Bendik Romundstad

Validity of the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) in Children and Adolescents with Pediatric Acquired Brain Injury

Hovedoppgave i Profesjonsstudiet i Psykologi
Veileder: Stian Solem
Desember 2020

Hovedoppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for psykologi

NTNU
Kunnskap for en bedre verden

Bendik Romundstad

Validity of the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) in Children and Adolescents with Pediatric Acquired Brain Injury

Hovedoppgave i Profesjonsstudiet i Psykologi
Veileder: Stian Solem
Desember 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for psykologi

NTNU
Kunnskap for en bedre verden

Validity of the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) in Children and Adolescents with Pediatric Acquired Brain Injury

Bendik Romundstad

Department of Psychology

Norwegian University of Science and Technology (NTNU)

PSYPRO4700 Graduate Thesis

Supervisor: Stian Solem

Co-supervisor: Torun Gangaune Finnanger

December 2020

Acknowledgements

This graduate thesis for the 6-year clinical psychology program at Norwegian University of Science and Technology (NTNU) marks the end of the theoretical portion of my studies. Executive functions and ecological validity are methodologically controversial constructs in the field of neuropsychology, which makes them important topics of interest to me. Therefore, it has been a privilege for me having had the opportunity to write my thesis on these topics, based on quality data from an ambitious research project (CORE pABI) connected to St. Olavs Hospital in Trondheim and Oslo University Hospital, Rikshospitalet. Previously, I have been involved as one of the test technicians in this project, and I appreciate the opportunity to write my thesis based on the acquired data.

The process of writing this thesis has both been challenging and rewarding. Being able to engage myself in-depth in a field of interest has been motivating and educational. However, formulating research questions based on the current gaps in the literature, in addition to handling the data for statistical analyses, has been especially challenging. But through the invaluable support from my supervisors and my own efforts, these obstacles were overcome. While my learning curve has been steep, conducting this project has made me more interested in pursuing a research career.

First and foremost, I want to thank my supervisor Stian Solem for all the encouragement and guidance he has provided. His thorough and constructive feedback, constant availability, and enthusiasm for the project, has been of great importance to me, both in terms of my motivation along the way and for the outcome of the thesis. I also want to express a heartfelt thanks to my co-supervisor Torun Gangaune Finnanger, first, for making it possible for me to write my thesis based on data from the CORE pABI project, and second, for providing important guidance and help throughout the process. In addition, I want to thank Anne Elisabeth Brandt, who, together with Torun, helped me decide on my research, presenting BADS-C in a way that sparked my interest. Finally, I want to thank researchers from the CORE pABI project, namely Jan Stubberud and Torstein Baade Rø, who together with Anne and Torun, provided valuable feedback on my thesis during a project meeting. The different perspectives have been greatly enriching, all having contributed to improve the quality of my thesis.

Abstract

Impaired executive functions (EF) in everyday life is common among children and adolescents with pediatric acquired brain injury (pABI). Yet, it is controversial whether everyday EF can be adequately measured using neuropsychological tests. The Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) was developed in order to address the need for a standardized ecologically valid test of EF among children and adolescents. However, research on its psychometric properties is rather sparse. The aim of this study was therefore to investigate the discriminant, concurrent, and ecological validity of the BADS-C in 74 Norwegian children and adolescents aged 10-17 years with pABI, in addition to 60 healthy controls. Results showed that the pABI group performed significantly worse than the control group on all of the BADS-C measures, except for the Playing Cards Test. BADS-C total score, Key Search Test, and Zoo Map Test 1, were significantly correlated with established neuropsychological tests and teacher questionnaire ratings of EF, but not with parent or self-report. Also, Key Search Test and Zoo Map Test 1 significantly predicted teacher ratings of EF, beyond IQ and established EF tests. Findings indicate that BADS-C is a solid tool in terms of discriminant validity, and that parts of it, namely BADS-C total score, Key Search Test, and Zoo Map Test 1 are robust in terms of concurrent and ecological validity. However, the ecological validity of these measures seems to be limited to the school context, based on teacher perceptions of the children's and adolescents' EF.

Keywords: Executive functions, ecological validity, discriminant validity, concurrent validity, pediatric acquired brain injury, Behavioural Assessment of the Dysexecutive Syndrome for Children, BADS-C

Sammendrag

Svekkede eksekutive funksjoner (EF) i hverdagen er utbredt blant barn og ungdommer med ervervet hjerneskade (pABI). Likevel er det omstridt hvorvidt hverdagslig EF kan måles tilstrekkelig med nevropsykologiske tester. The Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) ble utviklet for å adressere behovet for en standardisert, økologisk valid nevropsykologisk test av EF for barn og ungdommer. Testens psykometriske egenskaper er imidlertid mindre forsket på. Målet med denne artikkelen var derfor å undersøke den diskriminerende, samtidige, og økologiske validiteten til BADS-C i et utvalg 74 barn og ungdommer mellom 10 og 17 år med pABI, i tillegg til 60 friske kontroller. Resultatene viste at gruppen med pABI presterte signifikant dårligere enn kontrollgruppen på alle BADS-C variabler, utenom Playing Cards Test. BADS-C total score, Key Search Test og Zoo Map Test 1, var signifikant korrelert med etablerte nevropsykologiske tester og lærerutfylte spørreskjema av EF, men ikke med spørreskjema utfylt av foreldre eller ved selvrapport. Key Search Test og Zoo Map Test 1 predikerte også lærerutfylte spørreskjema av EF, utover IQ og etablerte tester av EF. Funnene indikerer at BADS-C er et godt verktøy med hensyn til diskriminerende validitet, og at deler av det, nærmere bestemt BADS-C total score, Key Search Test, og Zoo Map Test 1 er robust med hensyn til samtidig og økologisk validitet. Den økologiske validiteten til disse målene ser imidlertid ut til å være begrenset til skolekonteksten, basert på lærers oppfatning av barnet- eller ungdommens EF.

Nøkkelord: Eksekutive funksjoner, økologisk validitet, diskriminerende validitet, samtidig validitet, ervervet hjerneskade, Behavioural Assessment of the Dysexecutive Syndrome for Children, BADS-C

Validity of the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) in Children and Adolescents with Pediatric Acquired Brain Injury

Executive functions (EF) is an umbrella term that involves a constellation of high-level cognitive functions that enable goal-related behavior (Jurado & Rosselli, 2007), including skills such as effective problem solving, planning, and regulation of emotion and behavior. Well-functioning EF is associated with favorable outcomes in domains like mental and physical health, quality of life, and occupational performance (see Diamond, 2013 for a review). A common theoretical understanding is that executive functioning is dependent on three core cognitive processes: inhibition, mental flexibility, and working memory (Diamond, 2013; Miyake et al., 2000). However, although the overarching conceptualization of EF is generally agreed upon, it is defined and understood differently by different researchers, theories, and measurement methods (Chan et al., 2008; Welsh et al., 2006). Further, EF are seen as especially important in novel situations where routine-based or overlearned cognitive skills are insufficient. Thus, in a constantly changing world, EF can be regarded as crucial for adaptive human behavior, making us able to quickly shift our mindsets and adapt to new situations.

Children and adolescents with pediatric acquired brain injury (pABI) often show impairments in EF, especially in the more severe cases (Babikian & Asarnow, 2009; Koekkoek et al., 2008; Levin & Hanten, 2005; Long et al., 2011; Mangeot et al., 2002; Parrish et al., 2007). ABI is defined as damage to the brain occurring after birth, such as traumatic brain injury (TBI), stroke, infections, brain tumors, vestibular dysfunction, and postsurgical complications (Ciuffreda et al., 2012). Several studies have found that EF deficits in patients with pABI predict negative functional outcomes at a later stage in childhood (Fulton et al., 2012; O'Keeffe et al., 2014), adolescence (Arnett et al., 2013; Muscara et al., 2008), and into adulthood (Koskiniemi, 1999; Nybo et al., 200). Thus, accurate instruments for EF assessment in children and adolescents with pABI are important in order to guide rehabilitation and treatment.

EF can be measured by different approaches, including neuropsychological performance-based tests (PBTs), ecological tests, and questionnaires (Chevignard et al., 2010). PBTs typically measure how fast or precisely the examinee performs under highly structured conditions (e.g., Trail Making Test from Delis-Kaplan Executive Functioning System; D-KEFS; Delis et al., 2001), while the more open-ended ecological tests simulate real-life situations with less behavioral restraints (e.g., Children's Cooking Task; Chevignard et al., 2010). Questionnaires are typically based on parent, teacher, or caregiver ratings of the child's behavior in different contexts of daily life, such as home or school (e.g., Behavioural Rating Inventory of Executive Function; BRIEF; Gioia et al., 2000). While ecological tasks are less developed, PBTs and questionnaires of EF are commonly used in research and clinical practice (Chevignard et al., 2010).

