Received: 31 October 2016 | Revised: 29 January 2017 | Accepted: 13 March 2017

DOI: 10.1002/mpr.1567

WILEY

ORIGINAL ARTICLE

Adult executive functioning inventory (ADEXI): Validity, reliability, and relations to ADHD

Ylva Holst | Lisa B. Thorell

Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Correspondence
Lisa B. Thorell, Department of Clinical Neuroscience, Division of Psychology, Nobels Väg 9, SE-17165 Solna, Sweden
Email: lisa.thorell@ki.se

Funding information
Swedish Research Council

Abstract

This study examined the psychometric properties of the Adult Executive Functioning Inventory (ADEXI). This new executive functioning (EF) rating instrument has the advantage of being brief (14 items) and focusing specifically on working memory and inhibitory control. Results showed that scores on the ADEXI had high internal consistency and adequate test–retest reliability, but low concurrence between self-ratings and other ratings. High and statistically significant correlations were found between ADEXI scores and scores from another EF rating instrument, whereas the correlations between ADEXI scores and neuropsychological test scores were weak and often non-significant. Furthermore, with regard to discriminant validity, individuals with attention deficit hyperactivity disorder (ADHD) had significantly higher scores on both the inhibition and working memory subscales compared to clinical as well as non-clinical controls. The results showed high specificity, but relatively low sensitivity, when discriminating between adults with ADHD and non-clinical controls. Conclusively, the ADEXI can be a valuable screening instrument for assessing deficits in working memory and inhibitory control. However, similarly to other EF ratings, the ADEXI should be used as a complement rather than as a replacement for neuropsychological tests, and the low interrater reliability suggests that ratings from multiple sources is preferable compared to relying solely on self-ratings.

KEYWORDS

ADHD, executive functions, inhibition, self-ratings, working memory

1 | INTRODUCTION

Executive functioning (EF) can be described as an umbrella term for various complex cognitive processes responsible for cognitive control of actions and thoughts that are necessary to maintain goal-directed behaviour in order to reach future goals such as working memory, inhibition, set-shifting, and planning (Welsh & Pennington, 1988). Furthermore, with regard to the link between EF deficits and psychiatric disorders, the strongest link has been found with attention deficit hyperactivity disorder (ADHD). Most previous research has been conducted on children (e.g. Barkley, 2014; Nigg, 2006, for reviews), but there is also evidence showing that adults with ADHD perform significantly worse than controls on EF tests (for reviews and meta-analyses, see Alderson, Kasper, Hudec, & Patros, 2013; Boonstra, Oosterlaan, Sergeant, & Buitelaar, 2005; Doyle, 2006; Hervey, Epstein, & Curry, 2004; Seidman, 2006).

Several neuropsychological test batteries that target different types of EF deficits have been developed of which the "Delis–Kaplan Executive Function System" (D-KEFS; (Delis, Kaplan, & Kramer, 2001) is probably the most well-known of those used for both children and adults. However, one important issue regarding the use of neuropsychological tests for measuring EF deficits is that the tests often have low ecological validity. This has been demonstrated in studies showing only weak relations between EF tests and everyday abilities that are believed to be dependent on well-functioning executive skills such as academic achievement, social relations, work performance, criminality, and substance abuse (Barkley & Fischer, 2011; Barkley & Murphy, 2010, 2011; Szuromi, Bitter, & Czobor, 2013). However, EF deficits measured through self-ratings have been shown to be strongly related to functional impairments, and EF tests and EF ratings show only weak correlations (Barkley & Fischer, 2011; Barkley & Murphy, 2010).

The findings presented earlier have been interpreted to mean that EF tests and EF ratings do not measure the same underlying mental constructs and that EF tests should therefore not be relied on as the sole source for measuring EF deficits (Toplak, West, & Stanovich, 2013). The most well-known EF rating instruments for adults are the

Behaviour Rating Inventory of Executive Functions – Adult version (BRIEF-A; Roth, Isquith, & Gioia, 2005), Barkley's Deficits in Executive Function Scale (BDEFS; Barkley, 2011) and the Dysexecutive Questionnaire (DEX; Burgess, Alderman, Wilson, Evans, & Emslie, 1996). However, as further explained later, these ratings suffer from limitations. The present study therefore aimed at introducing a new instrument, the Adult Executive Functioning Inventory (ADEXI).

The most important limitation of existing EF rating instruments is that most of them include items measuring ADHD symptoms. For example, the BRIEF-A includes items that measure difficulties in sitting still, or having a short attention span. These items are essentially identical to the symptom criteria for ADHD as presented in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association [APA], 2013). The DEX also contains questions concerning ADHD symptoms such as impulsivity and restlessness.

Another aspect that needs to be taken into consideration is what constructs the questionnaire intends to capture. When examining individual items, it becomes evident that some available rating instruments are very broad as they include items measuring cognitive functioning in general (e.g. not being able to process information quickly or adequately) or items that are related to EF but do not measure this construct directly (e.g. easily becoming angry or upset, which measures emotional reactivity/regulation). There are also examples of items reflecting how others behave towards the rater rather than characteristics the rater sees in him/herself (e.g. that other people tell the patient that he/she is lazy or unmotivated). The inclusion of these items makes it difficult to draw conclusions regarding to what extent a patient has specific problems with EF or more general neuropsychological deficits.

