Article

Ten Items to Find Them All: Shortening Scales for the Screening of Executive Function in Children With Attention Deficit/Hyperactivity Disorder

Hilario Blasco-Fontecilla[1,2,3,*]
Marcos Bella-Fernández[4,5]

[1] Instituto de Transferencia e Investigación (ITEI), Universidad Internacional de La Rioja (UNIR), 26006 Logroño, Spain
[2] Department of Psychiatry, Emooti, 28010 Madrid, Spain
[3] Center of Biomedical Network Research on Mental Health (CIBERSAM), 28029 Madrid, Spain
[4] Faculty of Health, Universidad Internacional de La Empresa (UNIE), 28015 Madrid, Spain
[5] Department of Psychology, Universidad Pontificia de Comillas, 28108 Alcobendas, Spain

Abstract

Background: Attention Deficit/Hyperactivity Disorder (ADHD) is the most common neurodevelopmental disorder. The affectation of executive functions is stressed in the most recent research on ADHD, and many tests are used to assess it in ADHD, but they are usually time- and effort-consuming.

Methods: From a database of a total of 222 children with ADHD, 59 of them suffering executive dysfunction, we took the most widely used tests for executive functions in ADHD (Behavior Rating Inventory of Executive Function (BRIEF), Swanson, Noland, and Pelham (SNAP)-IV, and Conners' Parent Rating Scale (CPRS-R)) and applied several methods of test shortening: Item-total correlations from the Classical Test Theory, factor analysis and their subsequent factor loadings, elastic nets, and the Graded Response Model from the Item Response Theory models. Using the parameters or indicators provided by each of these methods, we selected the most discriminative items to develop a brief screening tool.

Results: Our results show that different selection methods select different items. More importantly, we found that the shortened tests obtained this way are in general capable of discriminating between patients with and without ADHD. More precisely, all the shortened tests show high sensitivity, but relatively low specificity.

Conclusions: Shortened tests can be used for screening purposes without having to administer full test versions.

Keywords

test shortening; screening; executive dysfunction; ADHD

Introduction

Attention Deficit/Hyperactivity Disorder (ADHD) is the most common neurodevelopmental disorder worldwide [1]. Suffering ADHD has many negative effects in daily life, such as difficulty in adapting to a variety of tasks and jobs, additional risks of accidents and substance abuse, or weaker interpersonal relationships. ADHD diagnosis is clinical, being to date no definitive tests for assessing ADHD. The diagnosis of ADHD is made after taking a medical history and performing a clinical assessment, which typically includes interviews with the patient and their relatives, as well as psychological tests, all of which are examined by a psychiatrist or a neurologist. In the last years, emphasis has been placed on the affectation of executive function in ADHD patients [2–4], and some useful tests have been developed or adapted to help therapists correctly diagnose ADHD, such as the Behavior Rating Inven-

Submitted: 19 November 2024 Revised: 3 March 2025 Accepted: 13 March 2025 Published: 5 May 2025
*Corresponding author details: Hilario Blasco-Fontecilla, Instituto de Transferencia e Investigación (ITEI), Universidad Internacional de La Rioja (UNIR), 26006 Logroño, Spain; Department of Psychiatry, Emooti, 28010 Madrid, Spain; Center of Biomedical Network Research on Mental Health (CIBERSAM), 28029 Madrid, Spain. Email: hilariomanuel.blasco@unir.net

tory for Executive Function (BRIEF) inventory [5] or Conners' Continuous Performance Test (CPT) [6]. The time and effort required to complete the tests may be challenging for many ADHD potential patients, characterized by lower levels of intrinsic motivation and a greater ease of becoming bored [2]. For instance, CPT lasts 14 minutes, whereas the BRIEF test takes 10 to 15 minutes [7,8], and these time lapses may be too exigent for ADHD children.

A variety of methods have been developed to develop brief versions of psychometric tests. These brief versions often consist of a selection of items from the test based on one or more criteria. Several approaches, including psychometric indicators [9] and some complex metaheuristics and algorithms [10,11], have been proposed and used to make these item selections. Applying this diversity of item selection methods results in heterogeneous item sets with comparable properties, even from the same tests and the same database [9,10]. In particular, for the case of ADHD diagnosis, to date, no brief, comprehensive tests have been developed. The most used tests take too much time, such as the BRIEF [5] or the CPT [6], or only contain items for inattention and hyperactivity, such as the Swanson, Noland, and Pelham (SNAP)-IV scale [12] or the Conners' Parent Rating Scale (CPRS-R) [13].

The aim of this work is to develop and assess a brief test from the items of the most common screening tests for ADHD. We hypothesize that it is possible to reduce the number of items without losing significantly the capability to discriminate between patients with and without affectation on executive function. Our goal is to develop a short test, taking no more than 10 items, which preserves the original form's discriminative power.