However, assessing EF in children is challenging, especially with respect to ecological validity, which refers to the extent the test results can be generalized to real-life situations (Gioia & Isquith, 2004). Research has shown inconsistencies between normal performance on PBTs and contrasting challenges with EF in daily life activities, especially after early pABI (Anderson et al., 2002; Chaytor & Schmitter-Edgecombe et al., 2003; Chevignard et al., 2012; Eslinger et al., 2004; Toplak et al., 2013). For example, Topak et al. (2013) found a weak median correlation of r = .19 across 20 studies examining the correlation between PBTs and questionnaire measures of EF. Thus, there seems to be a lack of ecological validity of PBTs of EF.

It has been argued that the highly structured nature of PBTs of EF masks potential executive difficulties, thus not representing the relatively unstructured situations of daily life (Chevignard et al., 2012; Roy et al., 2015). In contrast, more open-ended tasks would demand the use of EF to generate strategies and the use of feedback to evaluate different approaches to the task. Others have argued that PBTs and questionnaire measures of EF capture different aspects of cognitive functioning: the efficiency in cognitive processing and success in goal-pursuit, respectively (Toplak et al., 2013). Thus, the conclusions based on correlations between these methods are not necessarily good representations of ecological validity.

Chaytor and Schmitter-Edgecombe (2003) addressed other methodological issues regarding research on ecological validity of PBTs. It is unclear whether the instrument of comparison to the PBT of EF accurately reflects everyday EF functioning, because these methods (usually questionnaire) also has its limitations. Additionally, there is a need for standardized means assessing the relation between everyday function and corresponding PBTs. The current literature makes comparison between studies difficult, because different PBTs and comparison measures are used in different studies (e.g., see the variety of PBTs in the studies assessed by Toplak et al., 2013). However, despite methodological limitations in this area of research, it seems to be a general agreement that PBTs of EF are inadequate in terms of ecological validity.

The Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) was developed in order to address the need for a standardized, ecologically valid PBT of EF for children and adolescents (Emslie et al., 2003). It consists of six subtests investigating different aspects of EF and a questionnaire measure of EF (Dysexecutive Questionnaire for Children; DEX-C). BADS-C aims to assess the so-called dysexecutive syndrome, a term used to describe the constellation of functional deficits related to different aspects of EF (Baddeley, 1986). It is largely influenced by the Working Memory model of Baddeley and Hitch (1974) and the Supervisory Attentional System model of Shallice (1982). Based on these theoretical frameworks and the goal of achieving ecological validity, BADS-C includes open-ended tasks assessing flexibility, novel problem solving, impulsivity, planning, and the ability to use feedback in order to moderate behavior.

BADS-C is one of the most cited (N = 7) assessments during the last decade of research on ecologically valid measures of EF in children (Wallisch et al., 2018). However, research investigating its validity has been regarded as sparse, especially in samples of pABI (Chevignard et al., 2012; Longaud-Valès et al., 2016). Thus, in spite of increased use of the test, more research is needed to determine whether BADS-C is indeed as ecologically valid as intended, or if it shows adequate validity overall.

Several studies have demonstrated discriminant validity (i.e., to what extent a measure of a construct diverges from a theoretically unrelated measure or phenomenon) of BADS-C for different types of neurological conditions. These have included TBI (Chevignard et al., 2010), brain tumors (Longaud-Valès et al., 2016; Ward et al., 2009), attention deficit hyper-activity disorder (ADHD; Shimoni et al., 2012; Siu & Zhou, 2014) and autistic spectrum disorder (White et al., 2009). In these studies, the clinical groups performed worse on BADS-C compared to the normal population, either in terms of a control group or compared to norm data. However, only parts of BADS-C were used for analyses in two of the pABI studies, namely the Six Parts subtest (Chevignard et al., 2010) and BADS-C total score (Ward et al., 2009). The whole test battery was used for in the study of Longaud-Valès et al. (2016), but this was done on a specific type of pABI. Thus, there is no direct evidence of the discriminant validity of the whole battery of BADS-C on a more heterogenous pABI group.

Concurrent validity (i.e., to what degree a measure is similar to established measures of the same construct) has been demonstrated with moderate to strong correlations between BADS-C subtests and established PBTs of EF (r's between .49 and .76) for children and young adults treated for frontal lobe tumors (Longaud-Valès et al., 2016). However, the study had a small sample size (*N* = 21), limiting the generalizability of the findings. In contrast, weak correlations were found between BADS-C and commonly used PBTs of EF in the initial standardization of BADS-C (Baron, 2007). Thus, the mixed evidence regarding the concurrent validity of the BADS-C tests calls for more research, especially with larger sample sizes. However, the concurrent validity of the questionnaire measure of BADS-C (i.e., DEX-C) has been demonstrated in terms of a strong correlation (*r* = .78) with BRIEF (Roy et al., 2015).

The concurrent validity of BADS-C in terms of its relationship with IQ is complex. While EF and IQ are two theoretically different constructs, it has been argued that parts of the constructs are related to each other (Ardila, 2013). This is mainly because they are both tapping into a broad range of cognitive abilities, and that Working Memory is a core component of both constructs. However, the evidence regarding the link between BADS-C and IQ is mixed. Correlations between BADS-C and IQ has ranged from moderate (Emslie et al., 2003) and strong (Longaud-Valès et al., 2016), to weak and mostly non-significant (Roy et al., 2015). Thus, more knowledge is needed to clarify the relationship between BADS-C and IQ.

With respect to ecological validity, some significant relationships were found between BADS-C subtests and DEX-C parent ratings in the initial standardization of BADS-C (Emslie et al., 2003). However, the relationships were generally weak (*r*'s between -.03 and -.22). Another study found non-significant correlations (*r*'s between .01 and .19) between all of the BADS-C subtests and two EF questionnaires (DEX-C and the BRIEF), as rated by parents (Roy et al., 2015). These results are in line with the weak correlations between conventional PBTs and rating measures of EF (Toplak et al., 2013), thus indicating that BADS-C is not a more ecologically valid measure in this respect. However, the results of Roy et al. (2015) were based on a neurologically healthy sample of children with few executive difficulties, which might explain the lack of correlations with BRIEF and DEX-C. Also, the narrow age range (7-13 years) does not capture the potential ecological validity of BADS-C with respect to adolescents.

Other studies have shown findings supporting the ecological validity of the BADS-C to a larger degree. A significant moderate correlation between BADS-C total score and the parent rated Metacognition Index of BRIEF (*r* = -.37) has been shown (Shimoni et al., 2012), and the Six Parts subtest of BADS-C has been significantly weakly correlated with the parent rated DEX-C (*r* = -.28; Siu and Zhou, 2014). These studies indicate some degree of ecological validity of the BADS-C. However, the results are partial (only components of BADS-C and BRIEF) and limited to the ADHD population.

In contrast to the above findings, a more recent study found strong correlations between BADS-C and BRIEF teacher ratings (*r*'s between .60 and .81), and weaker and more non-significant correlations with BRIEF parent ratings (Longaud-Valès et al., 2016). However, the sample sizes were low for BRIEF teacher (n = 11) and parent (n = 16) and heterogenous in terms of age (including both children and young adolescents), representing methodological limitations to this finding. Nevertheless, the difference between strongly correlated teacher ratings and weakly correlated parent ratings, indicates that the ecological validity of BADS-C might vary according to different contexts. It is also the first study examining the ecological validity of BADS-C including teacher questionnaire ratings. However, because of the low sample sizes, this should be investigated further in more representative samples. Also, as BADS-C is intended to be more ecologically valid than traditional PBTs of EF, it should predict questionnaire measures of EF beyond these and IQ. However, this is currently unexplored.

In summary, there is a need for more research investigating several aspects of validity of BADS-C. This includes discriminant, concurrent, and ecological validity in a representative and heterogenous pABI sample, the relationship between BADS-C and IQ, and the predictive capacity of BADS-C on EF questionnaires beyond established PBTs. Also, questionnaires of EF filled out by different types of raters should be included, in order to examine potential variation in ecological validity with respect to different contexts. Therefore, the aim of the present study was to examine these aspects of validity of BADS-C in a sample of children and adolescents with pABI and a healthy control group. Exploring these areas will be important for the further use of BADS-C in research and clinical assessment. Ultimately, it could contribute to determine if BADS-C can be regarded as a better alternative to conventional neuropsychological tests of EF in terms of ecological validity.

Based upon the research reviewed, the hypotheses were as follows:

1. The pABI group will score lower on BADS-C compared to the control group.
2. BADS-C scores will be positively correlated with D-KEFS and IQ.
3. DEX-C parent ratings will be positively correlated with BRIEF parent ratings, and DEX-C teacher ratings will be positively correlated with BRIEF teacher ratings.
4. BADS-C scores will be negatively correlated with parent, teacher, and self-report questionnaire measures of EF (DEX-C and BRIEF). These correlations will be stronger with teacher ratings compared to parent and self-report.