A final limitation is that the rating instruments described earlier are quite extensive (i.e. 75–89 items except for the DEX and the short-version of the BDEFS). Adults who are being referred to a clinic for a neuropsychiatric assessment often have difficulties answering long rating scales. This emphasizes the need for a short screening instrument that can assess different aspects of EF deficits in a valid and reliable way. The ADEXI is intended to serve such a purpose and to address the earlier-mentioned limitations.

The ADEXI has been used in one previous study (Holst & Thorell, 2017) in which significant group differences were found between adults with ADHD and a clinical control group of adults with other psychiatric disorders. However, the reliability of the questionnaire was not examined, and the previous study did not include a non-clinical control group. Thus, there is a need to know more about the psychometric properties of the ADEXI and to what extent it can be used to discriminate between adults with ADHD and healthy controls. The ADEXI is an adult version of the Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008). Furthermore, similar to the CHEXI, the ADEXI is freely available for use by both researchers and clinicians (www.chexi.se). CHEXI scores have been shown to have good test–retest reliability (r = 0.89) and factor analysis has identified two factors tapping working memory and inhibition (Thorell & Nyberg, 2008). In addition, scores from the CHEXI have been shown to discriminate well between children with ADHD and controls (Catale, Meulemans, & Thorell, 2015; Thorell, Eninger, Brocki, & Bohlin, 2010).

1.1 | Aims of the present study

The overall aim of the present study was to examine the psychometric properties of ADEXI scores. More specifically, the present study aimed to address the following aspects:

1. Use factor analysis to investigate whether the two major factors found for the CHEXI (i.e. inhibition and working memory) can be replicated using the ADEXI.
2. Study the reliability of ADEXI scores in terms of both test–retest and interrater reliability.
3. Study the convergent validity of the ADEXI by investigating the association between ADEXI scores and scores from both another EF rating instrument and laboratory measures of EF.
4. Examine to what extent ADEXI scores can be used to discriminate between adults with ADHD, and adults either with or without another psychiatric disorder.

2 | METHOD

2.1 | Participants and procedure

The present study included altogether 202 participants from three groups: a clinical group of adults with ADHD (n = 51, 39% men); a clinical group of adults diagnosed with other psychiatric disorders (n = 46, 28% men); and a non-clinical sample consisting of individuals from a random sample and a sample of university students (n = 105, 42% men). The groups did not differ significantly from one another with regard to age, F = 1.23 [not significant (n.s.)] [ADHD group: mean (M) = 27.43, standard deviation (SD) = 6.31; Clinical controls: M = 25.65, SD = 4.76; Non-clinical controls: M = 26.50, SD = 5.57) or gender (χ^2 = 2.56, n.s.). However, there was a significant difference with regard to educational level. The two clinical groups had a similar educational level, but they both had lower educational level compared to the non-clinical controls.

2.1.1 | Clinical samples

The participants in the two clinical samples were recruited through advertisements at three outpatient psychiatric clinics, and they visited the clinic on two occasions to perform the neuropsychological testing. Questionnaire data was also collected from a close relative or friend of each participant. The participants in the ADHD group underwent a neuropsychiatric assessment conducted by a licensed psychologist. The assessment included a clinical judgement using the second version of the Diagnostic Interview for ADHD in Adults (DIVA 2.0; (Kooij & Francken, 2010). This semi-structured interview consists of two parts: one for assessing the presence of all 18 criteria in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; APA, 1994) during childhood and in the present; the other for assessing impairment in five areas of functioning (i.e. education, work, family, social/relationships, and self-confidence) in childhood and in the present. In addition, current levels of ADHD symptoms were assessed using self-report on the Adult ADHD Self-Report Scale (ASRS-v1.1; Kessler

et al., 2005). The psychologist also interviewed a close relative of the participant, in most cases his/her mother, to obtain a detailed anamnesis. All participants in the ADHD group met the full diagnostic criteria according to the DSM-5 (APA, 2013). Finally, all participants underwent testing of global intellectual ability using the fourth edition of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008). Exclusion criteria were an IQ score of <80 on WAIS-IV and the presence of substance-related disorders. In addition to a primary ADHD diagnosis, the participants in the ADHD group also met the criteria for the following comorbid diagnoses: mood disorders including "major depression" (15.8%), bipolar disorder (5.3%), unspecified anxiety disorder (UNS) (5.3%), panic disorder (3.5%), obsessive compulsive disorder (1.7%), social phobia (1.7%), and personality disorders (5.3%). Five of the participants had more than one comorbid diagnosis.

The diagnoses in the clinical control group were the following: mood disorders including "major depression" (43.4%), bipolar disorder (11.3%), anxiety disorder UNS (15.1%), social phobia (9.4%), panic disorder (1.8%), obsessive compulsive disorder (5.7%), general anxiety disorder (5.7%), post-traumatic stress disorder (5.7%), eating disorders (1.8%), and personality disorders (11.3%).

2.1.2 | Non-clinical sample

The study also included two non-clinical groups – a random sample and a sample of university students – which were combined in all analyses included in the present study. The random non-clinical sample of adults was recruited from a larger random sample recruited from a national population based register. From this sample, a subsample of 44 adults (41% men) matched to the clinical ADHD sample with regard to age and gender were included in the present study. In addition, a convenience sample of university students (n = 61, 43% men) was recruited through advertisements on the university campus. Exclusion criteria in the two non-clinical control groups were the presence of any psychiatric disorder.