Methods

Participants

A total of 222 participants were recruited for the present study. The inclusion criteria were: being admitted to the Child and Adolescent Psychiatry Service at the Hospital Universitario Puerta de Hierro Majadahonda, Spain; being under 18 years; having received a diagnosis of ADHD; having been assessed on executive dysfunction. Ages varied from 3 to 17 (mean = 11.43 years, standard deviation = 3.49). All the participants were assessed by an expert child and adolescent psychiatrist. 59 patients from the sample were diagnosed with executive dysfunction. All the participants signed an adapted reported consent, and their parents or legal guardians signed a written informed consent, before participating in the study. The study was previously approved by the Hospital Universitario Puerta de Hierro Majadahonda Ethics Committee (code 15.17).

Measurements

BRIEF

The BRIEF inventory [5] is a well-known assessment inventory for executive function. It has been widely used in samples of ADHD patients [14,15]. It has adequate psychometric properties [5,7,8]. It consists of 86 items and takes up to 15 minutes to complete. For every item, 3 response options are available: never, sometimes, and often, depending on the frequency of the events described on the items. The inventory contains 8 clinical scales, converging in two second-order factors (behavioral regulation index and metacognition index) and a global score.

SNAP-IV

The SNAP-IV test [12] is a short test for assessing ADHD and its two subtypes, inattention and hyperactivity/impulsivity. It consists of 18 items scored on a 4-point Likert scale, 9 items to assess inattention, and 9 items to assess hyperactivity/impulsivity. It is one of the most widely used scales for ADHD assessment and it shows acceptable psychometric properties [16]. Its adaptation to Spanish [17] also has good psychometrical properties.

CPRS-R

The CPRS-R [13] is a useful, well-validated screening test for ADHD in children. It consists of 10 items with a 4-point Likert scale. CPRS-R has acceptable psychometrical properties and good validity for screening for ADHD patients [13,18].

Statistical Analysis

We performed a variety of methods proposed for shortening tests. In all cases, the predictors include the 86 items from the BRIEF inventory, the 18 items from the SPAN-IV test, and the 10 items from the CPRS-R test.

These methods are:

Item-test correlations [9]. This is one of the most common discrimination measures in Classical Test Theory. In our case, because we are taking the items from several tests, we cannot simply use the total score from one of them or correlate each item with the total score on the test that item belongs to. Instead, we calculated the sum of all items and correlated each item with that total score. We then selected the items with the largest correlations.

Factor loadings from Confirmatory Factor Analysis (CFA) [9,10]. Kleka & Paluchowski [9] used an exploratory factor analysis and took factor loadings from it. In contrast, Schroeders *et al*. [10] used what they named "stepwise confirmatory factor analysis": they iteratively performed a confirmatory factor analysis by removing the item with the smallest factor loading after each factor analysis until the desired number of items remained. We simply fitted a confirmatory factor model with all the items loading in one single factor and then selected the items with the largest loadings.

Regression through elastic nets [11]. Kleka & Paluchowski [9] and Artieda-Urrutia *et al*. [19], among others, used logistic or linear regression coefficients, but when predictors are highly correlated, such as items from the same inventory, classical regression models are less adequate. In these cases, elastic nets are a plausible alternative. It has been successfully used to select items from several tests for a brief screening tool [20]. For every trial, the sample was randomly split into two subsamples; the first subsample served to train the elastic net and obtain the parameter estimations, and the second subsample was devoted to regression coefficient estimations. We performed 100 iterations and summed the 100 regression coefficients for each item. Then, we selected the items with the largest sum of coefficients. Packages "glmnet" [21] and "caret" [22] were used for parameter estimation.

Graded Response Model [23] from Item Response Theory, using "mirt" package [24] for R. For each item, the Graded Response Model estimates as many parameters as the number of response options, plus a discriminability parameter, a. The higher the parameter a, the higher the item discriminability. We are taking the items with the largest a values.

We discarded using metaheuristics like Genetic Algorithms [10] and Ant Colony Optimization [25] due to the large sample sizes required.

Once the coefficients or parameters were obtained, we performed Receiver Operating Characteristic (ROC) curves to estimate the accuracy, sensitivity, and specificity of the items selected for each procedure to discriminate between patients with and without executive dysfunction. ROC curves are statistical tools widely used to assess the performance of a given classifier against a criterion, which is the variable that the classifier should predict; in this case, each one of the shortened tests obtained acts as a separate classifier, and the criterion is the diagnosis of executive dysfunction. ROC curves provide an estimation of the optimal classifier threshold and three measures of prediction quality: sensitivity, specificity, and accuracy. Applied to this case, the sensitivity is the ratio between the cases of executive dysfunction correctly detected by the classifier and the total number of cases in the sample (detected or undetected); a sensitivity of 0 means that the classifier does not detect any case of executive dysfunction, and a sensitivity of 1 means that the classifier detects all of them. Analogously, the specificity is the ratio between the non-patients of executive dysfunction correctly labeled and all the non-patients (either correctly labeled as such or falsely detected as patients); again, a specificity of 0 means that the classifier does not detect any non-patient, and a specificity of 1 means that the classifier correctly labels all non-patients as such. Last, the accuracy is the ratio between all the individuals correctly labeled and the whole sample. We used the "pROC" package [26] for R to calculate the ROC curves corresponding to each of the shortened tests.