5. BADS-C will predict questionnaire measures of EF (DEX-C and BRIEF) beyond IQ and D-KEFS.

The first hypothesis examined discriminant validity, and the second and third hypotheses examined concurrent validity. The fourth and fifth hypotheses examined ecological validity of the BADS-C.

Method

Participants

The participants were 74 children and adolescents with pABI and 60 healthy controls aged 10-17 years, from a Norwegian intervention study (Cognitive Rehabilitation in paediatric Acquired Brain Injury [CORE pABI]; Hypher et al., 2019). The data was collected pre-intervention. The pABI group consisted of participants with TBI, brain tumor, stroke, anoxia, and brain infection. There was no significant age difference between the pABI group (*M* = 13.43, *SD* = 2.31) and the control group (*M* = 12.75, *SD* = 1.94). Categorical demographic data is presented in Table 1. While the groups did not differ significantly in terms of gender, parents of the control group had significantly higher education than parents of the pABI group.

Table 1
Categorical Demographic Variables for the pABI Group and the Control Group

Measure	*n (%)*		*p*
	pABI (*N* = 74)	Control group (*N* = 60)	
Gender			.435
Female	42 (56.8)	30 (50)	
Male	32 (43.2)	30 (50)	
Mother's education			<.001
Gymnasium or less	27 (38.1)	2 (3.3)	
One or more years of university	44 (61.9)	58 (96.7)	
Father's education			<.001
Gymnasium or less	33 (51.6)	10 (17.2)	
One or more years of university	31 (48.4)	48 (82.8)	
pABI subgroups			
TBI	18 (24.3)		
Tumor	27 (36.5)		
Other	29 (39.2)		

Note. pABI = paediatric Acquired Brain Injury; TBI = Traumatic Brain Injury; Other = Stroke (*n* = 17), brain infection (*n* = 7), and anoxia (*n* = 5).

Inclusion criteria for the pABI group were: at least 12 months since injury/illness, or more than 12 months since finished cancer therapy; and an experience of EF deficits in daily life determined by a self-made, semi-structured clinician interview. Exclusion criteria were: injury acquired before 2 years of age; cognitive, physical, sensory or language impairment affecting the capacity to attend to regular school; neurological disease pre-injury; recently detected relapse in brain tumor; unfit for outcome evaluation (evaluated independently by two investigators); and not fluent in Norwegian.

Inclusion criteria for the control group were signed informed consent; and age between 10 and 18. Exclusion criteria were: brain tumor; brain injury with unconsciousness surpassing 5 minutes; other brain illnesses; serious psychiatric disorders; drug problems; diabetes; and gluten, lactose, or egg allergies.

Materials

Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C)

BADS-C assesses a broad range of EF's in children and adolescents in an ecological manner (Emslie et al., 2003). It was developed as a more child friendly version of the adult-oriented Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1997). Different aspects of EF are examined by six performance-based subtests (Playing Cards Test; Water Test; Zoo Map Tests 1 and 2; Key Search Test; and the Six Parts Test), and a questionnaire measure of EF (DEX-C) which provides additional ecological validity. The DEX-C can be completed by parents, teachers, or other informants who knows the child well.

For the present study, BADS-C and DEX-C were administered to both the pABI group and the control group. For the pABI group, DEX-C was rated by both parents and teachers. For the control group, DEX-C was rated by parents. Raw scores in BADS-C can be converted to scaled scores ranging from 1 to 19, designed to be normally distributed with a mean of 10 and a standard deviation (SD) of 3 (Emslie et al., 2003). The BADS-C total score is obtained by summing the scaled scores from the six subtests. These are converted into an overall scaled score, standardized with a mean of 100 and a SD of 15. Low scores on BADS-C and high scores on DEX-C indicate executive difficulties. The norms are provided for children and adolescents aged 8 to 15 years and 11 months, based on a representative sample of 259 children balanced for gender, mean estimated IQ, and socio-economic background (Emslie et al., 2003). The scaled scores can be adjusted for eight age groups and three IQ groups. In this study, scores were only adjusted for age, not for IQ.

The following descriptions of the BADS-C subtests are shortened from the BADS-C manual (Emslie et al., 2003). Playing Cards test assesses mental flexibility in terms of the ability to change an established pattern of responding. It uses 21 spiral-bound non-picture playing cards that are turned over one at the time. In the first part of the test, participants are instructed to say “yes” to red cards and “no” to black cards. In the second part of the test, the rules are changed and explained to the participants. Now the children are instructed to say “yes” if the card is the same color as the previous card, and “no” if the card is a different color from the previous card. In addition to measuring flexibility, the second task also assesses working memory in the sense of the child’s ability to keep track of the color of the previous card and the changed rule.

Water Test is a novel, practical task that assesses the ability to develop a plan of action in order to solve a problem. The children are asked to physically manipulate different tools and objects, in order to extract a cork from a tube.

Key Search Test examines the ability to plan an efficient, systematic, implementable plan of action, monitoring of one’s own performance, and to take into account factors not explicitly stated. The examinee is presented with a piece of paper with a square in the middle that represents a large field where they have lost their keys. The objective is to draw a pathway through the square, representing a searching strategy with the goal of finding the lost keys in the most efficient manner.

Zoo Map Test 1 is an open-ended test with little structure, examining the ability to plan an adequate route. The examinee is asked to plan a path through a given zoo map, where they shall pass through a list of drawn animals and places according to a set of rules. Zoo Map Test 2 is a more structured version of the Zoo Map Test 1, asking the child to follow more specific rules.

Six Parts Test examines planning, task scheduling and performance monitoring. The examinee is given three different color-coded tasks to perform: a green task (arithmetic), blue test (picture naming), and a red task (sorting). Each of the tasks have two parts. The objective is to schedule one’s time to attempt something from all six parts over a five-minute period, with restrictions on the order in which the parts can be attempted.

The DEX-C is a 20-item questionnaire that can be rated by parents or teachers (Emslie et al., 2003). The items cover four broad areas of potential executive dysfunction: emotional/personality, motivational, behavioral, and cognitive. Each question is sensitive to a specific characteristic of the dysexecutive syndrome. The items are scored on a Likert scale ranging from 1 (“never”) to 5 (“very often”). Examples of items are: “Has difficulty thinking ahead or planning when undertaking a task or activity”; “Acts without thinking”. The DEX-C is regarded as a measure of everyday EF in home and school. Raw total scores were used for analyses in this study, which were compared with scaled scores for the other measures.

Behavior Rating Inventory of Executive Function (BRIEF)

The BRIEF is a standardized rating measure for assessing EF in children and adolescents aged 5-18 years (11-18 years for self-reporting) in home and school environments (Gioia et al., 2000). It was used in this study as a criterion measure for ecological validity of BADS-C, and it can be rated by parents, teachers and as self-report. The BRIEF Global Executive Composite (GEC) is an overarching summary score that is based on different executive component scales (inhibition, shifting, emotional control, initiation, working memory, planning, organization of materials, and monitoring). A total of 86 statements describe different behaviors which is rated by their frequency on a Likert scale ranging from 1 (“rare”) to 3 (“often”). The total scores are converted into standardized T scores with a mean of 50 and a standard deviation of 10. The higher the score, the more difficulties of EF in daily life is indicated.

Within the last decade, BRIEF has been one of the most widely cited (N = 17) rating measures for assessing ecological validity of EF (Wallisch et al., 2018). It shows good psychometric properties, and it stands out as the preferred caregiver rating measure of EF in studies on pABI populations (Roth et al., 2014). In this study, BRIEF GEC was rated by parents, teachers, and as self-report for the pABI group. It was not administered to the control group.

Delis-Kaplan Executive Functioning System (D-KEFS)

D-KEFS is a neuropsychological test battery, which includes nine individually administered tests, covering a full range of EF (Delis et al., 2001). It was used in this study as a criterion measure for concurrent validity of the BADS-C. Norms are provided for both children and adults (ages 8-89 years). Raw scores are converted into scaled scores with a mean of 10 and a standard deviation of 3. Higher scores indicate better performance. D-KEFS is one of the most widely used performance-based test batteries of EF, and it shows good psychometric properties (Stephens, 2014). The test battery is administered in a traditional pen-and-pencil manner, and it is therefore not explicitly made with an effort to provide ecological validity. In this study, two of the subtests from D-KEFS were administered to the pABI group and used for analyses: Color Word Interference Test (condition 3 and 4) and Trail Making Test (condition 4). The reason for this choice was that these conditions measure EF, while the other conditions measure more basic cognitive processing.

Color Word Interference Test measures inhibition of verbal responses by reading words printed in ink which is dissonant to the color of the written word (Delis et al., 2001). This is assessed in condition 3, while condition 4 also involves switching between reading the word or naming the color. These are tests of EF because they involve inhibition and switching (i.e., mental flexibility), which are two of the components of EF (Diamond, 2013). Condition 1 and 2 involves merely reading the written words and naming the ink color, respectively.

