111/jar.12257.
11. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. https://doi.org/10.1080/10705510701301834.
12. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
13. Climie, E. A., Cadogan, S., & Goukon, R. (2014). Test review: Comprehensive Executive Function Inventory. *Journal of Psychoeducational Assessment*, 32(2), 173–177. https://doi.org/10.1177/0734282913494169.
14. Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, 102(4), 874–888. https://doi.org/10.1037/a0027403.
15. Dawson, P. (2014). Best practices in assessing and improving executive skills. In P. Harrison & A. Thomas (Eds.). *Best practices in school psychology: Student-level* services (pp. 269–286). National Association of School Psychologists.
16. Dekker, M. C., Ziermans, T. B., Spruijt, A. M., & Swaab, H. (2017). Cognitive, parent and teacher rating measures of executive functioning: Shared and unique influences on school achievement. *Frontiers in Psychology*, 8. https://doi.org/10.3389/fpsyg.2017.00048.
17. Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D., & Young, J. (2012). The Loneliness Questionnaire–short version: An evaluation of reverse-worded and non-reverse-worded items via Item Response Theory. *Journal of Personality Assessment*, 94(4), 427–437. https://doi.org/10.1080/00223891.2012.662188
18. Ercikan, K., & Lyons-Thomas, J. (2013). Adapting tests for use in other languages and cultures. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.). *APA handbook of testing and assessment in psychology*, Vol. 3: Testing and assessment in school psychology and education. (pp. 545–569). American Psychological Association. https://doi.org/10.1037/14049-026.
19. Eslinger, P. J. (1996). Conceptualizing, describing, and measuring components of executive function: A summary. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 367–395). Brookes.
20. Fenwick, M., & McCrimmon, A. W. (2015). Test Review: Comprehensive Executive Function Inventory by J. A. Naglieri and S. Goldstein. *Canadian Journal of School Psychology*, 30(1), 64–69. https://doi.org/10.1177/0829573514546318.
21. Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360. https://doi.org/10.1093/poq/nfp031.
22. Gilbert, S. J., Bird, G., Brindley, R., Frith, C. D., & Burgess, P. W. (2008). Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: An fMRI study of two executive function tasks. *Neuropsychologia*, 46(9), 2281–2291. https://doi.org/10.1016/j.neuropsychologia.2008.03.025.
23. Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology*, 8(2), 121–137. https://doi.org/10.1076/chin.8.2.121.8727.
24. Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205. https://doi.org/10.1177/1094428104263675.

25. Hickey, A. J., & Flynn, R. J. (2019). Effects of the TutorBright tutoring programme on the reading and mathematics skills of children in foster care: A randomised controlled trial. *Oxford Review of Education*, 45(4), 519–537. https://doi.org/10.1080/03054985.2019.1607724.
26. Holmes, C. J., Kim-Spoon, J., & Deater-Deckard, K. (2016). Linking executive function and peer problems from early childhood through middle adolescence. *Journal of Abnormal Child Psychology*, 44(1), 31–42. https://doi.org/10.1007/s10802-015-0044-5.
27. Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling: A Multidisciplinary Journal*, 10(3), 435–455. https://doi.org/10.1207/S15328007SEM1003_6.
28. Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5(1), 64–86. https://doi.org/10.1037/1082-989X.5.1.64.
29. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118.
30. Huizinga, M., Baeyens, D., & Burack, J. A. (2018). Editorial: Executive function and education. *Frontiers in Psychology*, 9, 1357. https://doi.org/10.3389/fpsyg.2018.01357.
31. Jabłoński, S., Kaczmarek, I., & Kleka, P. (2018). *Test Sortowania Kart dla Dzieci (SKD)*. PTPiP.
32. Jacob, R., & Parkinson, J. (2015). The Potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*, 85(4), 512–552. https://doi.org/10.3102/0034654314561338.
33. Jaworowska, A., Matczak, A., & Fecenec, D. (2019). *IDS-2 Podręcznik 2. Przeprowadzenie badania i obliczanie wyników.* Pracownia Testów Psychologicznych.
34. Kavanaugh, B. C., Tuncer, O. F., & Wexler, B. E. (2019). Measuring and improving executive functioning in the classroom. *Journal of Cognitive Enhancement*, 3(3), 271–280. https://doi.org/10.1007/s41465-018-0095-y.
35. Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13(3), 223–248. https://doi.org/10.1080/15305058.2012.703734.
36. Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163–191. https://doi.org/10.1016/j.jsp.2006.11.005.
37. Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, 94(2), 196–204. https://doi.org/10.1080/00223891.2011.645936.
