Revista Interamericana de Psicología/Interamerican Journal of Psychology
2025, Vol., 59, e2186

Evidence of Validity Based on the Content of the Child Executive Function Assessment Battery (BAFE-inf)

Andriely Cordeiro [1] **Analice Maria da Silva Albuquerque, Antonio Simeão Sobrinho Neto, Paloma Victória de Sales Alves, & Carla Alexandra da Silva Moita Minervino** [2]

Universidade Federal da Paraíba, João Pessoa, Brasil.

ABSTRACT

The objective of the study was to analyze the validity based on the content of the Child Executive Function Assessment Battery, composed of three computerized instruments that assess inhibition, working memory and cognitive flexibility, in addition to a scale based on parents' perception of children's difficulties in executive function. Seven experts and 22 people from the target population participated (15 children aged six to 10 years and seven parents). They evaluated the clarity, relevance and design/layout of the items. Data collection was remote with experts, and in person with children and parents. The application was individual. The data were analyzed descriptively and by estimating agreement and content validity, with percentages of agreement and appropriate coefficients, above 0.90 and 0.80, respectively. Suggestions and modifications to the instruments were discussed, it was concluded that the battery demonstrated evidence of content-based validity. It was concluded that the battery demonstrated adequate evidence of content validity, and this study stands out for transparently presenting the first and fundamental process of validity analysis, which is essential for the instrument to progress to subsequent stages and achieve effective application in the future.

Keywords
executive function, neuropsychological tests, neuropsychology, psychometrics, reproducibility of results

RESUMO

O objetivo do estudo foi analisar a validade baseada no conteúdo da Bateria de Avaliação das Funções Executivas Infantis, composta por três instrumentos informatizados que avaliam inibição, memória de trabalho e flexibilidade cognitiva, além de uma escala baseada na percepção dos pais sobre as dificuldades das crianças na função executiva. Participaram sete especialistas e 22 pessoas da população-alvo (15 crianças de seis a 10 anos e sete pais). Eles avaliaram a clareza, relevância e design/layout dos itens. A coleta de dados foi remota com especialistas, e presencial com crianças e pais. A aplicação foi individual. Os dados foram analisados descritivamente e por estimativa de concordância e validade de conteúdo, com percentuais de concordância e coeficientes adequados, acima de 0,90 e 0,80, respectivamente. Sugestões e modificações nos instrumentos foram discutidas. Concluiu-se que a bateria apresentou adequadas evidências de validade de conteúdo, e este estudo se destaca por evidenciar de forma transparente o primeiro e fundamental processo de análise de validade, essencial para que o instrumento avance para as etapas subsequentes e alcance uma aplicação eficaz no futuro.

Palavras-chave
função executiva, testes neuropsicológicos, neuropsicologia, psicometria, reprodutibilidade dos resultados

[1] Correspondence about this article should be addressed **Andriely Cordeiro**: andrielyacademico@gmail.com

[2] ***Conflicts of Interest*:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Evidência de validade com base no conteúdo da bateria de avaliação da função executiva infantil (BAFE-inf)

Introduction

Executive functions (EF) are essential skills for directing behaviors and actions to achieve goals (Canet-Juric et al., 2021; Elage & Seabra, 2021; Nogues & Nunes, 2023; Perone et al., 2021). They are related to physical and mental health, academic and social performances, and other cognitive abilities (Chami et al., 2022; Dai et al., 2021; Jiang et al., 2022; Likhitweerawong et al., 2022; Nogues & Nunes, 2023; Pereira et al., 2020; Ruffini et al., 2021). It is crucial to assess whether they progress as expected from early years of life.

The scarcity of executive function (EF) assessment instruments with evidence of validity and reliability has been highlighted in several studies (Berardi et al., 2021; Kusi-Mensah et al., 2022b; Santana et al., 2019), especially in low- and middle-income countries (Kusi-Mensah et al., 2022a). Beyond Brazil, this scarcity is also observed in Latin American countries such as Argentina, Chile, Colombia, and Mexico, where instruments adapted from high-income contexts predominate (Canet-Juric et al., 2021; Kusi-Mensah et al., 2022b).

The BAFE-inf offers a regional contribution by supporting rigorous and culturally relevant psychometric development in Latin America. Consistent with this shortage, a 2020 systematic review identified 37 EF measures available for Brazilian children, but these assessed only working memory and inhibition, excluding cognitive flexibility (Guerra et al., 2022). In this context, combining performance-based measures and parental perceptions is particularly relevant, as socioeconomic inequalities may influence children's engagement and familiarity with formal assessment demands. This integration increases ecological validity and supports more accurate clinical and educational decision-making.

Three fundamental components constitute EF: inhibition, working memory, and cognitive flexibility (Diamond, 2013, 2020; Doğru et al., 2023; Nweze & Nwani, 2020; Souissi et al., 2022; Theodoraki et al., 2020). Therefore, it is essential that assessment measures include these processes and can be applied simply, sensitively, and quickly (Canet-Juric et al., 2021; Kusi-Mensah et al., 2022b).

In executive function assessments, few tools are authorized for use by psychologists, and of the 13 authorized measures, only eight are designed for children

(Guerra et al., 2022). Many are adaptations of adult tests (Venturieri et al., 2023), not considering childhood developmental stages or motivations, and have small normative samples by age group (Guerra et al., 2022). A study conducted in Brazil investigating psychologists' use of psychological tests revealed that, out of 248 instruments used, 105 were not included in the SATEPSI-approved list, indicating a substantial use of unapproved tests (Reppold et al., 2020).

Computerized instruments for assessing basic EF are even rarer, both in Brazil and globally (Martins, 2020). Identified measures include the Visual Attention Test (TAVIS-4), Revised D2 Test, Magic Cards Game, Closet Arrangement Test, Semantic Generation Test, Conners Kiddie Continuous Performance Test, and Continuous Performance Task (Guerra et al., 2022). However, many do not meet psychometric evidence criteria (Martins, 2020), which motivates the creation of new computerized tasks.