Trail Making Test condition 4 measures cognitive flexibility in the visual-motor domain (Delis et al., 2001). The examinee is instructed to write a line sequentially, switching between number and letter (i.e., 1 – A – 2 – B etc.). The other conditions in TMT measure basic cognitive processing, like visual scanning, number/letter sequencing and motor speed, and they are accordingly not direct measures of EF.

Wechsler Intelligence Scale for Children – fifth edition (WISC-V)

The intellectual levels of the participants were assessed using the WISC-V, and it was used in this study for analyses of concurrent validity of the BADS-C. WISC-V is the fifth and most recent edition of an intelligence test for children between the ages of 6 and 16 (Wechsler, 2014), and it is widely validated in terms of psychometric properties (Kaufman et al., 2015). It includes five major indexes: Verbal Comprehension Index, Visual Spatial Index, Fluid Reasoning Index, Working Memory Index, and Processing Speed Index. It is also a possibility to generate a Full Scale IQ, based on these indexes. Raw scores are converted into scaled scores for the subtests, with a mean of 10 and a standard deviation of 3. The indexes and Full Scale IQ are scored based on the subtests, with a mean of 100 and a standard deviation of 15. Higher scores indicate better performance.

In this study, a sample of subtests from WISC-V were administered to the pABI participants in order to examine Verbal Comprehension Index (Similarities and Vocabulary), Working Memory Index (Digit Span and Picture Span), Processing Speed Index (Coding and Symbol Search), Block Design, Matrix Reasoning, and an estimated Full Scale IQ. In the healthy controls, Verbal Comprehension Index was the only index administered, in addition to the Block Design and Matrix Reasoning subtests.

Procedure

The participants with pABI were referred to the CORE pABI study (Hypher et al., 2019) based on hospital discharge diagnosis and record information. The study was approved by the Regional Committees for Medical and Health Research Ethics Norway (2017/772). The trial registration number was NCT03215342. Data were collected between 2018 and 2019 at two test sites in Norway, namely St. Olavs Hospital in Trondheim and Oslo University Hospital, Rikshospitalet. Written informed consent was signed by participants (>16 years) or primary caregivers (<16 years). Primary caregivers were interviewed in order to assess EF symptoms and inclusion/exclusion criteria, that were not included in the hospital records. Participants that had surpassed 16 years of age were also able to attend this interview. The control group was recruited as a convenience sample from local public schools in Trondheim and Oslo. Written informed consent was assessed with the same procedure as with the pABI group. The CORE pABI study was set up to meet the GCP quality and safety standard (Baber, 1994). A secure, web-based system (WebCRF) was used to ensure storage of sensitive information and management of data files.

The test situation differed both between and within the groups. Participants in the control group were tested during a 2-3-hour period, at daytime or evening, and in the hospital or at school. These factors increase the variation in testing environment within the control group. However, the participants were tested by the same test technician, which reduced this variation. Participants in the pABI group were tested at the hospital during the course of one workday (7-8 hours), completing more tests than the control group. In order to alleviate tiredness, frequent breaks for the participants were included during the testing day. The pABI participants were tested at two test sites within the hospital and by different test technicians, increasing the variation in testing environment. However, a Standard Operating Procedure was made as a guide for the test technicians in order to minimize variation in the testing situation. In addition, test technicians were trained by experienced clinical neuropsychologists, and the scores and data were quality checked after testing and before it was entered into WebCRF.

Statistical Analyses

All of the variables used for statistical analyses were checked for normality by examining the ratio between skewness/kurtosis and standard error. Water Test, Zoo Map Test 2, DEX-C teacher, and D-KEFS Color-Word Interference Test 3, were transformed into variables with more acceptable normality distributions.

IBM SPSS statistics (Version 26) was used for statistical analyses. T-tests were performed comparing results between the pABI group and control group with respect to BADS-C, DEX-C parent ratings, and WISC-V (Verbal Comprehension Index, Block design and Matrix Reasoning). Chi square analyses were used for examining potential differences between the pABI group and control group in terms of gender and parental education. Analysis of variance (ANOVA) tests were performed when comparing pABI subgroups and the control group in terms of BADS-C performance. The relationships between BADS-C, DEX-C, BRIEF, WISC-V, and D-KEFS, were examined using Pearson correlations.

Hierarchical multiple regression was performed to investigate the predictive value of BADS-C on questionnaire ratings of EF, beyond established neuropsychological measures. BRIEF GEC and DEX-C were used as dependent variables. FSIQ and the three D-KEFS measures were entered on step 1 and step 2, and BADS-C measures were entered on step 3. Only the BADS-C measures with significant correlations with the dependent variables were used. The variance inflation factor (VIF) and the tolerance statistic were examined, in order to assess multicollinearity. A VIF level ranging between 1 and 10, and a tolerance level above 0.2, were considered adequate (Bowerman & O'connell, 1990; Menard, 2002). Also, the Durbin-Watson test was performed to examine autocorrelation between residuals. A value as close to 2 as possible was considered acceptable, and it should not be less than 1 or larger than 3 (Durbin & Watson, 1950).

Reported *p* values were two-sided. Effect sizes were calculated as Cohen's *d*, based on the pooled standard deviation between the pABI group and the control group (d_{pooled}). The strength of the effect sizes were interpreted based on Cohen's (1988) conventions, namely small (d = 0.2), medium (d = 0.5), and large (d = 0.8).

Some of the participants and caregivers did not complete all of the tests or questionnaires, and no tests or data were added to account for these missing values. Thus, the specific sample sizes for the different measures with missing values were as follows for the pABI group: BADS-C total score and Six Parts Test (N = 73); DEX-C teacher (N = 69);

BRIEF parent (*N* = 73) and teacher (*N* = 69); WISC-V Block Design and Matrix Reasoning (*N* = 73), Verbal Comprehension Index (*N* = 72), Working Memory Index (*N* = 71), and Processing Speed Index and Full Scale IQ (*N* = 70). For the control group, the exact sample sizes were: All of the BADS-C variables (*N* = 59); DEX-C parent (*N* = 58); WISC-V Verbal Comprehension Index (*N* = 59). The remaining measures were completed by all of the participants or caregivers.

Results

Preliminary Analyses

Reliability analyses showed that correlations between the five BADS-C subtests were mostly weak or very weak, both for the control group (*r*'s ranging from -.28 to .38) and the pABI group (*r*'s ranging from -.06 to .30). For the control group, there were significant correlations between Playing Cards Test and Six Parts Test (r = .38, p = .003), and between Water Test and Zoo Map Test 2 (r = -.28, p = .03). All other correlations were non-significant. For the pABI group, only the correlation between Key Search Test and Zoo Map Test 1 was significant (r = .30, p = .01). There were no significant gender differences in BADS-C scores. DEX-C items were found to have a strong internal consistency, both for the parent ratings (α = .94) and the teacher ratings (α = .97).

Group Comparisons

Table 2 presents test and questionnaire scores for the pABI group and the control group, with the exception of BADS-C and DEX-C parent scores, which is presented separately in the next section. The control group had significantly higher scores compared to the pABI group in Verbal Comprehension, Block Design, and Matrix Reasoning. Scores were slightly above average for the control group on these tests. All of the D-KEFS and WISC-V scores were slightly below average for the pABI group. BRIEF GEC teacher and parent scores were slightly higher than average. BRIEF GEC self-report scores were significantly lower than both parent (p = .001; d = .44) and teacher (p = .009; d = .40) ratings.

Table 2

Test and Questionnaire Scores for the pABI Group and the Control Group

Measure	*M (SD)*		*p*
	pABI (*N* = 74)	Control group (*N* = 60)	
WISV-V			
Full Scale IQ	92.60 (13.46)		
Working Memory	94.04 (14.11)		
Processing speed	89.57 (17.62)		
Verbal Comprehension	96.36 (11.95)	108.73 (12.11)	<.001
Block Design	8.55 (3.06)	10.90 (2.34)	<.001
Matrix Reasoning	9.42 (2.90)	11.42 (2.82)	<.001
D-KEFS			
CWIT 3	7.84 (3.23)		
CWIT 4	7.69 (3.61)		
TMT 4	8.01 (3.72)		
BRIEF GEC			
Parent	59.53 (10.59)		
Teacher	59.75 (14.04)		
Self	54.36 (12.84)		
DEX-C Teacher	17.75 (13.46)		

Note. pABI = paediatric Acquired Brain Injury; TBI = Traumatic Brain Injury; Other = Stroke (*n* = 17), brain infection (*n* = 7), and anoxia (*n* = 5); BRIEF = Behavioral Rating Inventory of Executive Function; GEC = Global Executive Composite; WISC-V = Wechsler Intelligence Scale for Children – fifth edition; D-KEFS = Delis-Kaplan Executive Functioning System; CWIT 3 = Color Word Interference Test condition three; CWIT 4= Color Word Interference Test condition four; TMT 4 = Trail Making Test condition four; DEX-C = Dysexecutive Questionnaire for Children.