Last, to assess structural validity, we performed parallel analyses on each of the four shortened tests to determine the number of factors to extract, and then we performed exploratory factor analysis with "promax" as the rotation method. We used the package "paran" [27] for the parallel analysis.

For all the statistical analyses described above, we used R, version 4.3.2. (The R Core Foundation, Vienna, Austria) [28].

Results

Fig. 1 summarizes the method process from the original tests to the four shortened versions.

Table 1 shows the parameters or coefficients obtained from each procedure. Only one item, Item 7 from CPRS-R, was selected in the four procedures. Another item, Item 13 from SNAP-IV, was selected in three of the four procedures. The content of the selected items is described in the **Supplementary Material**.

Table 2 shows the accuracy, sensitivity, and specificity for each item selection procedure and the total item bank. In general, sensitivities are high, and specificities are low, which means that these tools are generally able to detect cases of executive dysfunction, but they also tend to falsely detect patients without it. Accuracy values, as global mea-

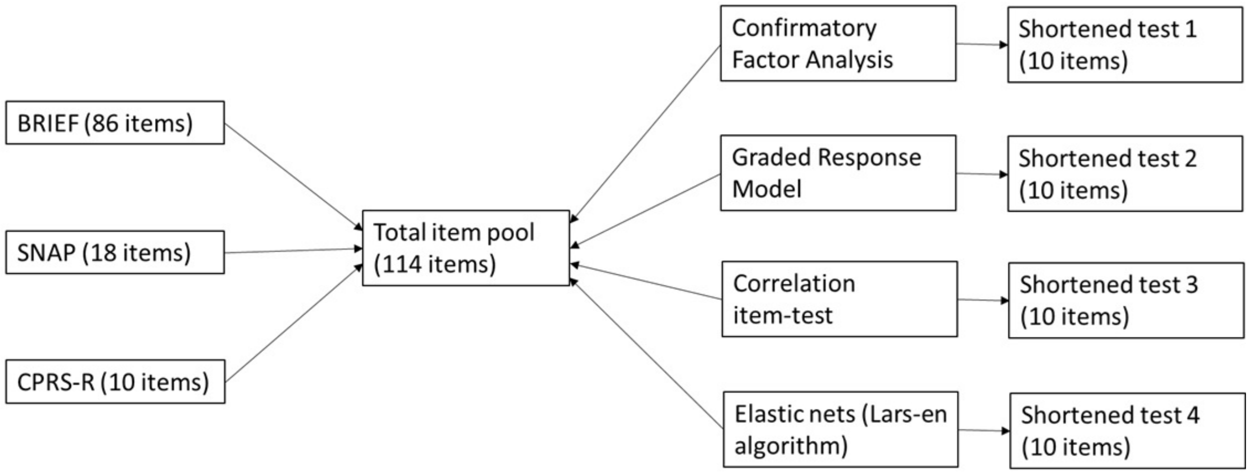

Fig. 1. Flowchart summarizing the process from the three original tests to the four shortened versions. BRIEF, Behavior Rating Inventory of Executive Function; SNAP, Swanson, Noland, and Pelham; CPRS-R, Conners' Parent Rating Scale.

sures of classification performance, are low, likely because the sample has many more controls (patients without executive dysfunction) than patients, and thus the false positives overweight the correct positives in the global accuracy measure.

We then divided the sample into two subsamples of patients with inattentive (n = 185) or mixed (n = 414) ADHD. We then also assessed the accuracy, sensitivity, and specificity of the shortened tests in the two subsamples. Table 3 shows the result of this differentiated assessment. In general, we can see that the shortened versions obtained through CFA, Graded Response Model(GRM), and item-test correlations show better performance in patients with inattentive ADHD than in patients with mixed ADHD. The shortened version obtained through CFA shows a large sensitivity in patients with inattentive ADHD and an acceptable specificity in patients with mixed ADHD. The shortened tests obtained through GRM and item-test correlation show acceptable sensitivities in patients with inattentive ADHD.

Reliability and Content Validity

Confirmatory Factor Analysis

The items selected through factor loadings from CFA were: items 3, 6, 9, 13, 16, and 18 from SNAP-IV, items 6 and 7 from CPRS-R, and items 27 and 58 from BRIEF. The factor loadings selected ranged from 1.44 and 1.13. The internal consistency of the shortened test is high (alpha = 0.85, omega = 0.85).

The items refer to impulsivity ("He or she tends to talk excessively"), distractibility ("He or she forgets daily activities", "He or she is easily distracted"), and difficulties in planning and persevering without external motivation ("He or she has difficulties to persevere in necessary actions to reach a certain goal, such as saving money to buy a special item or studying to get good grades").

Graded Response Model From Item Response Theory

The items selected were: Items 17, 27, 37, 42, 65, and 79 from the BRIEF, items 13 and 18 from the SNAP test, and items 6 and 7 from the CPRS-R test. Their discriminability parameters ranged from 1.81 to 2.27. The internal consistency of the shortened test is high (alpha = 0.88, omega = 0.88).

Similarly to the CFA, the items selected through GRM refer to impulsivity and lack of inhibition, distractibility, and planning difficulties, but there is also one item referred to inability to detect negative reactions to his/her behavior.