2.2 | Measures

2.2.1 | Self-ratings

ADEXI

The ADEXI is a 14-item questionnaire (see Table 1) measuring working memory and inhibition. As mentioned earlier, it is an adult version of the CHEXI, and both the child and adult version are freely available in several languages (see www.chexi.se). Several of the items from the CHEXI are also included in the ADEXI. However, a few items that were considered irrelevant to adults (e.g. "Has difficulty following through on less appealing tasks unless he/she is promised some type of reward for doing so") were deleted. We also tried to minimize item overlap in order to make the ADEXI as quick as possible to complete (14 items in the ADEXI versus 24 in the CHEXI), although still including enough items to be able to capture the most central aspects of working memory and inhibition.

B-DEFS

In addition to the ADEXI, the non-clinical sample also completed the BDEFS (Barkley, 2011). The BDEFS includes the following six subscales: Self-organization/Problem Solving (24 items), Self-management of Time (21 items), Self-restraint/Inhibition (19 items), Self-regulation of Emotion (13 items), and Self-motivation (12 items). There is also a short-version of the BDEFS, which includes 20 items.

2.2.2 | Laboratory measures of EF deficits

Working memory

Working memory was measured by two subtests from the WAIS-IV (Wechsler, 2008): Letter-Number Sequencing and Digit Span. In Letter-Number Sequencing, participants have to repeat a series of randomly mixed letters and numbers starting with the numbers in ascending order followed by the letters in alphabetical order. For Digit Span,

TABLE 1 Result of the factor analysis for self-ratings

Item		Working memory	Inhibition
5.	When someone asks me to do several things, I sometimes remember only the first or last.	.915	
1.	I have difficulty remembering lengthy instructions.	.890	
12.	I have difficulties with tasks or activities that involve several steps.	.845	
7.	I have difficulty coming up with a different way of solving a problem when I get stuck.	.813	
2.	I sometimes have difficulty remembering what I am doing in the middle of an activity.	.726	
13.	I have difficulty thinking ahead or learning from experience.	.634	
8.	When someone asks me to fetch something, I sometimes forget what I am supposed to fetch.	.578	
11.	I sometimes have difficulty understanding verbal instructions unless I am also shown how to do something.	.537	.319
9.	I have difficulty planning for an activity (e.g. remembering to bring everything necessary when going on a trip/to work/to school).	.440	
14.	People that I meet sometimes seem to think that I am more lively/wilder compared to other people my age.	– .318	.938
4.	I sometimes have difficulty stopping myself from doing things that I like even though someone tells me that it is not allowed.		.724
6.	I sometimes have difficulty refraining from smiling or laughing in situations where it is inappropriate.		.706
3.	I have a tendency to do things without first thinking about what could happen.		.663
10.	I sometimes have difficulty stopping an activity that I like (e.g. I watch television or sit in front of the computer in the evening even though I know that it is time to go to bed).	.369	.325
Explained variance (%)		48.84	9.18

Note: Factor loadings ±0.25 are not displayed in the table.

the mean raw score (i.e. number of correct trials) for Digit Span Backwards and sequencing was included. In the backwards condition, participants have to repeat the series in a backwards order, and in Digit Span Sequencing, the numbers are randomly presented and must be repeated in the correct number order.

Inhibition

Inhibition was measured using the Colour Word Test from the D-KEFS (Delis et al., 2001). Only the third trial (i.e. interference trial) was used. In this trial, participants are presented rows of words printed in dissonant colours and are instructed to inhibit reading the words and instead name the dissonant colours in which the words are printed. The number of seconds (i.e. raw score) needed to complete the trial was used as a measure of inhibition.

2.3 | Statistical analyses

All statistical analyses were performed using SPSS 23.0 statistical package. In order to first determine whether the two factors found in the CHEXI could be replicated using the ADEXI (i.e. working memory and inhibition), factorial validity was investigated using the maximum likelihood procedure. Kaiser's measure of sampling adequacy was 0.93 and Bartlett's test of sphercity was significant, χ^2 = 3170, $p < 0.0001$.

In order to study reliability, the internal consistency for the scores of each subscale was first examined. In line with recommendations (Nunnaly, 1978), an α-value of 0.70 to 0.79 was considered as fair, 0.80 to 0.89 as good, and ≥0.90 as excellent. Second, test–retest reliability using ratings collected 2–3 weeks apart were examined using both Pearson correlation analyses and intra-class correlations (ICCs). Third, Pearson correlation analyses and ICCs were used to investigate interrater reliability between the scores obtained from the self-ratings and those collected from a close relative/friend. In line with recommendations (Cohen, 1988), the following guidelines were used for adequate (0.70–0.79), good (0.80–0.89), and excellent (≥ 0.90) reliability when examining correlations. As regards to ICC, values less than 0.40 were considered poor, 0.40–0.59 were considered fair, 0.60–0.74 were considered good, and above 0.75 were considered excellent (e.g. Cicchetti, 1994). Paired *t*-tests were used to investigate to what extent self-ratings and rating completed by a significant other differed significantly from one another.