Table 3 shows group comparisons between the pABI group and control group in terms of BADS-C and DEX-C scores. The control group performed significantly better than the pABI group on all of the measures, except for the Playing Cards Test. The effect sizes ranged from medium to large. ANOVA group comparisons between the pABI subgroups and the control group in terms of BADS-C and DEX-C scores were also performed (not included in table). Analogous to the t-tests in Table 3, the control group performed significantly better than the pABI subgroups with medium to large effect sizes (*d's* ranging from 0.78 to 1.32) on most measures, but there were no significant differences between the pABI subgroups.

Table 3

Comparisons Between the pABI Group and the Control Group on BADS-C and DEX-C

Measure	*M (SD)*		*t*	*p*	d_{pooled}
	pABI	Control group			
BADS-C Total	83.89 (19.94)	102.08 (16.33)	5.64	<.001	1.00
Playing Cards	8.15 (3.68)	9.24 (3.39)	1.75	.082	0.31
Water	10.01 (3.18)	11.42 (2.39)	2.92	.037	0.50
Key Search	10.47 (4.28)	12.27 (3.71)	2.55	.012	0.45
Zoo Map 1	8.31 (4.11)	9.85 (3.55)	2.27	.025	0.40
Zoo Map 2	8.68 (3.5)	10.25 (2.40)	3.08	.003	0.52
Six Parts	7.14 (3.08)	9.20 (2.63)	4.09	<.001	0.72
DEX-C Parent	26.53 (12.37)	12.66 (8.99)	-7.46	<.001	1.28

Note. N_{pABI} = 73-74. $N_{Control\ Group}$ = 58-59. pABI = paediatric Acquired Brain Injury; BADS-C = Behavioural Assessment of the Dysexecutive Syndrome for Children; DEX-C = Dysexecutive Questionnaire for Children.

Correlations Between BADS-C, Questionnaires, and Established PBTs

Table 4 presents correlations between BADS-C, DEX-C, and BRIEF GEC scores for the pABI group. BADS-C total score, Key Search Test, and Zoo Map Test 1 had significant negative correlations with both DEX-C and BRIEF teacher ratings. Correlations were weak to moderate (*r*'s ranging from -.37 to -.46). The BADS-C tests were weakly and mostly non-significantly correlated to DEX parent, and BRIEF parent and self-report ratings. There were significant strong correlations between DEX-C and BRIEF parent ratings, and DEX-C and BRIEF teacher ratings.

Table 4

Correlations Between BADS-C, DEX-C, and BRIEF for the pABI Group

Measure	DEX-C		BRIEF GEC		
	Parent	Teacher	Parent	Teacher	Self
BADS-C Total	-.17	-.38***	-.24*	-.42***	-.10
Playing Cards	-.01	.06	.04	.08	.05
Water	-.12	-.25*	-.19	-.07	-.09
Key Search	-.14	-.38***	-.29*	-.37**	-.09
Zoo Map 1	-.14	-.46***	-.09	-.44***	-.10
Zoo Map 2	-.15	.05	.01	-.14	.14
Six Parts	-.06	-.04	-.08	-.19	-.13
DEX-C					
Parent		.17	.75***	.13	.39***
Teacher	.17		.35**	.77***	.13

Note. $N_{\text{BADS.C}}$ = 73-74. $N_{\text{DEX-C}}$ = 74 (Parent); 69 (Teacher). N_{BRIEF} = 73 (Parent); 69 (Teacher); 74 (Self). pABI = paediatric Acquired Brain Injury; BADS-C = Behavioural Assessment of the Dysexecutive Syndrome for Children; DEX-C = Dysexecutive Questionnaire for Children; BRIEF = Behavioral Rating Inventory of Executive Function; GEC = Global Executive Composite.

*$p < .05$. **$p < .01$. ***$p < .001$

Table 5 presents correlations between BADS-C and DEX-C scores, and D-KEFS and WISC-V scores for the pABI group. There were significantly weak to moderate positive correlations between BADS-C total score, Key Search Test, and Zoo Map Test 1, and all of the D-KEFS and WISC-V scores. Some of the DEX-C parent and teacher ratings were significantly correlated with some of the D-KEFS and WISC-V variables, but most were not significant and the correlations were weak.

Table 5

Correlations Between BADS-C, DEX-C, D-KEFS, and WISC-V for the pABI Group

Measure	D-KEFS			WISC-V Full scale IQ and Index				WISC-V subtests	
	CWIT 3	CWIT 4	TMT 4	FSIQ	WMI	PSI	VCI	BD	MR
BADS-C Total	.45***	.54***	.56***	.60***	.45***	.51***	.31**	.45***	.55***
Playing Cards	.16	.21	.31**	.21	.03	.17	.07	.26*	.25*
Water	.19	.27*	.29*	.15	.03	.18	.04	.11	.30*
Key Search	.31**	.41***	.36**	.57***	.45***	.37**	.26*	.34**	.58***
Zoo Map 1	.36**	.32**	.28*	.43***	.35**	.27*	.29*	.37**	.31**
Zoo Map 2	.10	.10	.01	.07	.08	.03	.12	.05	.06
Six Parts	.15	.30*	.28*	.20	.14	.36**	.14	.15	.10
DEX-C									
Parent	-.16	-.18	-.19	-.21	.01	-.03	-.25*	-.11	-.18
Teacher	-.20	-.25*	-.22	-.31*	-.26*	-.22	-.02	-.23	-.37**

Note. $N_{BADS\text{-}C}$ = 73-74. $N_{\text{DEX-C}}$ = 74 (Parent); 69 (Teacher). $N_{\text{D-KEFS}}$ = 74. $N_{\text{WISC-V}}$ = 70 (FSIQ and PSI); 71 (WMI); 72 (VCI); 73 (BD and MR). pABI = paediatric Acquired Brain Injury; BADS-C = Behavioural Assessment of the Dysexecutive Syndrome for Children; DEX-C = Dysexecutive Questionnaire for Children; D-KEFS = Delis-Kaplan Executive Functioning System; CWIT3 = Color Word Interference Test condition three; CWIT4= Color Word Interference Test condition four; TMT4 = Trail Making Test condition four; WISC-V = Wechsler Intelligence Scale for Children – fifth edition; FSIQ = Full Scale Intelligence Quotient; WMI = Working Memory Index; PSI = Processing Speed Index; VCI = Verbal Comprehension Index; BD = Block Design; MR = Matrix Reasoning

*p < .05. **p < .01. ***p <.001

Predictive Capacity of BADS-C on Questionnaire Ratings of EF

Two regression models were conducted in order to assess the predictive value of BADS-C on questionnaire measures of EF (DEX-C and BRIEF GEC). In the first regression model, Zoo Map Test 1 and Key Search Test were used as predictor variables in step 3. The second model was analogous to the first, with the exception of having BADS-C total score in step 3 instead of Zoo Map Test 1 and Key Search Test. In all of the regression analyses performed, multicollinearity was not an issue. VIF values ranged between 1 and 2.52, and no tolerance level was below 0.2. Also, the Durbin-Watson values were all close to 2, ranging between 1.78 and 2.33. This indicated no issue with autocorrelation in the residuals.

As shown in Table 6, Zoo Map Test 1 and Key Search Test explained significant amounts of the variance of both DEX-C (23%) and BRIEF GEC (12%) teacher ratings, beyond FSIQ and D-KEFS. Both Zoo Map Test 1 and Key Search Test were significant predictors of DEX-C teacher ratings (β = -.45 and β = -.28, respectively). Only Zoo Map Test 1 significantly predicted BRIEF GEC teacher ratings (β = -.34).

BADS-C total score did not predict neither DEX-C or BRIEF GEC teacher ratings, beyond FSIQ and D-KEFS (not included in table). Analogous regression analyses were also performed with BRIEF GEC parent ratings as dependent variable, with none of the BADS-C tests predicting this questionnaire measure of EF.

Table 6

Regression Analyses Predicting DEX-C and BRIEF GEC Teacher Scores for the pABI Group (N = 65)

Step	DEX-C Teacher			BRIEF GEC Teacher		
	R^2	R^2_{Adj}	R^2 *cha*	R^2	R^2_{Adj}	R^2 *cha*
1. FSIQ	.09	.08	.09	.15	.13	.15**
2. D-KEFS	.10	.04	.01	.21	.16	.06
3. Zoo Map 1 Key Search	.33	.26	.23***	.32	.25	.12*
Final step	*B (SE)*	*β*	*t*	*B (SD)*	*β*	*t*
FSIQ	0.16 (0.03)	.12	0.65	0.08 (0.20)	.08	0.41
D-KEFS						
CWIT3	-0.39 (0.47)	-.13	-0.83	-0.75 (3.82)	-.03	-0.20
CWIT4	-0.06 (0.08)	-.12	-0.72	-1.28 (0.68)	-.31	-1.87
TMT4	-0.63 (0.07)	-.13	-0.87	-0.09 (0.59)	-.02	-0.16
Zoo Map 1	-0.20 (0.05)	-.45	-3.72***	-1.22 (0.44)	-.34	-2.75**
Key Search	-0.12 (0.05)	-.28	-2.15*	-0.62 (0.44)	-.19	-1.40

Note. pABI = paediatric Acquired Brain Injury; BRIEF = Behavioral Rating Inventory of Executive Function; GEC: Global Executive Composite; FSIQ = Full Scale Intelligence Quotient; D-KEFS = Delis-Kaplan Executive Functioning System; CWIT 3 = Color Word Interference Test condition three; CWIT 4 = Color Word Interference Test condition four; TMT 4 = Trail Making Test condition four; BADS-C = Behavioural Assessment of the Dysexecutive Syndrome for Children.