Correlation Item-Test

The items selected were: items 37, 42, 65, and 79 from BRIEF, the items 2, 6, 8, 9, and 13 from SNAP test, and item 7 from CPRS-R. The item-total correlations of the selected items ranged from 0.667 to 0.610. The internal consistency of the shortened test is high (alpha = 0.89, omega = 0.89).

The items selected are also referred to impulsivity and lack of inhibition, distractibility, planning difficulties, and inability to detect negative reactions.

Table 1. Coefficients or parameters obtained.

Item	Factor loading from CFA	Discrimination (a) parameters from GRM	Item-total correlations	Regression coefficients from elastic nets
SNAP 1	1.000	1.361	0.560	0.258
SNAP 2	1.113	1.671	**0.610**	0.093
SNAP 3	**1.146**	1.503	0.599	0.029
SNAP 4	1.123	1.456	0.542	0.086
SNAP 5	1.096	1.312	0.535	0.093
SNAP 6	**1.188**	1.510	**0.610**	0.002
SNAP 7	1.087	1.308	0.491	0.040
SNAP 8	1.117	1.654	**0.641**	0.038
SNAP 9	**1.256**	1.619	**0.620**	0.077
SNAP 10	0.816	1.077	0.454	0.119
SNAP 11	1.065	1.553	0.537	0.026
SNAP 12	1.067	1.488	0.556	0.045
SNAP 13	**1.144**	**2.044**	**0.651**	0.035
SNAP 14	0.881	1.205	0.487	0.006
SNAP 15	0.972	1.114	0.493	**0.337**
SNAP 16	**1.440**	1.446	0.412	0.121
SNAP 17	1.040	1.557	0.566	0.036
SNAP 18	**1.184**	**1.808**	0.514	0.182
CPRS-R 1	0.865	1.180	0.396	0.013
CPRS-R 2	0.527	0.752	0.363	0.051
CPRS-R 3	1.060	1.479	0.577	0.055
CPRS-R 4	0.898	1.421	0.504	0.141
CPRS-R 5	0.826	1.301	0.467	0.125
CPRS-R 6	**1.168**	**1.698**	0.604	0.199
CPRS-R 7	**1.220**	**1.751**	**0.634**	**0.901**
CPRS-R 8	0.685	1.000	0.450	0.026
CPRS-R 9	0.830	1.351	0.526	0.073
CPRS-R 10	0.864	1.401	0.499	0.019
BRIEF 1	0.385	0.742	0.399	0.224
BRIEF 2	0.752	1.350	0.517	0.033
BRIEF 3	0.361	0.419	0.243	0.125
BRIEF 4	0.491	0.901	0.409	0.306
BRIEF 5	0.625	1.044	0.279	0.030
BRIEF 6	0.542	0.984	0.455	0.143
BRIEF 7	0.400	0.852	0.406	0.064
BRIEF 8	0.466	0.801	0.374	0.125
BRIEF 9	0.816	1.448	0.555	0.015
BRIEF 10	0.771	1.398	0.562	0.026
BRIEF 11	0.837	1.423	0.515	0.022
BRIEF 12	0.456	0.788	0.422	0.215
BRIEF 13	0.407	0.849	0.369	0.077
BRIEF 14	0.693	1.492	0.526	0.073
BRIEF 15	0.597	0.868	0.401	0.156
BRIEF 16	0.482	0.759	0.371	0.038
BRIEF 17	0.735	**1.757**	0.559	0.039
BRIEF 18	0.651	1.189	0.438	0.022
BRIEF 19	0.737	1.627	0.549	0.085
BRIEF 20	0.221	0.573	0.252	0.229
BRIEF 21	0.607	1.411	0.521	0.041

Table 1. Continued.