The literature discusses that many instruments used to assess EF are performance tests. In Argentina, Canet-Juric et al. (2021) justify the creation of a new EF assessment tool by reflecting on the main components and commonly employed assessment methods, highlighting that although these measures are internally valid, they face ecological validity limitations. In Brazil, all instruments approved by SATEPSI to assess EF are performance tests. Although rating instruments such as scales and questionnaires are used in research—for example, the Childhood Executive Functioning Inventory (CHEXI), the Behavior Rating Inventory of Executive Function (BRIEF), and the Working Memory Rating Scale—none have been approved for clinical use (Guerra et al., 2022).

Questionnaires or inventories are structured and objective techniques that do not have right or wrong answers and assess behavior through statements with predetermined response options (Canet-Juric et al., 2021). These instruments have high ecological validity and, together with performance tests, offer balanced advantages and disadvantages (Berardi et al., 2021). Moreover, the geopolitical, socioeconomic, and cultural diversity directly influences the assessment of EF (Berardi et al., 2021).

The generalization of standardization data to different locations is hindered, especially in Brazil, where there is significant socioeconomic disparity between regions. Standardization research is generally conducted in the South and Southeast regions, so the interpretation of results in other areas should be approached with caution (Guerra et al., 2021; Guerra et al., 2022).

Considering these challenges, the Child Executive Functions Assessment Battery (Bateria de Avaliação das Funções Executivas Infantis: BAFE-inf) was developed, which contains three computerized instruments for direct application with children. These are performance tasks with playful elements and game formats that take into account childhood characteristics, along with a scale completed by parents to investigate everyday EF difficulties. The data collection took place in the Northeast region of Brazil.

For the inhibition assessment test, the Simon (Cevada et al., 2019; Simon & Rudell, 1967) and Go/no Go (Brocki & Bohlin, 2010) paradigms were used as the basis. Two inhibition components were considered: response inhibition and interference suppression (Isherwood et al., 2023; Wang et al., 2024). The working memory test was based on the multi-component model proposed by Baddeley (2000, 2012), including a part that recruits only fluid systems and a part that requires the joint action of fluid and crystallized systems. Finally, the cognitive flexibility test was based on the task switching paradigm (Jersild, 1927) and inspired by the trail making test as described by Bowie and Harvey (2006).

For the assessment tools of EF to be valuable, psychometric evidence must be analyzed (Berardi et al., 2021). Content validity evidence checks how well the items reflect the construct and involves expert item review to ensure their suitability for the instrument, thus requiring expertise in the research field (AERA et al., 2014; Almanasreh et al., 2018).

Experts' responses are analyzed both quantitatively and qualitatively. Quantitative analyses include content validity indexes and overall agreement calculations (Alexandre & Coluci, 2011; Almanasreh et al., 2018). In addition to expert reviews, analyzing the instruments with the target population, using validity and agreement indexes, proves beneficial (Leal, 2022). The goal of this study was to investigate the agreement and content validity measures of the BAFE-inf, considering the expert analyses and the assessments of the instruments by the target population.

Method

Participants

Seven experienced research experts evaluated the BAFE-inf, with at least two experts per subtest, not exceeding ten, as recommended by Almanasreh et al. (2018). The experts had between three and 25 years of experience in the field, and some evaluated

multiple instruments, with each instrument being reviewed by three experts. In addition, 15 children aged six to 10 (nine girls and six boys) from public schools participated, reviewing the instructions, items, response options, and difficulty level of the computerized instruments. Seven parents (five mothers and two fathers, aged 31 to 51 years), responsible for children aged six to 10, evaluated the items of a scale on everyday executive function, which is part of the BAFE-inf.

Inclusion criteria: Judges: a) sign the Informed Consent Form (ICF); b) work in and/or conduct research on neuropsychology, specifically EF, and/or psychometrics; c) be affiliated with an educational or research institution. Children: a) be within the pre-established age range; b) have parental authorization via the ICF; c) sign the Assent Form (AF). Parents: a) consent to the use of data through the ICF; b) reside with the child in the same physical space.

Exclusion criteria: Judges: a) fail to fully complete the subtest analysis questionnaire. Children: a) have a psychiatric or neurodevelopmental disorder that would prevent responding as requested; b) have uncorrected visual and/or auditory problems. Parents: a) have a severe cognitive level or pathology that would prevent responding to the questionnaires.

Design

This is a psychometric study with a cross-sectional design using a non-probabilistic sample. The variables analyzed were clarity, relevance, and the design of the instruments, through the adjusted content validity coefficient and the percentage of agreement of the sample.

Materials

Sociodemographic and Health Considerations Questionnaire

Completed by the parents/guardians of the children. It covered information such as age, sex, school year, residential area, previous access to electronic devices, psychiatric or neurodevelopmental diagnoses, and uncorrected visual or auditory problems. It included questions about the respondent's relationship to the child and whether any parent has a psychiatric or neurodevelopmental disorder.

Sociodemographic Questionnaire for Judges

Developed by the author. It includes questions for characterizing the sample regarding educational background, profession, educational institution, field of study or research, years of experience in the field, and region of residence.

BAFE-inf subtests: Organized Backpacks (Mochilas Organizadas - MO)

Based on the Simon task (Cevada et al., 2019; Simon & Rudell, 1967) and the Go/No Go paradigm (Brocki & Bohlin, 2010), this subtest assesses inhibitory control skills. Implemented on a computer, it does not require the use of a mouse. In the test, children assist the character Julia in organizing items into virtual backpacks: classroom materials go into an orange backpack (key L) and sports materials into a blue backpack (key A). Backpacks are located at the lower part of the screen, while objects appear at the top. The test comprises 40 items, 25% of which are "No Go" items. Of the "Go" items, 15 are congruent (the object appears on the same side as the corresponding backpack) and 15 are incongruent (the object appears on the opposite side). For "No Go" items, children must inhibit their response, not touching anything when, instead of an object, a cat appears on the screen. Each item remains on the screen until the child responds, for up to 2000 milliseconds. Between items, a small cross appears at the center of the screen for 800 milliseconds followed by a blank screen for 250 milliseconds. The test records response time, total correct responses, and errors, categorized into substitution errors (choosing the wrong backpack), omission errors (not responding when needed), and false alarms (responding when not required). The maximum score is 40.