*$p < .05$. **$p < .01$. ***$p < .001$

Discussion

This is the first study to examine the discriminant, concurrent, and ecological validity of BADS-C in a representative, heterogenous pABI sample. It is also the first study to examine the ecological validity using a representative sample of teacher questionnaire ratings of EF, in addition to examining the predictive capacity of BADS-C on these ratings after controlling for established neuropsychological EF tests. The results showed that the pABI group performed significantly worse on all of the BADS-C tests compared to the control group, except for the Playing Cards Test. Effect sizes were moderate to large. Further, BADS-C total score, Key Search Test, and Zoo Map Test 1 were significantly weakly to moderately correlated with all of the D-KEFS and WISC-V variables, in addition to with DEX-C and BRIEF teacher ratings. However, the correlations of the other BADS-C subtests were mostly non-significant and weak, and there were almost no significant correlations between any of the BADS-C subtests and parent or self-report ratings of EF. Key Search Test and Zoo Map Test 1 significantly predicted DEX-C and BRIEF teacher ratings, beyond established PBTs.

Discriminant Validity

The lower scores for the pABI group compared to the control group on all of the BADS-C measures except Playing Cards Test, supports the first hypothesis. These results indicate that BADS-C is sensitive to detecting executive dysfunction in children and adolescents with pABI, and this is in line with previous research on discriminant validity of BADS-C for several clinical groups (Chevignard et al., 2010; Shimoni et al., 2012; Siu & Zhou et al., 2014; Ward et al., 2009; White et al., 2009). A non-significant difference for the Playing Cards Test was also found in the study of Shimoni et al. (2012), but another study found it to be significant and with the largest difference of all the BADS-C subtests (Siu & Zhou et al., 2014), both studies on ADHD samples. Thus, the discriminant validity of the Playing Cards Test is mixed. These results can be related to differences in executive functioning between pABI and ADHD populations, and also within the heterogenous ADHD population. However, it can also indicate that the discriminant validity of the Playing Cards Test is not as robust compared to the rest of the BADS-C test battery.

The present study did not find significant differences in BADS-C performance between the pABI subgroups. This indicates that type of injury does not differentiate EF performance. However, relatively low sample sizes in the different groups (TBI: n = 18; tumor: n = 27; other: n = 29) might have led to type 2 errors, thus not capturing potential differences in EF. Although type of injury might not differentiate EF performance, it has been found that injury severity does (Babikian & Asarnow, 2009; Mangeot et al., 2010). However, it has not yet been examined whether BADS-C is able to capture such differences, which merits further research.

Concurrent Validity

In accordance with the second hypothesis, BADS-C total score, Key Search Test, and Zoo Map Test 1, were significantly correlated with all of the D-KEFS and WISC-V measures. Correlations ranged from low to moderate. There were also significant correlations for other BADS-C subtests, but these were few and weak. The results of the present study are in line with the results of Longaud-Valès et al. (2016), who found strong significant correlations between most of the BADS-C measures, IQ and D-KEFS Trail Making Test condition 4. In the present study, Trail Making Test condition 4 was the only measure significantly correlated with all of the BADS-C tests (except Zoo Map Test 2). Thus, the BADS-C test battery overall seems to be concurrently valid with the Trail Making Test condition 4 of D-KEFS, in particular. However, the general pattern shown by the present study indicates concurrent validity for BADS-C total score, Key Search Test, and Zoo Map Test 1, but mixed results regarding the other subtests of BADS-C.

WISC-V does not primarily measure EF, but rather cognitive functions in general. However, the significant correlations between BADS-C total score, Key Search Test, Zoo Map Test 1, and the Working Memory Index of WISC-V, indicate concurrent validity for these BADS-C measures. This can be inferred because Working Memory is one of the core components of EF theoretically (Diamond, 2013), and the constructs have been found to correlate strongly in other research (McCabe et al., 2010). However, the other parts of BADS-C did not show such a relationship. The link between EF and IQ is controversial, which is reflected in a discussion whether the BADS-C scaled scores should be calculated according to IQ or not (Roy et al., 2015), which is the case in the current manual (Emslie et al., 2003). The mixed results of the present study may not clarify this issue, but rather indicate a complex relationship between BADS-C and IQ.

Looking at studies examining the relationship between BADS-C and IQ, the studies showing significant correlations used pABI samples (the present study and Longaud-Valès et al., 2016) or a mixed clinical and normal sample (Emslie et al., 2003). In contrast, the study which found non-significant correlations used a sample representing the normal population

(Roy et al., 2015). Thus, the link between IQ and BADS-C might be stronger in clinical populations compared to the normal population. However, the lack of significant correlations in the study of Roy et al. (2015) can also be attributed to not using a Working Memory Index in their IQ measure, thus losing an IQ component heavily linked with EF (Diamond, 2013; McCabe et al., 2010).

BADS-C assesses a broad range of EF's. Some aspects of EF might be more related to IQ than others (Ardila, 2013), which might be reflected in the different degrees of correlations with IQ between the BADS-C subtests in the present study. Thus, the subtests not significantly correlated with most WISC-V measures (i.e., Playing Cards Test, Water Test, Zoo Map Test 2, and Six Parts Test) does not necessarily lack concurrent validity. The subtests might measure parts of EF not overlapping with IQ, with the results indicating discriminant validity instead. However, the same BADS-C subtests were not correlated with most of the D-KEFS measures, which indicates a lack of concurrent validity. Yet, as EF is a broad construct, it is unclear whether these subtests are concurrently valid with other PBTs of EF than the D-KEFS measures used in this study. Thus, more knowledge is needed in order to determine the concurrent validity for the specific BADS-C subtests in terms of the relationship with IQ.

The strong correlations between DEX-C parent and BRIEF parent ratings, and between DEX-C teacher and BRIEF teacher ratings, indicate strong concurrent validity for the DEX-C. This supported the third hypothesis of the present study, and it is in line with results from Roy et al. (2015). However, the correlations between parent and teacher ratings were only low to moderate. Moderate inter-rater agreement between parents and teachers in BRIEF has also been shown in other research, which has been attributed to expectations for different environmental settings (Gioia et al., 2000). In the present study, DEX-C showed adequate internal consistency for both the parent ratings (α = .94) and the teacher ratings (α = .97), also in line with Roy et al. (2015). Thus, these results suggest preferable psychometric properties for the DEX-C.

Ecological Validity

With respect to ecological validity, BADS-C total score, Key Search Test, and Zoo Map Test 1, were significantly weakly to moderately correlated with DEX-C and BRIEF teacher ratings. These results supported our fourth hypothesis for these BADS-C measures, but not for the remaining subtests. However, correlations with parent or self-report ratings were mostly weak and non-significant for all of the BADS-C measures. These results are in line with the results of Longaud-Valès et al. (2016), thus indicating ecological validity for the BADS-C in terms of teacher rather than parent ratings.

The present study also examined the ecological validity of BADS-C one step further than Longaud-Valès et al. (2016), finding that Key Search Test and Zoo Map Test 1 predicted DEX-C and BRIEF teacher ratings of EF, beyond D-KEFS and IQ. This is in line with our fifth hypothesis. The results support the evidence of ecological validity of Key Search Test and Zoo Map Test 1 further than the mere correlations with DEX-C and BRIEF teacher ratings, because they explain variation in these questionnaires even after controlling for conventional PBTs. However, BADS-C total score did not predict DEX-C and BRIEF teacher ratings of EF. This can be explained by the fact that BADS-C total score is based on the results from all of the BADS-C subtests, including the subtests that did not correlate well with DEX-C and BRIEF teacher ratings.

The present study stands out from the majority of the literature examining the ecological validity of the BADS-C (Emslie et al., 2003; Roy et al., 2015) and other PBTs of EF (Toplak et al., 2013), which has almost exclusively examined parent questionnaire ratings as the criterion measure for ecological validity. Of the 13 studies that used BRIEF reviewed by Toplak et al. (2013), only one study included teacher ratings of EF. If looking at parent ratings only, the present study converges with the majority of evidence suggesting weak ecological validity for BADS-C and other PBTs of EF. However, the results are notably different when including teacher ratings, indicating that BADS-C is ecologically valid in terms of teacher perceptions of EF.