Item	Factor loading from CFA	Discrimination (a) parameters from GRM	Item-total correlations	Regression coefficients from elastic nets
BRIEF 22	0.759	1.147	0.479	0.192
BRIEF 23	0.459	0.770	0.404	**0.415**
BRIEF 24	0.839	1.594	0.582	0.097
BRIEF 25	0.476	1.059	0.454	0.014
BRIEF 26	0.342	0.756	0.352	0.036
BRIEF 27	**1.249**	**1.690**	0.490	0.242
BRIEF 28	0.672	1.201	0.490	0.203
BRIEF 29	0.309	0.569	0.246	0.308
BRIEF 30	0.327	0.538	0.288	0.036
BRIEF 31	0.862	0.749	0.361	0.069
BRIEF 32	0.734	1.281	0.505	0.049
BRIEF 33	0.645	1.058	0.461	0.118
BRIEF 34	0.737	1.491	0.565	0.172
BRIEF 35	0.703	1.416	0.526	0.015
BRIEF 36	0.646	1.502	0.515	0.031
BRIEF 37	0.915	**2.085**	**0.624**	0.035
BRIEF 38	0.790	1.485	0.572	0.160
BRIEF 39	0.374	0.675	0.336	**0.328**
BRIEF 40	0.608	1.142	0.474	**0.513**
BRIEF 41	0.836	1.656	0.604	0.044
BRIEF 42	0.850	**1.872**	**0.615**	0.023
BRIEF 43	0.685	1.472	0.582	0.095
BRIEF 44	0.685	1.560	0.566	0.068
BRIEF 45	0.605	1.247	0.516	0.038
BRIEF 46	0.776	1.326	0.521	0.078
BRIEF 47	0.783	1.451	0.545	0.142
BRIEF 48	0.726	1.001	0.297	0.008
BRIEF 49	0.674	1.098	0.483	0.026
BRIEF 50	0.759	1.030	0.398	0.046
BRIEF 51	0.886	1.420	0.579	0.052
BRIEF 52	0.661	1.255	0.501	0.066
BRIEF 53	0.796	1.378	0.505	0.041
BRIEF 54	0.478	1.173	0.471	0.083
BRIEF 55	0.686	1.516	0.594	0.013
BRIEF 56	0.645	1.318	0.460	0.013
BRIEF 57	0.753	1.222	0.522	0.260
BRIEF 58	**1.134**	1.482	0.459	0.068
BRIEF 59	0.590	1.206	0.483	0.131
BRIEF 60	0.831	1.463	0.527	**0.314**
BRIEF 61	0.504	0.722	0.348	0.026
BRIEF 62	0.435	1.069	0.415	**0.339**
BRIEF 63	0.708	1.488	0.534	0.009
BRIEF 64	0.476	1.017	0.451	0.137
BRIEF 65	0.887	**1.921**	**0.661**	0.028
BRIEF 66	0.710	1.345	0.533	0.064
BRIEF 67	0.633	1.053	0.474	0.055
BRIEF 68	0.701	1.221	0.525	**0.318**
BRIEF 69	0.618	0.960	0.233	0.089
BRIEF 70	0.433	0.925	0.419	0.086

Table 1. Continued.

Item	Factor loading from CFA	Discrimination (a) parameters from GRM	Item-total correlations	Regression coefficients from elastic nets
BRIEF 71	0.219	0.253	0.201	0.126
BRIEF 72	0.456	0.702	0.351	0.194
BRIEF 73	0.724	1.513	0.539	0.037
BRIEF 74	0.740	1.437	0.507	0.091
BRIEF 75	0.677	1.208	0.517	0.276
BRIEF 76	0.647	0.820	0.373	0.048
BRIEF 77	0.878	1.606	0.587	0.041
BRIEF 78	0.781	1.399	0.511	0.080
BRIEF 79	0.866	**2.056**	**0.667**	0.079
BRIEF 80	0.751	1.398	0.556	0.064
BRIEF 81	0.581	1.053	0.463	0.112
BRIEF 82	0.747	1.434	0.577	0.199
BRIEF 83	0.789	1.475	0.588	0.214
BRIEF 84	0.699	1.385	0.559	0.043
BRIEF 85	0.783	1.235	0.575	**0.342**
BRIEF 86	0.826	1.232	0.531	**0.482**

The ten best values for each procedure are highlighted in bold. CFA, Confirmatory Factor Analysis; GRM, Graded Response Model.

Table 2. Discriminability measures of the items selected through each procedure.

	Area under ROC curve	Accuracy	Sensitivity	Specificity
CFA	0.6214	0.460	0.922	0.277
GRM	0.6274	0.507	0.839	0.369
Item-total correlation	0.6254	0.522	0.763	0.423
Elastic nets	0.6142	0.516	0.777	0.407
All items	0.5854	0.482	0.830	0.356

ROC, Receiver Operating Characteristic.

Elastic Net

The items selected were the item 15 from the SNAP-IV, the item 7 from the CPRS-R, and the items 23, 39, 40, 60, 62, 68, 85, and 86 from the BRIEF. The aggregated coefficients ranged from 0.901 to 0.314. The internal consistency of the shortened test is acceptable (alpha = 0.69, omega = 0.70).

The items selected refer to distractibility, planning difficulties, cognitive rigidity, emotional dysregulation, lack of inhibition, and impulsivity.

Structural Validity

To assess the structural validity, we performed parallel analyses to extract the number of dimensions and exploratory factor analyses to obtain the factor loadings from each item. Tables 4,5,6,7 show the factor loadings for each shortened test; only factor loadings above 0.3 are included. The number of factors extracted from each shortened test is determined by the result from the respective parallel analyses.

Factor 2 contains 2 items related to daily and time planification. Factor 1 includes items related to impulsivity, distractibility, and perseverance.

In this case, all the items load to a common factor including items of inattention and executive functions.

In this case, items from Factor 1 relate to distractibility and sustained attention, while items from Factor 2 are related to impulsivity and perseverance.

In this case, Factor 1 contains items related to talking too much and regarding closed topics, as well as one item pertaining to time estimation; Factor 2 contains items related to cognitive flexibility, emotional regulation, and detail orientation; and Factor 3 contains items pertaining to distractibility and difficulties to keep routines. In general,

Table 3. Discriminability measures of the shortened tests in patients with inattentive and mixed ADHD.