Computerized Working Memory Task for Children (Tarefa Informatizada de Memória de Trabalho para Crianças - TIMTraC)

This task, designed as a game, assesses visuospatial working memory using items that include animal figures, transportation means, and landscapes (Cordeiro et al., 2019; Minervino & Tomaz, 2021). It is performed on a computer and requires the use of a mouse. The task consists of two stages: in the first (fluid working memory), the child memorizes images for five seconds and then identifies them in the presented order; in the

second (integrated working memory), memorizes and positions elements according to rules. Each stage features eight items of increasing difficulty, one training item and seven test items, with the training item not being scored. Each item offers two attempts with different stimuli. If the child fails both, the subtest is terminated. Scoring for each item varies from zero (fails both attempts), one (succeeds on one attempt), to two (succeeds on both attempts). The maximum score is 28 points. TIMTraC evaluates correct responses, errors, raw scores for each stage, total score, and the longest correct stimulus sequence in each stage.

Going Home (Voltando para casa - VPC)

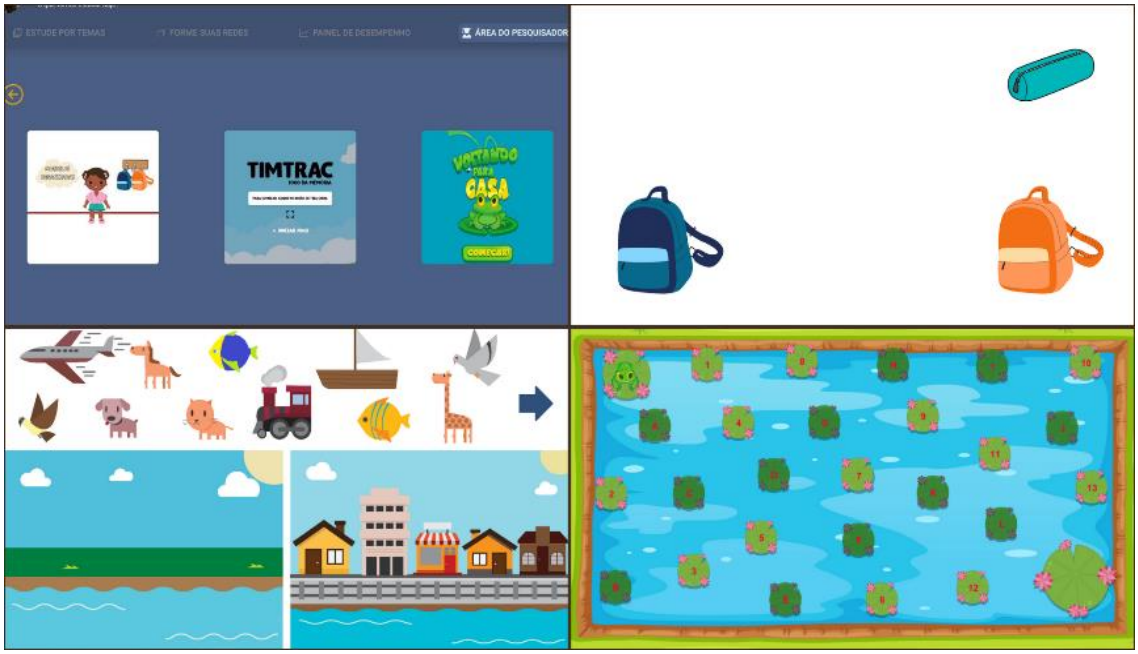
Based on the task switching paradigm (Jersild, 1927), specifically the Trail Making Test as described by Bowie and Harvey (2006), this subtest is designed to assess sequencing and cognitive flexibility skills and is administered on a tablet. It features the figure of a frog, "Zé, the Frog", and water lilies with numbers or letters displayed on the screen. In VPC, the goal is to guide the frog to its home. The child performs the task by touching and dragging on the tablet screen without lifting their finger. The assessment is based on the traces made on the tablet. Each correct connection earns one point. The task is divided into two parts.

In "Part A" of the test, the child must sequence numbers from one to 25 to help Zé, the Frog, move from a water lily in the top left corner of the screen to another in the bottom right corner, its home. In "Part B," the child alternates between numbers in ascending order and letters in alphabetical order towards the same goal. The scoring in "Part A" ranges from zero to 26 points and "Part B" also varies from zero to 26 points, with a total of up to 52 points. Scoring is halted at the first error or if the child lifts their finger from the screen.

This subtest evaluates total correct connections, correct connections until interruption, the number of times the child lifts their finger from the screen, the number of errors (out-of-order connections), response time for each stage, and total response time. Figure 1 illustrates the screens used in the BAFE-inf subtests.

Figure 1

Illustrative Screens of the Child Executive Function Assessment Battery



Note. In the top left corner is the cover of the assessment battery, where the test can be selected for execution; in the top right corner, a screen from the "Organized Backpacks" subtest; in the bottom left corner, a screen from the "Computerized Task of Working Memory in Children" subtest; and in the bottom right corner, an illustrative screen from the "Going Home" subtest.

Everyday Executive Function Scale: Parent Version (Escala da Função Executiva no Cotidiano: Versão para pais - EFEC-parents)

This is a 16-item multiple-choice instrument that uses a Likert-type scale, allowing parents to evaluate the frequency of their child's difficulties in EF such as working memory, inhibition, and cognitive flexibility based on described situations (0 = never; 1 = sometimes; 2 = often; 3 = always). Scores range from zero to 48 points, with higher scores indicating greater difficulties in executive functioning.

Developed during the COVID-19 pandemic, the EFEC-parents was designed for remote application. In an initial study, it showed a Cronbach's alpha of 0.882 (Cordeiro, 2020), indicating high internal consistency. In this paper, the scale has been refined and subjected to a validity evidence analysis process.

Questionnaire for Expert Analysis of Subtests

A form was used to evaluate each subtest of the BAFE-inf by experts. The forms included definitions related to the assessed construct, which facilitated clarity assessment. Each instruction screen and training or test item was evaluated for clarity, relevance, and design/layout. A Likert-type scale from one to four was used, where one indicated unsatisfactory and four satisfactory. Experts were able to suggest modifications, additions, or removals of items, as recommended by Almanasreh et al., (2018).