Convergent validity was first examined by studying correlations between scores from the ADEXI and the scores from the BDEFS. In addition, we studied correlations between ADEXI scores and scores from laboratory measures of EF. For discriminant validity, analyses of variance (ANOVAs) were used to study group difference for the ADEXI. Group differences were complemented with measures of effect sizes using Cohen's effect size formula, where 0.30 represents a small effect size, 0.50 a medium effect size, and 0.80 a large effect size (Cohen, 1992). In addition to ANOVAs, logistic regression analyses were conducted. In the first regression model, only ADEXI was included. In the second model, hierarchical regression analyses were used. Here, the executive function tasks were entered in step one and the ADEXI in step two. This allowed us to examine whether adding ADEXI scores would increase the sensitivity and/or specificity of belonging to the ADHD group beyond the influence of EF deficits. In all analyses addressing discriminant analyses, only the random non-clinical sample, not the student sample, was included as it was considered important to compare the ADHD group with an age-matched control group with a similar level of education.

3 | RESULTS

3.1 | Factorial validity of ADEXI scores

Both a two- and three-factor solution was examined when conducting the factor analysis in order to determine whether the same number of factors obtained for the childhood version of the questionnaire (i.e. the CHEXI) could also be found for the ADEXI. The three-factor solution did not appear to provide a good fit for the data as several items loaded on more than one factor. However, the two-factor solution showed two clear factors. The scree test and the eigenvalue criterion also supported a two-factor solution. An oblique rotation method was chosen as two factors were shown to be highly correlated, $r = 0.69$. The two-factor solution accounted for 50.02% of the variance. As expected (see Table 1), one factor was comprised of items from the a priori subscale tapping working memory and the other factor was comprised of items from the a priori subscale tapping inhibition. The exception was item 10 ("I sometimes have difficulty stopping an activity that I like") which had a similar factor loading on both factors. Based on the a priori subscales, this item was selected to be part of the inhibition subscale.

3.2 | Reliability of ADEXI scores

All reliability estimates are presented in Table 2. The results first of all showed high internal consistency for scores on the ADEXI full scale as well as for scores on the inhibition and working memory subscales. Second, the test–retest reliability was found to be adequate with estimates ranging between 0.68 and 0.72 for bivariate correlations and between 0.62 and 0.72 for ICCs. However, as can also be seen in Table 2, the interrater reliability for the full scale between self-ratings (i.e. ratings made by the participant him/herself) and other-ratings (i.e. ratings made by a close relative/friend) was low: 0.53 for bivariate correlations and 0.49 for the ICC. The concordance between self-ratings and other ratings was especially low with regard to inhibition but also relatively low for working memory. Paired-samples *t*-tests showed that the mean scores for the full scale as well as the two subscales were all significantly higher (i.e. more problems) for self-ratings compared to other ratings, all *t* values <2.88, *p* values <0.01.

3.3 | Convergent validity of ADEXI scores

As shown in Table 2, relatively strong relations (*r* values ranging between 0.48 and 0.72) were found for all three ADEXI scores and scores on the BDEFS subscales. As expected, scores on the working memory subscale of the ADEXI were most strongly related to scores on the BDEFS subscale "Self-organization/Problem Solving", whereas scores on the ADEXI Inhibition subscale were most strongly related to scores on the BDEFS subscale referred to as "Self-restraint/Inhibition".

TABLE 2 Reliability and convergent validity for the adult executive functioning inventory (ADEXI)

	ADEXI		
	Full scale	Inhibition	Working memory
Internal consistency			
Total sample (*n* = 461)	.91	.77	.90
Non-clinical sample (*n* = 364)	.89	.72	.88
Clinical sample (*n* = 97)	.90	.73	.89
Test–retest reliability			
Bivariate correlations (*n* = 105)	.71	.72	.68
Intra-class correlations (*n* = 105)	.67	.72	.62
Interrater reliability			
Bivariate correlations (*n* = 88)	.53	.38	.56
Intra-class correlations (*n* = 88)	.49	.34	.54
BDEFS subscales (n = 127)			
Self-organization/problem solving	.71***	.51***	.72***
Self-management of time	.66***	.51***	.65***
Self-restraint/inhibition	.64***	.62***	.57***
Self-regulation of emotion	.58***	.48***	.55***
Self-motivation	.64***	.57***	.59***
BDEFS short version	.70***	.58***	.66***

***$p < .001$

As a second measure of convergent validity, we examined the relation between ADEXI scores and neuropsychological test scores tapping inhibition or working memory (see Table 3). With regard to self-ratings, scores on the ADEXI full scale and ADEXI working memory subscale were significantly correlated with scores on the Colour Word Task measuring inhibition in both the clinical and the non-clinical sample. In addition, scores on the ADEXI working memory subscale were related to the two working memory measures within the non-clinical sample. As regards to other ratings, no significant relations were found between the three ADEXI measures and scores on the digit span task. However, all ADEXI measures rated by a significant other were related to scores on the number-letter sequencing task and the Colour Word Task except for a non-significant relation between scores on the ADEXI inhibition subscale and scores on the Colour Word Task. In summary, it can be concluded that although a number of significant relations were found between ADEXI scores and scores obtained from the laboratory tests, all correlations were of small effect sizes (i.e. *r* values <0.30).