	Area under ROC curve	Accuracy	Sensitivity	Specificity
Inattentive ADHD				
CFA	0.681	0.578	0.917	0.400
GRM	0.703	0.671	0.731	0.632
Item-total correlation	0.675	0.630	0.731	0.574
Elastic nets	0.574	0.575	0.577	0.574
Mixed ADHD				
CFA	0.602	0.638	0.478	0.709
GRM	0.600	0.605	0.543	0.634
Item-total correlation	0.599	0.642	0.543	0.686
Elastic nets	0.591	0.633	0.522	0.683

ADHD, Attention Deficit/Hyperactivity Disorder.

Table 4. Factor loadings for the shortened test from CFA.

	Factor 1	Factor 2
SNAP 16	0.405	
SNAP 9		1.000
BRIEF 27	0.745	
CPRS-R 7	0.774	
SNAP 6	0.748	
SNAP 18	0.673	
CPRS-R 6	0.479	
SNAP 3	0.568	
SNAP 13	0.697	
BRIEF 58		0.831

Table 5. Factor loadings for shortened tests from GRM.

	Factor 1
BRIEF 37	0.733
BRIEF 79	0.616
SNAP 13	0.551
BRIEF 65	0.806
BRIEF 42	0.454
SNAP 18	0.692
BRIEF 17	0.726
CPRS-R 7	0.742
CPRS-R 6	0.536
BRIEF 27	0.723

Table 6. Factor loadings for the shortened test from item-total correlations.

	Factor 1	Factor 2
BRIEF 79		0.575
BRIEF 65		0.627
SNAP 13		0.872
SNAP 8	0.753	
CPRS-R 7	0.975	
BRIEF 37	0.662	
SNAP 9	0.721	
BRIEF 42		0.653
SNAP 6	0.633	
SNAP 2		0.914

Table 7. Factor loadings for the shortened test from elastic nets.

	Factor 1	Factor 2	Factor 3
CPRS-R 7			0.363
BRIEF 40	0.567		
BRIEF 86			0.525
BRIEF 23		0.744	
BRIEF 85	0.396		
BRIEF 62		0.603	
SNAP 15	0.685		
BRIEF 39	0.439		
BRIEF 68			0.399
BRIEF 60	0.323	0.347	

factor structures change in the shortened tests related to the original tests.

Discussion

We used four procedures to shorten a battery of the most widely used tests to assess executive dysfunction in children with ADHD. Our results show that, in general terms, the four collections of items show acceptable psychometric and discriminant properties, despite containing substantially different items; only one item, item 7 from the CPRS-R test, is on the four item collections. In particular, the elastic nets tended to select different items than the three other procedures. This paradoxical effect of similar properties despite differences in selected items is consistent with other studies on test shortening [9].

The four item selection methods obtained generally large sensitivities (above or around 0.7) and low specificities. These results were particularly good in patients with inattentive ADHD, but none of these tests obtained good sensitivities or specificities in patients with mixed ADHD. In this context, these results mean that the four shortened tests have in general adequate properties to detect cases of executive dysfunction among patients with ADHD, especially with inattentive ADHD, leaving few cases undetected. However, they are also prone to falsely detect cases of executive dysfunction in patients who do not suffer from it. These results provide evidence of the adequacy of using these shortened tests as screening tools, but not as diagnostic tools. Rather, in the cases where any of these tests detect a possible case of executive dysfunction, a deeper assessment should be performed before diagnosing a patient with executive dysfunction.

Shortening tests usually has a negative impact on reliability. Shortened tests tend to have lower reliability than original tests [29]. In our case, the shortened versions obtained from CFA, GRM, and item-total correlations show large reliability coefficients, while the shortened version obtained through elastic nets has an acceptable internal consistency. This difference is not surprising: selecting the best item-total correlations is a good manner to optimize internal consistency [30], and the same logic might be applied to selecting the items with the largest factor loadings. On the contrary, GRM and elastic nets are based on discriminability. Despite this, GRM seems to outperform elastic nets in this regard.

Another common objection regarding shortened tests is the lack of validity that shortened versions suffer compared with the original tests [29,31], and our case is not an exception. As suggested by several authors [29,31,32], we assessed the validity of the shortened versions in three ways: content validity, examining the theoretical content of the selected items, structure validity, assessing the factor structure of the shortened versions, and criterion validity in the sense of the shortened tests' accuracy discriminating between individuals with and without executive dysfunction. We obtained substantial variations from the original contents and factor structure. The shortened tests only showed good sensitivity in detecting cases of executive dysfunction among patients with ADHD. Any other use for these shortened tests should be avoided without careful validity analysis.

The four procedures selected items which were related to impulsivity, lack of inhibition, and difficulties in planning and reaching long-term goals. GRM and item-total correlation also selected an item related to a lack of detection of negative reactions. The elastic nets included a wider variety of executive functions: apart from the features mentioned above, this method also selected items related to cognitive flexibility and emotional regulation. Recently, another study [33] used machine learning techniques to shorten the BRIEF-2 and found that Lasso algorithm (our elastic net) performed the best and selected items to predict ADHD. In our case, the goal was slightly different (to assess executive dysfunction in samples of patients already diagnosed with ADHD) and we fixed the number of items to be selected, whereas their study allowed a variable number of selected items.