Questionnaire for Children's Assessment of Instruments

The participating children evaluated all computerized instruments. In the MO subtest, they responded to the instructions and each of the 40 items; in TIMTraC, they assessed the instruction screens, mouse training, and stages one and two of the test; and in VPC, they answered about the instructions (phase one and phase two) and items (training phase, phase one and phase two). A Likert-type scale ranging from one to four was used to assess understanding, design/layout, and ease of the items or task stages. Space was provided for free comments about the instruments. The instructions were read by the applicators to all participants to ensure standardization in the application.

Questionnaire for Assessing the EFEC by Parents and Experts

This questionnaire was completed by experts and the parents/guardians of the participating children. A Likert-type scale was used to evaluate the clarity, relevance, and design/layout of the general instructions, instructions for each part, and each item. The scale ranged from one to four, where one indicated unsatisfactory and four satisfactory. Space was provided below each item for observations, suggestions, or free-form comments by the parents/guardians.

Procedure

Experts were selected through the Lattes platform, based on their research areas in neuropsychology, EF, and/or psychometrics. Invitations were sent via email to 20 researchers: seven accepted. The collection was conducted remotely, involving the sending of an invitation letter, support material for understanding the instruments, ICF, and a questionnaire for content analysis.

Parents/guardians were invited via WhatsApp. Ten were contacted, and seven agreed to participate. The collection took place in a clinic, where they filled out the material in the presence of the researcher to clarify any doubts. With the children, following approval from the Secretary of Education and parental authorization, the collection was carried out at two schools. In a controlled non-classroom setting, administration was fully standardized. Examiners, trained to apply the instruments identically—including instruction intonation—followed a manual detailing every required sentence. Even literate children had instructions read and responses recorded by the examiner. Sessions involved 15 children and lasted about 45 minutes each.

The separation between specialists, children, and parents was maintained due to their distinct roles in the instrument development process, with only the specialists providing evidence of content validity, whereas children and parents contributed information regarding clarity and usability.

Data analysis

Descriptive analyses of the sample characteristics were conducted using the Statistical Package for the Social Sciences (IBM SPSS) version 29. Content validity was analyzed through the Content Validity Coefficient (CVC) in Excel (Microsoft Office Professional Plus 2019). The CVC assesses three criteria: clarity, practical relevance, and item relevance, taking into account expert bias (Damásio, 2023).

Proposed by Hernández-Nieto (2002), the CVC quantitatively analyzes expert judgments, adjusted for possible standard error of judge bias. The CVC for each item (CVCi) was calculated by dividing the average values assigned by experts by the

maximum value on the Likert-type scale. The adjusted CVC (CVCa), which considers the standard error (SE), was calculated by subtracting SE from CVCi. A CVCa greater than 0.80 is considered adequate (Hernández-Nieto, 2002). The CVC for the complete instrument is the average of the adjusted CVCs.

For children, item understanding, design understanding, and item ease were analyzed. For parents/guardians and participating experts, clarity, relevance, and design/layout understanding of the items were analyzed. The percentage of agreement among each participant group was also calculated, considered adequate above 90% (Stemler, 2004). Disagreement was defined as a difference of more than one point between judges' ratings. Scores of three or four were considered adequate, while scores of one or two were considered inadequate.

Ethical considerations

The study followed the guidelines established by Resolutions No. 466/2012 and No. 510/2016, which govern research involving human subjects and the humanities and social sciences, respectively. The Universal Declaration of Ethical Principles for Psychologists, the International Ethical Guidelines for Biomedical Research Involving Human Subjects, and the current statements of the Interamerican Society of Psychology on ethical actions were followed. The authors of this article declare that they have no conflicts of interest.

Results

The instruments were analyzed based on the perceptions of experts and the target population, which consisted of children and adults who are parents. The results were divided into three sections: expert analysis, child analysis, and parent analysis. The description of the sample was conducted in the topics related to each sample group. The indices obtained from children and parents refer to the comprehension, clarity, and usability of the tasks, characterizing evidence of face validity and practical acceptability. Actual content validity was established exclusively based on expert judgment.

Quantitative and Qualitative Analysis of Experts

The experts participating in this study were professionals with expertise in neuropsychology, psychometrics, technological innovation in assessment, and/or executive functions. One participant was a master's level clinical psychologist, while the others were PhDs and professors at federal universities, with experience ranging from three to 25 years in the field, most with over 10 years of experience. Their residential and work locations included Paraíba (n=3 participants), Pernambuco (n=1), Maranhão (n=1), São Paulo (n=1), and Epsom-Surrey in the United Kingdom (n=1).

The experts evaluated the instruments for use with children and the scale for application with parents. The level of agreement among the experts and the adjusted Content Validity Coefficient (CVCa) for the instruments as a whole was considered adequate, with CVCa values ranging from 0.85 to 0.96, and agreement levels varying from 95.3% to 100%. See Table 1.

Table 1

Content Validity Coefficients and Percentage of Agreement Among Experts

	MO			TIMTraC			VPC			EFEC-parents		
	Cla.	**Rel.**	**Des.**	**Cla.**	**Rel.**	**Des.**	**Cla.**	**Rel.**	**Des.**	**Cla.**	**Rel.**	**Des.**
CVC_a	0.96	0.96	0.95	0.93	0.96	0.94	0.85	0.96	0.85	0.88	0.91	0.92
%	100	100	100	100	100	100	95.24	100	100	99.74	96.49	96.49

Note. MO: Organized Backpacks; TIMTraC: Computerized Task of Working Memory in Children; VPC: Going Home; EFEC-Parents: Everyday Executive Function Scale; Cla.: Clarity; Rel.: Relevance; Des.: Understanding of design/layout elements; $CVC_{a:}$ Adjusted content validity coefficient; %. Percent agreement.

The analysis of the indicators of agreement and content validity item by item showed that most instructions and items were evaluated with a CVCa of 0.80 or above and an agreement percentage above 90%. Exceptions occurred in clarity: Part B of VPC, and items 8, 11, and 16 of EFEC-parents; relevance: items 8 and 16 of EFEC-parents; understanding of design: items 6 and 16 of EFEC-parents. In these items, agreement was < 90%.