3.4 | Discriminant validity of the ADEXI

The results of the ANOVAs examining group difference (see Table 4) showed that the ADHD group reported higher scores (i.e. more EF

TABLE 3 Relations between the adult executive functioning inventory (ADEXI) and neuropsychological test measures (one-tailed)

	Clinical sample (*n* = 97)			Non-clinical sample (*n* = 105)		
	Digit span	Number-letter sequencing	Colour-word test[a]	Digit span	Number-letter sequencing	Colour-word test[a]
Self-ratings						
Full scale	−.13	−.14	.19*	−.10	−.09	.21*
Inhibition	−.15	−.11	.12	.09	.10	.09
Working memory	−.11	−.14	.20*	−.18*	−.18*	.24**
Other ratings						
Full scale	−.16	−.29**	.26*	NA	NA	NA
Inhibition	−.14	−.26**	.15	NA	NA	NA
Working memory	−.15	−.27**	.28**	NA	NA	NA

Note: NA, not available;

*$p < 0.05$;

**$p < 0.01$.

[a]Similar associations between the Colour Word Test and the Adult Executive Functioning Inventory (ADEXI) subscales were obtained when using the contrast score (i.e. subtracting the colour naming score from the interference score) rather than using the raw score for the interference trial.

TABLE 4 Group differences on the adult executive functioning inventory (ADEXI) between the attention deficit hyperactivity disorder (ADHD) group (*n* = 51), the clinical control group (*n* = 46), and the non-clinical group (*n* = 44)

	ADHD group (group I)	Clinical control group (group II)	Non-clinical control group (group III)			Effect sizes	
						Group I versus group III	Group I versus group II
	M (SD)	*M* (SD)	*M* (SD)	F	Post hoc	d	d
ADEXI full scale	3.41 (.71)	2.50 (.80)	2.03 (.61)	70.32	1 > 2 > 3	2.07	1.21
ADEXI inhibition	3.46 (.84)	2.72 (.92)	2.14 (.71)	48.06	1 > 2 > 3	1.69	.84
ADEXI working memory	3.39 (.78)	2.39 (.90)	1.97 (.68)	59.62	1 > 2 > 3	1.93	1.19

Note: *M*, mean; SD, standard deviation.

deficits) than both the clinical and the non-clinical control group for all three measures with large effect sizes. For the logistic regression analyses (see Table 5), we first compared the ADHD group with non-clinical controls. Only the ADEXI scores were included in this analysis. The result showed that the model was significant (χ^2 = 56.08, n = 95, p < 0.001), and the ADEXI scores classified 85% of the participants in the correct category with a sensitivity of 86% and a specificity of 84% (Model 1). Second, we compared the ADHD group with the clinical control group. The results showed that this model was also significant (χ^2 = 30.09, n = 97, p < 0.001) and the ADEXI scores classified 76% of the sample correctly with a sensitivity of 80% and a specificity of 72% (Model 2).

In a second set of logistic regression analyses, we wanted to examine to what extent ADEXI scores can increase the discriminatory ability beyond the influence of EF tests. When comparing the ADHD group with non-clinical controls (Model 3), the results showed that scores from the EF tasks correctly classified 71% of the participants with a sensitivity of 64% and a specificity of 77%. When adding ADEXI scores in the second step, this step was significant (χ^2 = 42.76, n = 95, p < 0.001) and classified 85% of the participants in the correct category with a sensitivity of 87% and a specificity of 84%. When comparing the ADHD group with the clinical control group (Model 4), the results showed that scores on the EF tasks correctly classified 67% of the participants with a sensitivity of 64% and a specificity of 70%. When adding ADEXI scores in the second step, this step was significant (χ^2 = 24.00, n = 97, p < 0.001) and could classify 81% of the participants in the correct category with a sensitivity of 84% and a specificity of 77%.

4 | DISCUSSION

In this study, we have emphasized the need for a short reliable rating scale focused on measuring EF deficits not including items measuring other related constructs or symptoms of ADHD. However, no instrument meeting these criteria has been made available, which is the reason why the ADEXI was created. The present study showed that scores from the ADEXI generally have relatively good psychometric properties. However, due to the low correlations between ADEXI scores and those obtained from laboratory EF tests, the ADEXI should be used as a complement rather than a replacement for neuropsychological tests. In addition, the low concordance between self-ratings and other ratings demonstrates the need to use ratings from multiple sources rather than relying solely on self-ratings.

TABLE 5 Results of the logistic regression analyses

	Sensitivity (%)	Specificity (%)	Total correctly classified (%)
Model 1: ADHD versus non-clinical controls			
Step 1. Only ADEXI	86.3	84.1	85.3
Model 2: ADHD versus clinical controls			
Step 1. Only ADEXI	80.4	71.7	76.3
Model 3: ADHD versus non-clinical controls			
Step 1. Only executive functioning tasks	64.4	77.3	70.8
Step 2. Executive functioning tasks + ADEXI	86.7	84.1	85.4
Model 4: ADHD versus clinical controls			
Step 1. Only executive functioning tasks	64.4	70.5	67.4
Step 2. Executive functioning tasks + ADEXI	84.4	77.3	80.9

Note: ADHD, attention deficit hyperactivity disorder; ADEXI, Adult Executive Functioning Inventory.