Regarding the weak correlations between BADS-C and self-report ratings, this is in line with a clear pattern in the literature showing that self-report is a weak measure of cognitive performance compared to clinician or informant ratings (Chaytor & Schmitter-Edgecombe, 2003). This can be related to denial, memory deficits, or lack of self-awareness that often follows brain damage, or other common self-report biases, such as social desirability, recall bias, context effects, or misunderstandings.

The stronger associations between BADS-C and teacher ratings of EF compared to parent ratings, suggest that the ecological validity of the BADS-C varies according to different contexts. These contexts present different environmental challenges to the child, with school arguably presenting more challenges to EF in terms of schoolwork, cooperation in the classroom, and social interactions with other pupils. As BADS-C is relatively similar to tasks given to children in school (Emslie et al., 2003), it should not be surprising that the correlations are stronger with teacher questionnaire ratings of EF compared to parent ratings. In addition, the teacher observes the child in a more neutral way (i.e., with no familial relationship). They also have a larger frame of reference in what to expect of the child in terms of EF performance, as they are able to compare the child with the other pupils in the classroom. However, it should be noted that many of the participants in this study, because of the impairments of their pABI condition, attend to more specialized teaching situations. Thus, the results based on this sample does not represent the school situation for the average pupil. Nevertheless, the results of the present study indicate that a measure can be more ecologically valid in one context rather than the other. These considerations are in line with the prompt to take into account the specific cognitive demands of the patients' environment when assessing the ecological validity of a PBT (Chaytor & Schmitter-Edgecombe, 2003). Following this line of thinking, including both parent and teacher perspectives on child executive functioning have also been addressed as important by other researchers (Longaud-Valès et al., 2016).

Limitations

There are several limitations for the present study that needs to be addressed. As there is no commonly agreed upon measure of injury severity for the pABI population as a whole, we were unable to examine whether BADS-C is sensitive to capturing different degrees of impairment. As EF difficulties increases with injury severity (Babikian & Asarnow, 2009; Mangeot et al., 2010), examining whether BADS-C is sensitive to this differentiation can be regarded as an important validity indicator, which is currently unexplored.

Another issue is that parents of the participants in the control group had more education compared to parents of the pABI group. Also, the IQ levels of the controls were above average. Thus, the control group seems to be biased towards being more resourceful compared to the normal population. This might have caused artificial differences between the pABI group and the control group. However, the effect sizes were moderate to large, indicating that the differences were large enough to also be present using an unbiased control group.

Differences in testing situation can also have affected the group comparison between pABI and controls, in addition to the results in other respects. Considering the pABI participants had a substantially longer testing period than the control group (7-8 hours vs 2-3 hours), they might have been more fatigued by the testing. This might have led to systematic bias. However, participants in the control group could have become tired from other factors, such as being tested in the evening after a regular school day. Other factors increased the testing variation between and within the groups, such as time of day, test location, test sites and different test technicians. This might have led to random measurement error. However, several means were done in order to reduce systematic bias and variation in terms of differences in test situation (see Procedure section).

Another limitation concerns the fact that DEX-C scores were raw scores, which were correlated with scaled scores of BADS-C and BRIEF. This might have led to systematic bias, because scaled scores are adjusted for age, while raw scores are not. Thus, the results involving the concurrent and ecological validity based on the results of DEX-C should be interpreted with caution. However, while the strength differed, the significance and valence of the results of the correlations between BADS-C and DEX-C mirrored the correlations between BADS-C and BRIEF, where the latter used both scaled scores. Thus, the DEX-C results had the same tendency as BRIEF, but when reviewing the exact strength of the correlations or predictions, the BRIEF results should be emphasized.

Regarding concurrent validity, there were relatively few PBTs assessing a limited range of executive functions that were compared to BADS-C, which assesses a broad range of EF's. The concept of EF is complex and understood in different ways by different researchers and theories (Chan et al., 2008; Welsh et al., 2006), which is mirrored in the wide number of existing PBTs of EF, targeting different aspects of the construct. For example, the D-KEFS battery also includes other tests that measure aspects of EF not being captured by Color-Word Interference Test or Trail Making Test (e.g., Verbal Fluency or Tower of London; Delis et al., 2001), which could have been interesting to include in this study. This notion is supported by evidence of BADS-C being more concurrently valid to some EF tests than others, for example by strong correlations with Trail Making Test but weak correlations with Tower of London (Longaud-Valès et al., 2016). If we had the possibility of including other PBTs of EF, it would be easier to compare results directly with other studies using the same PBTs. But ultimately, more agreement concerning the construct of EF, and more standardized means of assessing it, will make comparisons of results across studies less problematic.

There is no objective target measure of ecological validity, which means that it is unclear to what degree the measures of comparison to BADS-C are biased. Questionnaire measures are limited by its psychometric properties or the bias of the informant completing it

(Chaytor & Schmitter-Edgecombe, 2003). Despite this, the BRIEF which was used in this study has good psychometric properties (Roth et al., 2014), and it is one of the most widely used measures for research on ecological validity (Wallisch et al., 2018). Yet, the method of operationalizing ecological validity by correlating PBTs with questionnaire measures of EF in the first place, has been criticised. It has been argued that these types of measures are capturing different aspects of cognitive functioning (Toplak et al., 2013), which makes the premise of measuring ecological validity in this way questionable. Nevertheless, it has been argued that combining measures examining different aspects of EF by different methods, still provides a more comprehensive view of EF functioning than merely using PBTs (Chevignard et al., 2012).

Implications

BADS-C seems to be a valid neuropsychological test that captures EF among children and adolescents with pABI, in line with teacher perceptions of their executive functioning. Thus, it seems like a promising tool in clinical practice for assessing EF that is generalizable to the school context for pABI patients. However, this applies only to the BADS-C total score, Key Search Test, and Zoo Map Test 1. Thus, the differences in validity between different subtests point to evaluating whether the whole battery or the apparently more valid subtests should be used in clinical practice. Although BADS-C seems to be ecologically valid according to teacher perceptions, more knowledge is needed with respect to understanding its role in other contexts of everyday life.

The weak and mostly non-significant correlations between the BADS-C subtests, indicate that they measure different aspects of executive functioning. However, it could also be interpreted as a lack of internal reliability between the subtests. Thus, the potential weak internal reliability combined with the different degrees of validity between the subtests might imply that using BADS-C total score can be problematic in clinical practice, even though it showed adequate discriminant, concurrent, and ecological validity by itself.

Future research should examine the aspects of validity measured in the present study for other populations. This can contribute to determine whether the findings are specific to the pABI population, general to the normal population, or present in other clinical groups. We will also encourage future research on ecological validity to include instruments measuring EF in different contexts of everyday life for comparison to PBTs of EF. This will provide a more comprehensive understanding of ecological validity. Future research should also examine the discriminant validity of BADS-C with respect to injury severity of pABI. The relationship between IQ and BADS-C should be further investigated comparing clinical and healthy samples, in addition to examining potential differences between BADS-C subtests with respect to overlap with IQ. The concurrent validity should be examined further comparing BADS-C with established PBTs measuring other aspects of EF than the ones used in the present study. Finally, the validity of the BADS-C in children and adolescents with pABI should be examined using longitudinal designs.

Conclusion

The findings of the present study indicate that BADS-C is a solid tool in terms of discriminant, concurrent, and ecological validity in relation to the pABI population. While the results showed discriminant validity for the battery as a whole (except Playing Cards Test), BADS-C total score, Key Search Test and Zoo Map Test 1 stand out as the most valid measures with respect to concurrent and ecological validity. These measures seem to be ecologically valid in terms of the school context based on teacher perceptions of the children's and adolescents' EF, but not in terms of parent- or self-perceptions. Thus, the results of this study encourage further use of BADS-C in research and clinical practice with respect to measuring EF in an ecologically valid way, but more research is needed in order to achieve a comprehensive understanding of the validity of BADS-C.

References

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology*, *8*(4), 231-240. https://doi.org/10.1076/chin.8.4.231.13509

Ardila, A. (2013). Development of metacognitive and emotional executive functions in children. *Applied Neuropsychology: Child*, *2*(2), 82-87. https://doi.org/10.1080/21622965.2013.748388

Arnett, A. B., Peterson, R. L., Kirkwood, M. W., Taylor, H. G., Stancin, T., Brown, T. M., & Wade, S. L. (2013). Behavioral and cognitive predictors of educational outcomes in pediatric traumatic brain injury. *Journal of the International Neuropsychological Society: JINS*, *19*(8), 881–889. https://doi.org/10.1017/S1355617713000635

Baber, N. (1994). International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use (ICH). *British journal of clinical pharmacology*, *37*(5), 401–404. https://doi.org/10.1111/j.1365-2125.1994.tb05705.x

Baron, I. S. (2007). *Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C) by Emslie, H., Wilson, F. C., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003)*. Harcourt Assessment/The Psychological Corporation. https://doi.org/10.1080/09297040601112781

Babikian, T., & Asarnow, R. (2009). Neurocognitive outcomes and recovery after pediatric TBI: meta-analytic review of the literature. *Neuropsychology*, *23*(3), 283-286. https://doi.org/10.1037/a0015268

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press. http://dx.doi.org/10.1016/s0079-7421(08)60452-1

Baddeley, A. D. (1986). *Working memory*. Clarendom Press.