Qualitative Assessment of the "MO" Test

The qualitative analysis of the MO test brought forward relevant suggestions for its improvement. In the first instruction screen, it was recommended to replace 'arrumando' with 'organizando' for better cohesion with the task name and regional understanding. On screen six, concerning response inhibition with the image of a cat, it was suggested to add a contextual cue with the word 'CUIDADO' (CAUTION) in uppercase. It was also proposed to include a keyboard with the signal to press nothing and the identification of the cat as 'MEL'. These suggestions were accepted and implemented.

Another suggestion was to adjust the key layout, placing them closer together to facilitate responses from younger children. However, the current layout was maintained to assess the Simon effect (Simon & Rudell, 1967), which required one hand to be distant from the other. During the training and testing stages, suggestions were made to optimize item clarity and diversity, including avoiding the repetition of the cat images. However, the item order was based on randomization and was not changed. General comments highlighted the task's quality, praising its clarity, relevance, and layout. Constant attention to support levels needed for atypical participants was recommended, with the aim of refining the test and enhancing its acceptance among professionals.

Qualitative analysis of TIMTraC

In the mouse usage training interface, it was suggested to replace the term 'move' with 'drag,' and to change 'indicated location' to 'the white circle,' as the location might not be sufficiently explicit. In the instruction for the second phase, it was suggested to more explicitly expose the child's action. Throughout the task, minimal suggestions were given, such as standardizing the size of elements. These suggestions were accepted. General comments indicated that the task has items with clarity, relevance, and design suitable for its intended evaluation.

Qualitative analysis of VPC

In the qualitative analysis of the VPC test, some specific modifications were suggested for the screen designs. On instruction screen two, it was suggested to explicitly indicate the spot where the child should touch, above the frog. In the second stage of the test, which involves alternation, it was suggested to replace the red color of the letters with orange for improved contrast.

Qualitative analysis of EFEC-parents

The experts evaluating the EFEC-parents indicated several modifications that would enhance the clarity and precision of the instructions and items. The questionnaire was revised based on suggestions to improve clarity and precision in the instructions and questions. The general instruction now explicitly specifies that responses should pertain to the evaluated child's behavior to avoid misunderstandings.

Specific items were adjusted to provide context and eliminate ambiguities. For example, information was added that the class schedule is visible in the child's room, and responses about storytelling now include details about following a logical sequence. The second part, which previously mentioned "impulsivity and inhibition," now asks about attention and behavior control to prevent confusion with introversion or shyness.

Daily situations in the questions were adapted to be more culturally relevant and specific. Examples of dangerous activities and behaviors were revised to be more common and easily understood by parents. These changes ensure that parents can respond more accurately and better reflect the actual behavior of their children.

Quantitative and qualitative analysis with children

The study involved 15 children aged six to ten years (M=8.0; SD=1.5), with three from each age group and nine females. The children were in 1st (n=4), 2nd (n=4), 3rd (n=4), and 4th grade (n=3) of primary school. Four children were non-readers (two aged six, one seven, and one eight years), and three struggled with reading (aged six, seven, and eight years). The rest could read fluently. Regarding the use of electronics for the subtests, three children had never used a tablet but faced no difficulties. Six children,

especially those aged six to eight, struggled with the mouse, having never used such devices before. All children provided positive feedback on the instruments.

In assessing the CVCa for each instruction and item across all children, without age group separation, the only CVCa below 0.80 was for the comprehension of the first instruction screen of the VPC, which was valued at 0.78. When separated by age group and characteristics of the evaluated instruments (comprehension of instruction or item; design elements comprehension; and item ease), the only feature with a CVCa below 0.80 was the ease of the TIMTraC for the groups of children aged eight, nine, and ten years. Some agreement percentages also fell below 0.90 (see Table 2).

Table 2

Content Validity Coefficients and Percentage of Agreement Among Children

		MO			TIMTraC			VPC		
Age	**Index**	**Com.**	**Des.**	**Eas.**	**Com.**	**Des.**	**Eas.**	**Com.**	**Des.**	**Eas.**
6	CVC_a	0.96	0.93	0.96	0.90	0.96	0.88	0.85	0.94	0.85
	%	100	100	100	100	100	100	100	100	100
7	CVC_a	0.96	0.93	0.96	0.90	0.96	0.80	0.80	0.95	0.80
	%	100	100	100	100	100	100	83.22**	100	100
8	CVC_a	0.96	0.93	0.96	0.89	0.95	0.75*	0.91	0.96	0.82
	%	100	100	100	95.24	100	83.34**	100	100	88.89**
9	CVC_a	0.96	0.93	0.96	0.94	0.95	0.75*	0.85	0.96	0.94
	%	100	97.73	100	100	100	83.34**	94.43	100	100
10	CVC_a	0.96	0.95	0.96	0.95	0.96	0.75*	0.92	0.96	0.91
	%	100	100	100	100	100	100	100	100	100
All	CVC_a	0.99	0.97	0.99	0.95	0.99	0.82	0.90	0.99	0.90
	%	100	99.55	100	99.05	100	93.33	95.55	100	97.78

Note. MO: Organized Backpacks; TIMTraC: Computerized Task of Working Memory in Children; VPC: Going Home; Com.: Comprehension; Des.: Understanding of design/layout elements; Eas.: Instrument ease; TIMTraC: Computerized Task of Working Memory in Children; %. Percent agreement; *. CVCa < 0.80; **. Percent agreement below 90.

On a four-point scale, where closer to four indicates easier phase, of the nine children aged eight to ten, only two rated the instrument stages as level four for ease, with others varying between one and three. For each task, the CVCa was calculated as follows: 0.98 for MO, 0.92 for TIMTraC, and 0.93 for VPC, indicating excellent content validity. All feedback from the children regarding the instruments was positive.