The main limitation of this study is the potential use of the shortened tests obtained. Our goal was to obtain screening tests to detect executive dysfunction in children diagnosed with ADHD. Although this specific goal was achieved satisfactorily, especially in patients with inattentive ADHD, the tests obtained here showed that their usefulness beyond this application is scarce. In this regard, future research should cover the feasibility of these shortened test versions, focusing on completion times, and the experience from both patients and practitioners. Furthermore, our sample consisted of 222 patients with ADHD, 59 patients who also suffered executive dysfunction (cases), and 163 patients without executive dysfunction (controls). This relatively small sample did not allow us to use some meta-heuristic methods, such as Genetic Algorithms and Ant Colony Optimization, which could provide different shortened tests. Moreover, the sample size of this study highlights the need to generalize and replicate these findings in other samples of patients with ADHD.

Conclusions

We obtained four shortened tests by applying four different shortening methods to the same database. The four shortened tests obtained in this work are adequate to be used for executive dysfunction screening purposes in the context of child psychiatry and psychology. Their usefulness as a screening tool for executive dysfunction was assessed through ROC curves. Content validity assessments reveal that the content loss due to the test shortening strongly discourages their use for other purposes than executive dysfunction screening in patients with ADHD. Further research and analyses are required to assess their usability beyond the scope of this work.

Availability of Data and Materials

The datasets generated and/or analyzed during the current study are not publicly available due to confidentiality issues but are available from the corresponding author on reasonable request.

Author Contributions

Conceptualization: HBF; Methodology: HBF and MBF; Formal analysis: MBF; Data collection: HBF; Data curation: MBF; Writing: Original Draft: MBF; Writing: Review and Editing: HBF; Supervision: HBF. Both authors contributed to the drafting or important editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of the Hospital Universitario Puerta de Hierro Majadahonda (code 15.17). All the participants signed an adapted reported consent, and their parents or legal guardians signed a written informed consent, before participating in the study. This study also abides by the Declaration of Helsinki.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

HBF has received lecture fees from Takeda Pharmaceuticals and the laboratories BIAL, Rubio, and Rovi. He has also been granted 3 prizes for developing the serious video game "The Secret Trail of Moon" to treat attention-deficit/hyperactivity disorder: the Shibuya Prize by Takeda, first prize from the College of Psychologists of Madrid, and a prize for the best innovative health initiative within the Health Start campaign. He is the principal investigator of predoctoral contracts for training in health research (IFI16/00039); the coprincipal investigator of a Ministry of Economy, Trade, and Enterprise research grant (RTI2018-101857-B-I00); and the principal investigator of Sincronia and Discronia research projects, funded by the Start-up Bitsphi. Moreover, he is the recipient of (1) a Foundation for Innovation and Foresight in Health in Spain grant and (2) a Puerta de Hierro Segovia de Arana Institute of Health Research intensification grant and is involved in 2 clinical trials (Mensia Koala, Newrofeed study; ESKET-SUI2002). He is also a cofounder of Haglaia Solutions (www.haglaia.com). HBF is serving as one of the Guest Editors of this journal. We declare that HBF had no involvement in the peer review of this article and has no access to information regarding its peer review. MBF has no conflicts of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.62641/aep.v53i3.1883.

References

[1] Francés L, Quintero J, Fernández A, Ruiz A, Caules J, Fillon G, *et al*. Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: a systematic review in accordance with the PRISMA criteria. Child and Adolescent Psychiatry and Mental Health. 2022; 16: 27. https://doi.org/10.1186/s13034-022-00462-1.

[2] Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychological Bulletin. 1997; 121: 65–94. https://doi.org/10.1037/0033-2909.121.1.65.

[3] Castellanos FX, Sonuga-Barke EJS, Milham MP, Tannock R. Characterizing cognition in ADHD: beyond executive dysfunction. Trends in Cognitive Sciences. 2006; 10: 117–123. https://doi.org/10.1016/j.tics.2006.01.011.

[4] Boonstra AM, Oosterlaan J, Sergeant JA, Buitelaar JK. Executive functioning in adult ADHD: a meta-analytic review. Psychological Medicine. 2005; 35: 1097–1108. https://doi.org/10.1017/s003329170500499x.

[5] Gioia GA. Behavior Rating Inventory of Executive Function. Psychological Assessment Resources, Inc.: Lutz, Florida, USA. 2000.

[6] Conners CK. Conners' Continuous Performance Test. 3rd edn. Multi-Health Systems: North Tonawanda, NY, USA. 2014.

[7] Maldonado Belmonte MJ. Adaptación del BRIEF ("Behavior Rating Inventory of Executive Function") a población española y su utilidad para el diagnóstico del trastorno por déficit de atención-hiperactividad subtipos inatento y combinado [PhD's thesis]. Universidad Complutense de Madrid. 2016. (In Spanish)

[8] Maldonado Belmonte MJ, Fournier del Castillo MC, Martínez Arias R, Gioia GA. BRIEF2: Evaluación conductual de la función ejecutiva. TEA Ediciones: Madrid. 2017. (In Spanish)

[9] Kleka P, Paluchowski WJ. Shortening of psychological tests–assumptions, methods and doubts. Polish Psychological Bulletin. 2017; 516–522.