Bowerman, B. L., & O'connell, R. T. (1990). *Linear statistical models: An applied approach*. Brooks/Cole.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology*, *23*(2), 201-216. https://doi.org/10.1016/j.acn.2007.08.010

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review*, *13*(4), 181-197. https://doi.org/10.1023/B:NERV.0000009483.91468.fb

Chevignard, M. P., Catroppa, C., Galvin, J., & Anderson, V. (2010). Development and evaluation of an ecological task to assess executive functioning post childhood TBI: the children's cooking task. *Brain Impairment*, *11*(2), 125-143. https://doi.org/10.1375/brim.11.2.125

Chevignard, M. P., Soo, C., Galvin, J., Catroppa, C., & Eren, S. (2012). Ecological assessment of cognitive functions in children with acquired brain injury: A systematic review. *Brain injury*, *26*(9), 1033-1057. https://doi.org/10.3109/02699052.2012.666366

Ciuffreda, K. J., Kapoor, N., Taub, M., Bartuccio, M., & Maino, D. (2012). Acquired brain injury. In M. Bartuccio & D. M. Maino (Eds), *Visual diagnosis and care of the patient with special needs* (pp. 95-100). Lippincott Williams & Wilkins.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Academic Press. https://doi.org/10.4324/9780203771587

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis Kaplan Executive Function System.* Psychological Corporation.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*(1), 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, *37*(3/4), 409-428. https://doi.org/10.2307/2332391

Emslie, H., Burden, V., Nimmo-Smith, I., Wilson, B. A., & Wilson, C. (2003). *Behavioural assessment of the dysexecutive syndrome for children*. Thames Valley Test Company.

Fulton, J. B., Yeates, K. O., Taylor, H. G., Walz, N. C., & Wade, S. L. (2012). Cognitive predictors of academic achievement in young children 1 year after traumatic brain injury. *Neuropsychology*, *26*(3), 314-322. http://dx.doi.org/10.1037/a0027973

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238. https://doi.org/10.1076/chin.6.3.235.3152

Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental neuropsychology*, *25*(1-2), 135-158. https://doi.org/10.1080/87565641.2004.9651925

Hypher, R. E., Brandt, A. E., Risnes, K., Rø, T. B., Skovlund, E., Andersson, S., Finnanger, T. G., & Stubberud, J. (2019). Paediatric goal management training in patients with acquired brain injury: study protocol for a randomised controlled trial. *BMJ open*, *9*(8), e029273. http://dx.doi.org/10.1136/bmjopen-2019-029273

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review*, *17*(3), 213-233. https://doi.org/10.1007/s11065-007-9040-z

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2015). *Intelligent testing with the WISC-V*. John Wiley & Sons.

Koekkoek, S., de Sonneville, L. M., Wolfs, T. F., Licht, R., & Geelen, S. P. (2008). Neurocognitive function profile in HIV-infected school-age children. *European journal of paediatric neurology*, *12*(4), 290-297. https://doi.org/10.1016/j.ejpn.2007.09.002

Koskiniemi, T. N. M. (1999). Cognitive indicators of vocational outcome after severe traumatic brain injury (TBI) in childhood. *Brain Injury*, *13*(10), 759-766. https://doi.org/10.1080/026990599121151

Longaud-Valès, A., Chevignard, M., Dufour, C., Grill, J., Puget, S., Sainte-Rose, C., Valteau-Couanet, D., Dellatolas, G. (2016). Assessment of executive functioning in children and young adults treated for frontal lobe tumours using ecologically valid tests. *Neuropsychological rehabilitation*, *26*(4), 558-583. https://doi.org/10.1080/09602011.2015.1048253

Long, B., Anderson, V., Jacobs, R., Mackay, M., Leventer, R., Barnes, C., & Spencer-Smith, M. (2011). Executive function following child stroke: The impact of lesion size. *Developmental neuropsychology*, *36*(8), 971-987. https://doi.org/10.1177%2F0883073810380049

Mangeot, S., Armstrong, K., Colvin, A. N., Yeates, K. O., & Taylor, H. G. (2002). Long-term executive function deficits in children with traumatic brain injuries: Assessment using the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology*, *8*(4), 271-284. https://doi.org/10.1076/chin.8.4.271.13503

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222-243. http://dx.doi.org/10.1037/a0017619

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

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

Muscara, F., Catroppa, C., & Anderson, V. (2008). Social problem-solving skills as a mediator between executive function and long-term social outcome following paediatric traumatic brain injury. *Journal of neuropsychology*, *2*(2), 445-461. https://doi.org/10.1348/174866407X250820

Nybo, T., Sainio, M., & Müller, K. (2004). Stability of vocational outcome in adulthood after moderate to severe preschool brain injury. *Journal of the International Neuropsychological Society*, *10*(5), 719-723. https://psycnet.apa.org/doi/10.1017/S1355617704105109

O'Keeffe, F., Liégeois, F., Eve, M., Ganesan, V., King, J., & Murphy, T. (2014). Neuropsychological and neurobehavioral outcome following childhood arterial ischemic stroke: attention deficits, emotional dysregulation, and executive dysfunction. *Child Neuropsychology*, *20*(5), 557-582. https://doi.org/10.1080/09297049.2013.832740

Parrish, J., Geary, E., Jones, J., Seth, R., Hermann, B., & Seidenberg, M. (2007). Executive functioning in childhood epilepsy: parent-report and cognitive assessment. *Developmental Medicine & Child Neurology*, *49*(6), 412-416. https://doi.org/10.1111/j.1469-8749.2007.00412.x

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2014). Assessment of executive functioning using the Behavior Rating Inventory of Executive Function (BRIEF). In *Handbook of executive functioning* (pp. 301-331). Springer. https://doi.org/10.1007/978-1-4614-8106-5_18

Roy, A., Allain, P., Roulin, J. L., Fournet, N., & Le Gall, D. (2015). Ecological approach of executive functions using the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C): Developmental and validity study. *Journal of clinical and experimental neuropsychology*, *37*(9), 956-971. https://doi.org/10.1080/13803395.2015.1072138

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298*(1089), 199-209. https://doi.org/10.1098/rstb.1982.0082

Shimoni, M. A., Engel-Yeger, B., & Tirosh, E. (2012). Executive dysfunctions among boys with Attention Deficit Hyperactivity Disorder (ADHD): Performance-based test and parents report. *Research in developmental disabilities*, *33*(3), 858-865. https://doi.org/10.1016/j.ridd.2011.12.014

Siu, A. F., & Zhou, Y. (2014). Behavioral assessment of the dysexecutive syndrome for children: An examination of clinical utility for children with attention-deficit hyperactivity disorder (ADHD). *Journal of child neurology*, *29*(5), 608-616. https://doi.org/10.1177%2F0883073813516191

Stephens, T. L. (2014). The assessment of executive functioning using the delis-kaplan executive functions system (D-KEFS). In *Handbook of executive functioning* (pp. 209-222). Springer. https://doi.org/10.1007/978-1-4614-8106-5_13

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

Wallisch, A., Little, L. M., Dean, E., & Dunn, W. (2018). Executive function measures for children: A scoping review of ecological validity. *OTJR: occupation, participation and health*, *38*(1), 6-14. https://doi.org/10.1177%2F1539449217727118

Ward, C., Phipps, K., de Sousa, C., Butler, S., & Gumley, D. (2009). Treatment factors associated with outcomes in children less than 3 years of age with CNS tumours. *Child's Nervous System*, *25*(6), 663-668. https://doi.org/10.1007/s00381-009-0832-8

Wechsler, D. (2014). Wechsler intelligence scale for children-fifth edition. Pearson

Welsh, M. C., Friedman, S. L., & Spieker, S. J. (2006). Executive Functions in Developing Children: Current Conceptualizations and Questions for the Future. In K. McCartney & D. Phillips (Eds.), *Blackwell handbooks of developmental psychology* (pp. 167-187). Blackwell Publishing. https://psycnet.apa.org/doi/10.1002/9780470757703.ch9

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on “open-ended” executive function tests in autism. *Autism Research*, *2*(3), 138-147. https://doi.org/10.1002/aur.78

Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. In P. Rabbitt, (Ed.), *Methodology of frontal and executive function* (pp. *232*-239). Psychology Press.

NTNU

Kunnskap for en bedre verden