Quantitative and Qualitative Analysis with Parents/Caregivers

The participants of this study were seven parents/caregivers (two men and five women) aged between 31 and 51 years (M=40.57; SD=6.9). Regarding education, one had completed primary education, four had higher education, and two had postgraduate degrees. Occupations included a homemaker (n=1), psychologist (n=1), businesswoman (n=1), accountants (n=2), and public servants (n=2).

The CVCa indices for EFEC-parents showed clarity (0.93), relevance (0.92), and design comprehension (0.98), all above 0.80. The total CVCa for EFEC-parents was 0.94. As for the agreement percentage, the results were clarity (94.96%), relevance (94.12%), and design comprehension (100%). However, some individual items showed agreement below 90%: clarity (85.71% for items 1, 4, 5, 7, 11, and 12), relevance (85.71% for items 5, 7, 9, 12, and 15), and design comprehension (71.43% for item 11).

In the qualitative analysis, there was uncertainty regarding the term "supervision" in item 9, which was changed to "presence or assistance of an adult." Also, in item 8, the term "falls" was changed to "trips" to avoid ambiguity. Most participants commented at the end of the questionnaire that the questions and figures were comprehensible.

Discussion and Conclusion

The aim of this study was to analyze the content validity evidence of BAFE-inf. This was the first piece of validity evidence to be analyzed, as recommended (AERA et al., 2014; Almanasreh et al., 2018). The analysis involved expert judges and the target population, which consisted of children aged six to ten who evaluated the computerized instruments, and their parents who evaluated the EFEC-parents. The responses were subjected to both quantitative and qualitative analyses. The quantitative analyses included calculations of content validity indices and agreement indices (Alexandre & Coluci, 2011; Almanasreh et al., 2018).

Experts were provided with materials and response forms for analysis. The quantitative results from the judges' analysis showed an agreement percentage above 0.90 and a CVCa above 0.80 for clarity, pertinence, and comprehension of the design of the instruments, both considered adequate (Hernández-Nieto, 2002; Stemler, 2004). This suggests that BAFE-inf possesses appropriate instruments to assess working memory, inhibition, and cognitive flexibility (Diamond, 2013, 2020; Doğru et al., 2023; Nweze & Nwani, 2020; Souissi et al., 2022; Theodoraki et al., 2020) as per AERA et al. (2014).

All items and instructions had a CVCa above 0.80. However, there was disagreement among experts regarding clarity, pertinence, and comprehension of design in the Part B of VPC and some items of EFEC-parents, resulting in agreement indices below 0.90. Qualitative analysis revealed that the negative evaluation of part B of the VPC by one evaluator was due to an issue of contrast of the figures, which was modified for improvement. Regarding the EFEC-parents items, an expert who disagreed about the clarity, pertinence, and design of the items provided improvement suggestions, which were analyzed and accepted, resulting in modifications to enhance the instrument.

The target population also participated in the analyses, similar to in Leal (2022). The children evaluated all the computerized instruments, and the scoring was mostly adequate (Hernández-Nieto, 2002; Stemler, 2004). Children were also asked to evaluate the ease of each item, and in this regard, the CVCa indices and agreement percentages for TIMTraC and VPC among eight-, nine-, and ten-year-olds showed values below expectations. These results suggest the need to consider children's technological familiarity and attentional demands, especially in clinical contexts. In addition, the qualitative adjustments made during the instrument's development, including visual and linguistic changes, directly contributed to improving the comprehension and usability indices, thereby strengthening its practical applicability.

These discrepancies can be attributed to the variability of latent traits in working memory (Baddeley, 2012) and cognitive flexibility (Diamond, 2013) among the children. Those with better abilities found the tasks easier, while those with more difficulty found them harder. Additionally, social desirability may have influenced the results, as some children may have been reluctant to rate the items as difficult, since all provided positive feedback at the end.

When considering the CVCa values for each instruction and item, the only value below 0.80 was the comprehension of the instruction screen one of VPC, which had a value of 0.78. Qualitative analyses indicate that screen one presents only a part of the instruction, which is complemented on screen two. Upon reaching screen two, the children understood the instruction better and gave it a higher rating. The division of the instruction across screens was not modified, because, considering the costs and benefits, it was found that children could better understand the instruction presented in parts rather than on a single screen, where denser text and smaller font might divert attention.

When not separated by age groups, the CVCa indices for children ranged between 0.92 and 0.98, and the concordance indices were close to 1.0, indicating measures with

appropriate evidence of content validity for the intended constructs (AERA et al., 2014; Alexandre & Coluci, 2011; Almanasreh et al., 2018).

Adults with children between six and ten years old also evaluated the EFEC-parents. Suggestions for improvement, made by a mother on items considered understandable by experts, demonstrated the importance of evaluation by the target population as well (Leal, 2022). The CVCa indices among the parents, both in the analysis of the scale as a whole and among the instructions and items, were above 0.80, suggesting adequacy to the content-based validity evidence of EFEC-parents (Almanasreh et al., 2018; Hernández-Nieto, 2002).

The indices of clarity, pertinence, and design comprehension for the entire scale were above 0.90. However, some items showed inadequate concordance percentages (Stemler, 2004), mostly 85.71%. This percentage varies according to the size of the sample group; the disagreement was just one person in a group of seven. Nevertheless, the suggestions for improvement were implemented.

The results of this study reduce the scarcity (Berardi et al., 2021; Canet-Juric et al., 2021; Kusi-Mensah et al., 2022b; Santana et al., 2019) of computerized instruments that assess basic EF components in children, offering simple and quick application and good content validity indices, which supports subsequent validity analyses (AERA et al., 2014; Almanasreh et al., 2018) for future clinical and institutional use. The findings may also benefit other Latin American countries facing similar challenges in accessing adequate child-focused psychometric tools. The BAFE-inf structure can guide regional adaptations, enhancing ecological validity and clinical applicability across the region.

This study has limitations, including the small number of experts per subtest and the absence of chance-corrected indices, such as kappa, which may overestimate agreement. Moreover, evaluation with the target population does not represent content validity. Future studies should examine construct validity, reliability, regional norms, and cross-cultural adaptations.