[10] Schroeders U, Wilhelm O, Olaru G. Meta-Heuristics in Short Scale Construction: Ant Colony Optimization and Genetic Algorithm. PloS One. 2016; 11: e0167110. https://doi.org/10.1371/journal.pone.0167110.

[11] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology). 2005; 67: 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[12] Swanson JM. School-Based Assessment and Interventions for ADD Students. KC Publications: Irvine, CA, USA. 1992.

[13] Conners CK, Sitarenios G, Parker JD, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. Journal of Abnormal Child Psychology. 1998; 26: 257–268. https://doi.org/10.1023/a:1022602400621.

[14] Purper-Ouakil D, Blasco-Fontecilla H, Ros T, Acquaviva E, Banaschewski T, Baumeister S, *et al*. Personalized at-home neurofeedback compared to long-acting methylphenidate in children with ADHD: NEWROFEED, a European randomized noninferiority trial. Journal of Child Psychology and Psychiatry, and Allied Disciplines. 2022; 63: 187–198. https://doi.org/10.1111/jcpp.13462.

[15] Toplak ME, Bucciarelli SM, Jain U, Tannock R. 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. 2009; 15: 53–72. https://doi.org/10.1080/09297040802070929.

[16] Bussing R, Fernandez M, Harwood M, Wei Hou, Garvan CW, Eyberg SM, *et al*. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. Assessment. 2008; 15: 317–328. https://doi.org/10.1177/1073191107313888.

[17] Grañana N, Richaudeau A, Gorriti CR, O'Flaherty M, Scotti ME, Sixto L, *et al*. Assessment of attention deficit hyperactivity: SNAP-IV scale adapted to Argentina. Revista Panamericana De Salud Publica = Pan American Journal of Public Health. 2011; 29: 344–349. https://doi.org/10.1590/s1020-49892011000500007.

[18] Deb S, Dhaliwal AJ, Roy M. The usefulness of Conners' Rating Scales-Revised in screening for attention deficit hyperactivity disorder in children with intellectual disabilities and borderline intelligence. Journal of Intellectual Disability Research: JIDR. 2008; 52: 950–965. https://doi.org/10.1111/j.1365-2788.2007.01035.x.

[19] Artieda-Urrutia P, Delgado-Gómez D, Ruiz-Hernández D, García-Vega JM, Berenguer N, Oquendo MA, *et al*. Short Personality and Life Event scale for detection of suicide attempters. Revista De Psiquiatria Y Salud Mental. 2015; 8: 199–206. https://doi.org/10.1016/j.rpsm.2015.02.004.

[20] Blasco-Fontecilla H, Delgado-Gomez D, Ruiz-Hernandez D, Aguado D, Baca-Garcia E, Lopez-Castroman J. Combining scales to assess suicide risk. Journal of Psychiatric Research. 2012; 46: 1272–1277. https://doi.org/10.1016/j.jpsychires.2012.06.013.

[21] Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, *et al*. Package 'glmnet'. 2021; 33.

[22] Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, *et al*. Package 'caret'. The R Journal. 2020; 223: 7.

[23] Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika. 1969; 34: 1–97.

[24] Chalmers RP. mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software. 2012; 48: 1–29. https://doi.org/10.18637/jss.v048.i06.

[25] Leite WL, Huang IC, Marcoulides GA. Item Selection for the Development of Short Forms of Scales Using an Ant Colony Optimization Algorithm. Multivariate Behavioral Research. 2008; 43: 411–431. https://doi.org/10.1080/00273170802285743.

[26] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12: 77. https://doi.org/10.1186/1471-2105-12-77.

[27] Dinno A. Package 'paran'. R package version: Dortmund, Germany. 2018.

[28] R Core Team. R: a Language and Environment for Statistical Computing. Version 4.3.2. 2023. R Foundation for Statistical Computing: Vienna, Austria. Available at: https://cran.r-project.org/ (Accessed: 21 March 2025).

[29] Kruyen PM, Emons WH, Sijtsma K. On the shortcomings of shortened tests: A literature review. International Journal of Testing. 2013; 13: 223–248. https://doi.org/10.1080/15305058.2012.703734.

[30] Zijlmans EAO, Tijmstra J, van der Ark LA, Sijtsma K. Item-Score Reliability as a Selection Tool in Test Construction. Frontiers in Psychology. 2019; 9: 2298. https://doi.org/10.3389/fpsyg.2018.02298.

[31] Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. Psychological Assessment. 2000; 12: 102–111. https://doi.org/10.1037//1040-3590.12.1.102.

[32] Kleka P, Soroko E. How to avoid the sins of questionnaires abridgement? Survey Research Methods. 2018; 12: 147–160.

[33] Caselles-Pina L, Del Amo PS, Aguado D, López-Castromán J, Sanjuán-Antúnez JDD, Delgado-Gómez D. A data science approach to optimize ADHD assessment with the BRIEF-2 questionnaire. Translational Neuroscience. 2024; 15: 20220349. https://doi.org/10.1515/tnsci-2022-0349.