4.1 | Reliability of the ADEXI scores

With regard to the reliability of the ADEXI scores, the internal consistency was shown to be high (i.e. αs around 0.90 for the ADEXI full scale) and comparable or higher compared to those of other EF rating scales for adults such as the BDEFS (α values ≥0.91; Barkley, 2011) and the BRIEF-A (α = 0.98 for the full scale and 0.84 for the inhibit subscale and 0.90 for working memory subscale; Roth et al., 2005). The test–retest reliability for the ADEXI scores was shown to be within the range of what is generally considered as adequate (r values ranging between 0.68 and 0.72 and ICC between 0.62 and 0.72). The obtained bivariate correlation coefficients for test–retest reliability is somewhat lower compared to studies using the BRIEF-A (r values <0.82; Roth et al., 2005), but higher compared to estimates for some of the subscales included in BDEFS (r values <0.62; Barkley, 2011). The present study as well as the previous studies using the BRIEF-A and BDEFS included only non-clinical samples when assessing test–retest reliability. This is a limitation as these instruments are intended to be used in clinical populations. However, the test–retest reliability for scores on the Conners Adult ADHD Rating Scale (CAARS) has been shown to be higher in clinical compared to non-clinical samples (Christiansen et al., 2012). As the CAARS assess similar constructs as the ADEXI, this study could be taken to mean that the test–retest reliability of the ADEXI scores is likely to be adequate also in clinical samples, although this issue needs to be empirically examined.

The interrater reliability between participants and a close relative/friend was found to be low with regard to inhibition (r = 0.38) but also quite low for working memory (r = 0.56). These estimates are lower than those obtained for the BDEFS (r values =0.66–0.79; Barkley, 2011). However, the interrater reliability has been found to be very low (r < 0.50) for several of the subscales included in BRIEF-A (Roth et al., 2005). In addition, there are previous studies that have demonstrated very low concordance between self-ratings and other ratings also for ratings of ADHD symptoms (e.g. Mörtstedt, Corbisiero, Bitto, & Stieglitz, 2015).

In addition to studying correlations between self-ratings and those obtained by a significant other, it can also be of value to study mean differences. In the present study, the results showed that the scores obtained for the self-ratings were significantly higher compared to those obtained from a significant other. This finding is in line with two previous studies of EF ratings in adults using either the BRIEF-A (Roth et al., 2005) or the DEX (Rizzo, Steinhausen, & Dreschler, 2012) as well as with two previous studies investigating ratings of ADHD symptom levels (Katz, Petscher, & Welles, 2009; Kooij et al., 2008). Interestingly, Kooij et al. (2008) concluded that although patients rated their symptom levels as higher than significant others, both tended to under-report symptom levels in relation to investigator reports.

In summary, it appears as if interrater reliability is often relatively low when investigating EF deficits or related constructs in ADHD samples. This is problematic and points to the importance of including information from several sources in order to obtain a full picture of the patient's strength and weaknesses. Further research is clearly needed in order to determine how to best integrate information from different sources when interrater agreement is low.

4.2 | Validity of the ADEXI scores

Regarding convergent validity, the results showed that scores on all BDEFS subscales correlated significantly with scores on the ADEXI full scale as well as with scores on the two subscales, inhibition and working memory. As expected, scores from the subscales that aimed to assess the most similar constructs were also those that showed the highest correlations. Thus, these findings could be taken as support for the convergent validity of the ADEXI scores.

As a second measure of convergent validity, relations between ADEXI scores and neuropsychological test scores were investigated, but the results showed only small, and mostly non-significant, relations. These findings are in accordance with several previous studies examining the relation between scores obtained from EF ratings and EF tests using either a childhood or adult sample (for a review, see Toplak et al., 2013). This could be taken as an indication that ratings and tests capture at least partially different constructs. In line with this thinking, the present study was able to show that sensitivity was much increased when the scores from both the laboratory test and the ADEXI were entered into the logistic regression model. The fact that EF tests and EF ratings explain independent variance in relation to ADHD suggests that EF ratings should be used as a complement rather than as a replacement for laboratory tests. Previous studies comparing EF tests and EF ratings have also argued that EF ratings have higher ecological validity as they are more predictive of functional impairments in daily life such as occupational adjustment, antisocial behaviours, and social relations (Barkley & Murphy, 2010, 2011). It will therefore be of importance for future studies to investigate the link between ADEXI scores and measures of daily functioning.

4.3 | Discriminative validity

The ADEXI scores were shown to have high discriminative validity with large effect sizes being found when comparing ADHD patients to either non-clinical or clinical controls. The sensitivity was 0.76 and the specificity was 0.91 when comparing the ADHD group to the non-clinical control group and including only the ADEXI scores. Current models of heterogeneity within ADHD research (e.g. Nigg, 2006) have emphasized that only a subgroup of individuals with ADHD has significant EF deficits. It is therefore not surprising, that the present study showed that the sensitivity (i.e. the true positive rate) was lower compared to the specificity (i.e. the true negative rate) when comparing individuals with ADHD to non-clinical controls. The specificity and sensitivity have to our knowledge seldom been examined for scores obtained from other EF rating instruments for adults such as the BDEFS or the BRIEF-A. This is a serious limitation as group differences alone are insufficient indices of the discriminant validity of a measure (Doyle, Biederman, Seidman, Weber, & Faraone, 2000). However, one study showed that scores on the BRIEF-A had a sensitivity of 0.75 and a specificity of 0.79 when trying to discriminate between college students with or without ADHD (Hauser, Lukomski, & Samar, 2013). Another study showed that 89–98% of adults with ADHD were considered to be in the clinically significant range compared to only 8–11% of the community control group using scores on the BDEFS (Barkley, 2011). In summary, the results of the present study show that ADEXI scores can discriminate between individuals with ADHD and healthy controls as well as, or even better than, other available EF rating instruments for adults.