The analysis of BAFE-inf's content validity confirmed the quality and utility of the tools developed, demonstrating that they accurately measure the intended constructs. The findings emphasized the importance of expert and target population input, with quantitative data showing consistent evaluations. BAFE-inf, designed to assess EF in children aged six to 10, helps fill the gap in instruments with adequate validity. It provides a comprehensive measure through computerized tests and a parent-perception scale, capturing both controlled and everyday performance.

Referências

Alexandre, N. M. C., & Coluci, M. Z. O. (2011). Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*, *16*(7), 3061-3068. https://doi.org/10.1590/S1413-81232011000800006

Almanasreh, E., Moles, R., & Chen, T. F. (2018). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*. https://doi.org/10.1016/j.sapharm.2018.03.066

American Educational Research Association - AERA, American Psychological Association - APA, and National Council on Measurement in Education - NCME. (2014). Standards for educational and psychological testing. Washington, DC. https://www.testingstandards.net/open-access-files.html

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual review of psychology*, *63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Berardi, A., Panuccio, F., Pilli, L., Tofani, M., Valente, D., & Galeoto, G. (2021). Evaluation instruments for executive functions in children and adolescents: a systematic review. *Expert review of pharmacoeconomics & outcomes research*, *21*(5), 885-896. https://doi.org/10.1080/14737167.2021.1908889

Bowie, C., Harvey, P. (2006). Administration and interpretation of the Trail Making Test. *Nature Protocols, 1*, 2277-2281. https://doi.org/10.1038/nprot.2006.390

Brocki, K. C., & Bohlin, G. (2010). Executive Functions in Children Aged 6 to 13: A Dimensional and Developmental Study. *Developmental Neuropsychology, 26*(2), 571-593. http://dx.doi.org/10.1207/s15326942dn2602_3

Canet-Juric, L., del-Valle, M. V., Gelpi-Trudo, R., García-Coni, A., Zamora, E. V., Introzzi, I., & Andrés, M. L. (2021). Desarrollo y validación del Cuestionario de Funciones Ejecutivas en niños de 9 a 12 años (CUFE). *Avances En Psicología Latinoamericana*, *39*(1), 1-25. https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.9892

Cevada, T., Conde, E., Marques, D., & Deslandes, A. C. (2019): Test-retest reliability of the simon task: a short version proposal. *Somatosensory & Motor Research*, 1-8. https://doi.org/10.1080/08990220.2019.1689114

Chami, S., Charalambous, C., Knijnik, S. R., & Docking, K. (2022). Language and executive function skills as predictors of semantic fluency performance in pre-school children. *International journal of speech-language pathology*, *24*(6), 626-635. https://doi.org/10.1080/17549507.2021.2008005

Cordeiro, A. S. (2020). Análise das funções executivas infantis: antes e durante a pandemia por COVID-19. [Dissertação de mestrado. Universidade Federal da Paraíba]. https://repositorio.ufpb.br/jspui/handle/123456789/20876?locale=pt_BR

Cordeiro, A. S., Tomaz, D. F. de O., & Minervino, C. A. da S. M. (2019). Memória de trabalho infantil: informatização de uma tarefa avaliativa. *Revista Neuropsicologia Latinoamericana*, *11*(2), 21-33. https://www.neuropsicolatina.org/index.php/Neuropsicologia_Latinoamericana/article/view/493

Dai, D. W. T., Franke, N., Wouldes, T. A., Brown, G. T. L., Tottman, A. C., Harding, J. E., & PIANO Study Group (2021). The contributions of intelligence and executive function to behaviour problems in school-age children born very preterm. *Acta paediatrica (Oslo, Norway: 1992)*, *110*(6), 1827-1834. https://doi.org/10.1111/apa.15763

Damásio, B. (2023, Junho 09). Compreendendo os índices de validade de conteúdo. Psicometria Online. https://psicometriaonline.com.br/compreendendo-os-indices-de-validade-deconteudo/#:~:text=%E2%80%93%20Coeficiente%20de%20Validade%20de%20Conte%C3%BAdo,relev%C3%A2ncia%20te%C3%B3rica%20de%20cada%20item

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

Diamond A. (2020). Executive functions. *Handbook of clinical neurology*, *173*, 225-240. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

Doğru, Y., Carroll, D., & Blakey, E.P. (2023). Cognitive Flexibility in Early Childhood: A Contemporary View of the Development of Flexible Goal-Oriented Behavior. *Psikoloji Çalışmaları / Studies in Psychology,* *43*(2), 171-193. https://doi.org/10.26650/sp2022-1138580

Elage, G. K. C. de F., & Seabra, A. G. (2021). Desenvolvimento e Propriedades Psicométricas do Teste Informatizado de Avaliação das Funções Executivas. *Avaliação Psicológica*, *20*(1), 100-110. https://dx.doi.org/10.15689/ap.2021.2001.17491.11

Guerra, A., Hazin, I., Guerra, Y., Roulin, J. L., Le Gall, D., & Roy, A. (2021). Developmental Profile of Executive Functioning in School-Age Children From Northeast Brazil. *Frontiers in psychology*, *11*, 596075. https://doi.org/10.3389/fpsyg.2020.596075

Guerra, A., Hazin, I., Siebra, C., Rezende, M., Silvestre, I., Le Gall, D., & Roy, A. (2022). Assessing executive functions in Brazilian children: A critical review of available

tools. *Applied Neuropsychology: Child, 11*(2), 184-196. https://doi.org/10.1080/21622965.2020.1775598

Hernández-Nieto, R. A. (2002). Contributions to statistical analysis: The coefficients of proportional variance, content validity and kappa capa. Booksurge Publishing.

Isherwood, S. J. S., Bazin, P. L., Miletić, S., Stevenson, N. R., Trutti, A. C., Tse, D. H. Y., Heathcote, A., Matzke, D., Innes, R. J., Habli, S., Sokołowski, D. R., Alkemade, A., Håberg, A. K., & Forstmann, B. U. (2023). Investigating Intra-Individual Networks of Response Inhibition and Interference Resolution using 7T MRI. *NeuroImage, 271*, 119988. https://doi.org/10.1016/j.neuroimage.2023.119988

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, 14,* 5-81. https://archive.org/details/mentalsetshift00jers/page/6/mode/2up.