38. Lockwood, A. B., McClure, J., Sealander, K., & Baker, C. N. (2017). Measuring school psychology trainee self-efficacy. *Psychology in the Schools*, 54(6), 655–670. https://doi.org/10.1002/pits.22016.
39. Luca, C. R. D., Wood, S. J., Anderson, V., Buchanan, J.-A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the Cantab. I: Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology*, 25(2), 242–254. https://doi.org/10.1076/jcen.25.2.242.13639.
40. 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.
41. Mata, S., Gómez-Pérez, M. M., Molinero, C., & Calero, M. D. (2017). Interpersonal problem-solving skills, executive function and learning potential in preadolescents with high/ low family risk. *The Spanish Journal of Psychology*, 20, E56. https://doi.org/10.1017/sjp.2017.54.
42. Meltzer, L., & Krishnan, K. (2007). Executive function difficulties and learning disabilities. In L. Meltzer (Ed.). *Executive function in education: From theory to practice* (pp. 77–105). The Guilford Press.
43. Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. https://doi.org/10.1006/cogp.1999.0734.
44. Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide. Eighth edition*. Muthén & Muthén.
45. Naglieri, J. A., & Goldstein, S. (2013). *Comprehensive Executive Function Inventory (CEFI)*. Manual. Mental-Health Systems Inc.

46. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. McGraw-Hill.
47. Otero, T. M., Barker, L. A., & Naglieri, J. A. (2014). Executive function treatment and intervention in schools. Applied Neuropsychology: *Child*, 3(3), 205–214. https://doi.org/10.1080/21622965.2014.897903.
48. Panayiotou, A. (2004). Switching codes, switching code: Bilinguals' emotional responses in English and Greek. *Journal of Multilingual and Multicultural Development*, 25(2–3), 124–139. https://doi.org/10.1080/01434630408666525.
49. Poon, K. (2018). Hot and cool executive functions in adolescence: Development and contributions to important developmental outcomes. *Frontiers in Psychology*, 8, 2311. https://doi.org/10.3389/fpsyg.2017.02311.
50. Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559. https://doi.org/10.1080/00223891.2010.496477.
51. Sălăgean, C., Stan, C. & Crișan, C. (2021). Linguistic adaptation and Romanian validation of the "teachers' self-efficacy for Autism Spectrum Disorder Inclusion" Scale (TSE-ASDI). *Journal of Educational Sciences and Psychology*, 11(73), 123-134. https://doi.org/10.51865/JESP.2021.1.11.
52. Selemon, L. D. (2013). A role for synaptic plasticity in the adolescent development of executive function. *Translational Psychiatry*, 3(3), e238–e238. https://doi.org/10.1038/tp.2013.7.
53. Séguin, J. R., & Zelazo, P. D. (2005). Executive function in early physical aggression. In R. E. Tremblay, W. W. Hartup, & J. Archer (Eds.). *Developmental origins of aggression* (pp. 307–329). The Guilford Press.
54. Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, 79(2), 310–334. https://doi.org/10.1177/0013164418783530.
55. Sireci, S. G., & Berberoglu, G. (2000). Using bilingual respondents to evaluate translated-adapted Items. *Applied Measurement in Education*, 13(3), 229–248. https://doi.org/10.1207/S15324818AME1303_1.
56. Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102.
57. Sotelo-Dynega, M. (2017). Clinical reasoning in the assessment and intervention planning for a reading disability. *Canadian Journal of School Psychology*, 32(1), 87–100. https://doi.org/10.1177/0829573516659936.
58. Stevanovic, D., Brajkovic, L., Srivastava, M. K., Krgovic, I., & Jancic, J. (2018). Widespread cortical PET abnormalities in an adolescent related to a PNES dissociative state, PTSD, ADHD, and domestic violence exposure. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, 6(2), 98–106. https://doi.org/10.21307/sjcapp-2018-011.
59. Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2009). Executive functions: Performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 15(1), 53–72. https://doi.org/10.1080/09297040802070929.
60. Tripon, C. (2021). Helping teenagers reconnect to learning – the power of supportive teachers to design autonomy. *Journal of Educational Sciences and Psychology*, 11(1), 11-25. https://doi.org/10.51865/JESP.2021.1.03.
61. Tyupa, S. (2011). A theoretical framework for back-translation as a quality assessment tool. *New Voices in Translation Studies*, 7(1), 35–46.
62. Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. https://doi.org/10.1177/109442810031002.
63. Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. A. Nelson & M. Luciana (Eds.). *Developmental cognitive neuroscience. Handbook of developmental cognitive neuroscience* (pp. 553–574). MIT Press.
64. Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, 1(2), 198–226. https://doi.org/10.1037/1089-2680.1.2.198.