Previous studies using either EF test or ratings have shown that adults with other psychiatric disorders besides ADHD (e.g. depression and anxiety) also have EF deficits (e.g. Castaneda, Tuulio-Henriksson, Marttunen, Suvisaari, & Lönnqvist, 2008; Henin et al., 2009; Nilsson et al., 2016; Snyder, 2016). However, all these studies contrasted psychiatric patients with healthy controls. It is therefore very interesting to note that our study showed that EF deficits were more pronounced among individuals with ADHD compared to the clinical controls as demonstrated both by significant group differences and a classification rate above 0.80 when including both ratings and tests.

5 | CONCLUSIONS AND FUTURE DIRECTIONS

In conclusion, we have shown that ADEXI scores have adequate psychometric properties and that this instrument serves a partly different purpose compared to other available EF rating instruments for adults. Compared to the BRIEF-A, the ADEXI has a clear advantage of not including items that are more or less identical to the diagnostic criteria for ADHD, but instead focuses more specifically on EF deficits. In comparison with BDEFS, the ADEXI is much less comprehensive, which could be seen as a disadvantage. However, individuals with EF deficits often have difficulties completing long rating instruments, and it might therefore be an advantage to have a quick and easily administered screening instrument focusing specifically on two central EF aspects (i.e. inhibition and working memory), which is also freely available (www.chexi.se).

ACKNOWLEDGEMENTS

This study was supported by a grant from the Swedish Research Council to the last author. The authors wish to thank Emelie von Vogelsang Antonsson for her valuable help with data the collection.

DECLARATION OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

REFERENCES

Alderson, R. M., Kasper, L. J., Hudec, K. L., & Patros, C. H. (2013). Attention deficit/hyperactivity disorder (ADHD) and working memory in adults: A meta-analytic review. *Neuropsychology*, *27*, 287–302. https://doi.org/10.1037/a0032371

American Psychiatric Association (APA) (1994). *Diagnostic and Staistical Manual of Mental Disorders*. Washington, DC: APA.

American Psychiatric Association (APA) (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*. Washington, DC: APA.

Barkley, R. A. (2011). *Barkley Deficits in Executive Functioning Scale (BDEFS)*. New York: Guilford Press.

Barkley, R. A. (2014). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. New York: Guilford Publications.

Barkley, R. A., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, *36*(2), 137–161. https://doi.org/10.1080/87565641.2010.549877

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: The predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology*, *25*(3), 157–173. https://doi.org/10.1093/arclin/acq014

Barkley, R. A., & Murphy, K. R. (2011). The nature of executive function (EF) deficits in daily life activities in adults with ADHD and their relationship to EF tests. *Journal of Psychopathology and Behavioral Assessment*, *33*(2), 137–158. https://doi.org/10.1007/s10862-011-9217-x

Boonstra, A. M., Oosterlaan, J., Sergeant, J. A., & Buitelaar, J. K. (2005). Executive functioning in adult ADHD: A meta-analytic review. *Psychological Medicine*, *35*, 1097–1108. https://doi.org/10.1017/S003329170500499X

Burgess, P. W., Alderman, N., Wilson, B. A., Evans, J. J., & Emslie, H. (1996). *The Dysexecutive Questionnaire. Behavioral Assessment of the Dysexecutive Syndrome*. Bury St Edmunds: Thames Valley Test Company.

Castaneda, A. E., Tuulio-Henriksson, A., Marttunen, M., Suvisaari, J., & Lönnqvist, J. (2008). A review of cognitive impairments in depressive and anxiety disorders with a focus on young adults. *Journal of Affective Disorders*, *106*, 1–27. https://doi.org/10.1016/j.jad2007.06.006

Catale, C., Meulemans, T., & Thorell, L. B. (2015). The childhood executive function inventory: Confirmatory factor analyses and cross-cultural clinical validity in a sample of 8- to 11-year-old children. *Journal of Attention Disorders*, *19*(6), 489–495. https://doi.org/10.1177/1087054712470971

Christiansen, H., Kis, B., Hirsch, O., Matthies, S., Hebebrand, J., Uekerman, J., … Philipsen, A. (2012). German validation of the Conners adult ADHD rating scales (CAARS) II: Reliability, validity, diagnostic sensitivity and specificity. *European Psychiatry*, *27*, 321–328. https://doi.org/10.1016/j.eurpsy.2010.12.010

Cicchetti, D. V. (1994). Multiple comparison methods: Establishing guidelines for their valid application in neuropsychological research. *Journal of Clinical and Experimental Neuropsychology*, *16*, 155–161. https://doi.org/10.1080/01688639408402625

Cohen, J. (1988). Ser correlation and contingency tables. *Applied Psychological Measurement*, *12*, 425–434.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System*. San Antonio, TX: The Psychological Corporation.