Jiang, H., Johnstone, S. J., & Lu, J. (2022). Comparing the Efficacy of Two School-Based Approaches of Neurocognitive Training for Enhancing Executive Functions. *Children (Basel, Switzerland), 9*(10), 1501. https://doi.org/10.3390/children9101501

Kusi-Mensah, K., Nuamah, N. D., Wemakor, S., Agorinya, J., Seidu, R., Martyn-Dickens, C., & Bateman, A. (2022a). Assessment Tools for Executive Function and Adaptive Function Following Brain Pathology Among Children in Developing Country Contexts: a Scoping Review of Current Tools. *Neuropsychology review, 32*(3), 459-482. https://doi.org/10.1007/s11065-021-09529-w

Kusi-Mensah, K., Nuamah, N. D., Wemakor, S., Agorinya, J., Seidu, R., Martyn-Dickens, C., & Bateman, A. (2022b). A Systematic Review of the Validity and Reliability of Assessment Tools for Executive Function and Adaptive Function Following Brain Pathology among Children and Adolescents in Low- and Middle-Income Countries. *Neuropsychology review, 32*(4), 974-1016. https://doi.org/10.1007/s11065-022-09538-3

Leal, P. B. (2022). Instrumento para avaliação da qualidade do serviço médico de emergência sob perspectiva do usuário proposta e validação de conteúdo [Trabalho de Conclusão de Curso, Escola Bahiana de Medicina e Saúde Pública]. Repositório Institucional. http://www.repositorio.bahiana.edu.br/jspui/handle/bahiana/7181

Likhitweerawong, N., Khorana, J., Boonchooduang, N., Phinyo, P., Patumanond, J., & Louthrenoo, O. (2022). Association between executive function and excess weight in pre-school children. *PloS one, 17*(10), e0275711. https://doi.org/10.1371/journal.pone.0275711

Martins, C. R. (2020). Evidências de validade do Teste de Inibição e Flexibilidade Cognitiva - TIF: Analisando o desenvolvimento do controle inibitório e flexibilidade cognitiva

na infância. [Tese de Doutorado. Universidade Federal da Bahia]. https://repositorio.ufba.br/ri/handle/ri/32839

Minervino, C. A. da S. M., & Tomaz, D. F. de O. (2021). Habilidades precursoras da leitura: análise da tarefa informatizada de memória de trabalho para crianças. Em J. F. de S., & A. L. P. G. P. Navas. Avaliação da linguagem oral, escrita e de habilidades relacionadas: Panorama nacional de instrumentos. Vetor Editora.

Nogues, C. P., Nunes, D. M. (2023). A influência das funções executivas na matemática. Em L. V. Corso, E. F. de Assis, & C. P. Nogues (Orgs.), Matemática na Educação Infantil (pp. 93-110). CirKula. https://lume.ufrgs.br/handle/10183/265677

Nweze, T., & Nwani, W. (2020). Contributions of Working Memory and Inhibition to Cognitive Flexibility in Nigerian Adolescents. *Developmental neuropsychology*, *45*(3), 118-128. https://doi.org/10.1080/87565641.2020.1765169

Pereira, E. E. L. D., Minervino, C. A. da S. M., Cruz, L. F. P. da., Roama-Alves, R. J., & Andrade, J. M. (2020). Executive Functions in Children with ADHD and/or Reading Difficulty. *Psicologia: Teoria E Pesquisa*, *36*, e3623. https://doi.org/10.1590/0102.3772e3623

Perone, S., Simmering, V. R., & Buss, A. T. (2021). A Dynamical Reconceptualization of Executive-Function Development. *Perspectives on Psychological Science*, *16*(6), 1198-1208. https://doi.org/10.1177/1745691620966792

Reppold, C. T., Wechsler, S. M., Almeida, L. da S., Elosua, P., & Hutz, C. S. (2020). Perfil dos Psicólogos Brasileiros que Utilizam Testes Psicológicos: Áreas e Instrumentos Utilizados. *Psicologia: Ciência E Profissão*, *40*, e201348. https://doi.org/10.1590/1982-3703003201348

Ruffini, C., Marzocchi, G. M., & Pecini, C. (2021). Preschool Executive Functioning and Child Behavior: Association with Learning Prerequisites?. *Children (Basel, Switzerland)*, *8*(11), 964. https://doi.org/10.3390/children8110964

Santana, A. N. de., Melo, M. R. A., & Minervino, C. A. da S. M. (2019). Instrumentos de Avaliação das Funções Executivas: Revisão Sistemática dos Últimos Cinco Anos. *Avaliação Psicológica*, *18*(1), 96-107. http://dx.doi.org/10.15689/ap.2019.1801.14668.11

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *The Journal of applied psychology*, *51*(3), 300-304. https://doi.org/10.1037/h0020586

Souissi, S., Chamari, K., & Bellaj, T. (2022). Assessment of executive functions in school-aged children: A narrative review. *Frontiers in psychology*, *13*, 991699. https://doi.org/10.3389/fpsyg.2022.991699

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, *9*(4), 1-19. https://doi.org/10.7275/96jp-xz07

Theodoraki, T. E., McGeown, S. P., Rhodes, S. M., & MacPherson, S. E. (2020). Developmental changes in executive functions during adolescence: A study of inhibition, shifting, and working memory. *The British journal of developmental psychology*, *38*(1), 74-89. https://doi.org/10.1111/bjdp.12307

Venturieri, C., Silva, L. D. P. da, Lunkes, S., Stutz, L. P., & Dias, N. M. (2023). Avaliação de funções executivas em pré-escolares: revisão de escopo da literatura brasileira: Funções executivas em pré-escolares. *Neuropsicología Latinoamericana*, *15*(1), 31-44. https://neuropsicolatina.org/index.php/Neuropsicologia_Latinoamericana/article/view/7562

Wang, L., Li, J., Jia, F., Lian, L., & Li, L. (2024). The Development of Response and Interference Inhibition in Children: Evidence from Serious Game Training. *Children (Basel, Switzerland)*, *11*(2), 138. https://doi.org/10.3390/children11020138

Received: 2025-01-12
Accepted: 2025-11-27