Doyle, A. (2006). Executive functions in attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry*, *67*, 21–24.

Doyle, A. E., Biederman, J., Seidman, L. J., Weber, W., & Faraone, S. V. (2000). Diagnostic efficiency of neuropsychological test scores for discriminating boys with and without attention deficit-hyperactivity disorder. *Journal of Consulting and Clinical Psychology*, *68*(3), 477–488. https://doi.org/10.1037//0022-006X.68.3.477

Hauser, P. C., Lukomski, J., & Samar, V. (2013). Reliability and validity of the BRIEF-A for assessing deaf college students' executive functions. *Journal of Psychoeducational Assessment*, *31*, 363–374. https://doi.org/10.1177/0734282912464466

Henin, A., Micco, J. A., Wozniac, J., Briesch, J. M., Narayan, A. J., & Hirschfeld-Becker, D. R. (2009). Neurocognitive functioning in bipolar disorder. *Clinical Psychology: Science and Practice*, *16*, 231–250. https://doi.org/10.1111/j.1468-2850.2009.01162.x

Hervey, A. S., Epstein, J. N., & Curry, J. F. (2004). Neuropsychology in adults with attention-deficit/hyperactivity disorder: A meta-analytical review. *Neuropsychology*, *18*, 485–503. https://doi.org/10.1037/0894-4105.18.3.485

Holst, Y., & Thorell, L. B. (2017). Neuropsychological functioning in adults with ADHD and adults with other psychiatric disorders: The issue of specificity. *Journal of Attention Disorders*, *21*, 137–148. https://doi.org/10.1177/1087054713506264

Katz, N., Petscher, Y., & Welles, T. (2009). Diagnosing attention deficit hyperactivity in college students: An investigation of the impact of informant ratings on diagnosing and subjective impairment. *Jornal of Attention Disorders*, *13*, 277–283. https://doi.org/10.1177/1087054708326112

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., … Walters, E. E. (2005). The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245–256. https://doi.org/10.1016/j.biopsych.2005.04.001

Kooij, J. J. S., Boonstra, A. M., Swinkles, S. H. N., Bekker, E. M., De Noord, I., & Buitelaar, J. K. (2008). Reliability, validity, and utility of instruments for self-report and informant report concerning symptoms of ADHD in adults patients. *Journal of Attention Disorders*, *11*, 445–458. https://doi.org/10.1177/1087054707299367

Kooij, J. J. S., & Francken, M. H. (2010). Diagnostic interview for ADHD in adults 2.0 (DIVA 2.0). Adult ADHD. Diagnostic assessment and treatment. Amsterdam: Pearson.

Mörtstedt, B., Corbisiero, S., Bitto, H., & Stieglitz, R.-D. (2015). Attention-deficit/hyperactivity disorder (ADHD) in adulthood: Concordance and differences between self- and informant perspectives on symptoms and functional impairments. *PloS One*, *10*, 1–15. https://doi.org/10.1371/journal.pone.0141342

Nigg, J. T. (2006). *What Causes ADHD? Understanding What Goes Wrong and Why*. New York: Guilford Press.

Nilsson, J., Thomas, A. J., Stevens, L. H., McAllister-Williams, R. H., Ferrier, I. N., & Gallagher, P. (2016). The interrelationship between attentional and deficits in major depressive disorder. *Acta Psychiatrica Scandinavica*, *134*, 73–82. https://doi.org/10.1111/acps.12570

Nunnaly, J. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Rizzo, P., Steinhausen, H. C., & Dreschler, R. (2012). Self-perception and perception of executive functions impairment in adults with ADHD. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, *60*, 131–143. https://doi.org/10.1024/1661-4747/a000108

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *Behavior rating inventory of executive function—Adult version: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Seidman, L. J. (2006). Neuropsychological functioning in people with ADHD across the lifespan. *Clinical Psychology Review*, *26*, 466–485. https://doi.org/10.1016/j.cpr.2006.01.004

Snyder, H. R. (2016). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*, 81–132. https://doi.org/10.1037/a0028727

Szuromi, B., Bitter, I., & Czobor, P. (2013). Functional impairment in adults positively screened for attention-deficit hyperactivity disorder: The role of symtom presentation and executive functioning. *Comprehensive Psychiatry*, *54*, 974–981. https://doi.org/10.1016/j.comppsych.2013.04.002

Thorell, L. B., Eninger, L., Brocki, K. C., & Bohlin, G. (2010). Childhood executive function inventory (CHEXI): A promising measure for identifying young children with ADHD? *Journal of Clinical and Experimental Neuropsychology*, *32*(1), 38–43. https://doi.org/10.1080/13803390902806527

Thorell, L. B., & Nyberg, L. (2008). The childhood executive functioning inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Psychology*, *33*, 536–552. https://doi.org/10.1080/87565640802101516

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive functions assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*, 199–230.

How to cite this article: Holst Y, Thorell LB. Adult executive functioning inventory (ADEXI): Validity, reliability, and relations to ADHD. *Int J Methods Psychiatr Res*. 2018;27:e1567. https://doi.org/10.1002/mpr.1567