PTP
Psicologia
Teoria e Prática

The Stroop Paradigm in Brazilian Studies: A Scoping Review

Maria Eduarda de Oliveira Martins[1], Caio Marcos Garcia Tosi[1], Bernardo Peressoni Luz[1],
Luiz Henrique Toresan[1], Chrissie Ferreira de Carvalho[1] and Natália Martins Dias[1]

[1] Federal University of Santa Catarina (Universidade Federal de Santa Catarina [UFSC])

Author Note

Maria Eduarda de Oliveira Martins 🆔 https://orcid.org/0000-0002-8830-9496
Caio Marcos Garcia Tosi 🆔 https://orcid.org/0000-0003-1589-0459
Bernardo Peressoni Luz 🆔 https://orcid.org/0000-0002-7170-4350
Luiz Henrique Toresan 🆔 https://orcid.org/0000-0002-3381-4735
Chrissie Ferreira de Carvalho 🆔 https://orcid.org/0000-0002-1369-6188
Natália Martins Dias 🆔 https://orcid.org/0000-0003-1144-5657
Correspondence concerning this article should be addressed to Maria Eduarda de Oliveira Martins, A/C Cidade Universitária. Universidade Federal de Santa Catarina, campus Trindade. CFH, Bloco C – sala 15ª. 88040-900. Florianópolis/SC. Email: psi.eduardamartins@gmail.com

Abstract

The Stroop paradigm is widely used in the executive functions assessment. This study investigated the use of tasks based on the Stroop paradigm in national studies published between 2000 and 2022 by doing a scoping review. Empirical studies were selected in the following databases: Pubmed, Scielo Brazil, LILACS, Pepsic, Index Psi Periódicos, Index-Psi teses, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) and Portal CAPES de Teses e Dissertações. From the initial 1448 studies found, 147 were selected (90 theses/dissertations and 57 articles). Most articles were published in psychology and psychiatry journals, and the construct most targeted by the tasks was inhibitory control. Four versions were identified, and the classic Stroop Color and Word test was the most used (124 studies). Different scores were reported. There were few psychometric studies and few studies with task norms. Computerized versions are still infrequently used. The review helps to outline the versions used in the national context.

Keywords: Stroop; inhibitory control; neuropsychological assessment; executive functions; neuropsychology

O PARADIGMA DE STROOP NOS ESTUDOS BRASILEIROS:
UMA REVISÃO DE ESCOPO

Resumo

O paradigma do Stroop é amplamente utilizado na avaliação das funções executivas. Este estudo investigou o uso de tarefas baseadas no paradigma de Stroop em estudos nacionais publicados entre 2000 e 2022, por meio de uma revisão de escopo. Buscaram-se estudos empíricos nas bases de dados: Pubmed, Scielo Brazil, LILACS, Pepsic, Index Psi Periódicos, Index-Psi teses, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e Portal CAPES de Teses e Dissertações. Do montante inicial de 1448 estudos, 147 foram selecionados (90 teses/dissertações e 57 artigos). A maioria dos artigos foi publicada em revistas de psicologia e psiquiatria, e o controle inibitório foi o constructo mais mencionado como alvo das tarefas. Quatro versões foram identificadas, sendo a mais utilizada o Stroop de cores e palavras clássico (124 estudos). Diferentes escores/índices foram relatados. Houve poucos estudos psicométricos e com normas das tarefas. Versões computadorizas ainda são pouco utilizadas. A revisão colabora para mapeamento das versões utilizadas no âmbito nacional.

Palavras-chave: Stroop; controle inibitório; avaliação neuropsicológica; funções executivas; neuropsicologia

EL PARADIGMA DE STROOP EN LOS ESTUDIOS BRASILEÑOS:
UNA REVISIÓN DEL ALCANCE

Resumen

El paradigma de Stroop se usa ampliamente en la evaluación de las funciones ejecutivas. Este estudio investigó el uso de tareas basadas en esto paradigma en estudios nacionales publicados entre 2000 y 2022, haciendo una revisión de alcance. Los estudios empíricos fueron seleccionados en las bases de datos: Pubmed, Scielo Brasil, LILACS, Pepsic, Index Psi Periódicos, Index-Psi teses, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) y el Portal CAPES de Teses e Dissertações. De los 1448 estudios iniciales encontrados, se seleccionaron 147 (90 tesis / disertaciones y 57 artículos). La mayoría de los artículos se publicaron en revistas de psicología y psiquiatría, y el constructo más citado de las tareas fue el control inhibitorio. Se identificaron cuatro versiones, siendo la prueba clásica Stroop Colores y Palabras la más utilizada (124 estudios). Se informaron diferentes puntuaciones. Hubo pocos estudios psicométricos y pocos estudios con normas de tareas. Las versiones computarizadas todavía se utilizan poco. La revisión ayuda a identificar las versiones utilizadas a nivel nacional.

Palabras claves: Stroop; control inhibitorio; evaluación neuropsicológica; funciones ejecutivas; neuropsicología

Neuropsychology is an interdisciplinary area of science; the primary practices are the Neuropsychological Assessment (ANP – Brazilian abbreviation for *Avaliação neuropsicológica*) and intervention. The pillars of the ANP are the interview, observation, scales, and neuropsychological tests (or tasks). The application of instruments, namely neuropsychological testing, is part of the process and allows the detailed investigation and quantification of measures of specific components of human cognition (e.g., attention, memory, executive functions) (Bertrand et al., 2019; Dias & Seabra, 2019; Goldstein & Mcneil, 2013; Malloy–Diniz et al., 2016).

For assessment quality, it is essential to construct and use adequate neuropsychological tests (Bilder & Reise, 2019; Dias & Seabra, 2019). The ANP is being consolidated with the help of psychometrics, which contributes to the quality of measurement instruments. The current neuropsychological practice results from significant advances in recent decades, using different assessment approaches, contributions from psychometrics, and, more recently, a growing trend toward computerized tests (Bilder & Reise, 2019; Casaletto & Heaton, 2017). However, there is still some resistance to integrating technology into the area (Bilder & Reise, 2019; Miller & Barr, 2017).

Among the most used neuropsychological tests, those based on classical paradigms stand out, with one example being the Stroop Paradigm, which assesses cognitive control by keeping a goal in mind and suppressing a dominant response to perform a less habitual response (Strauss et al., 2006). When reviewing the national production related to the ANP, Ramos and Hamdan (2016) identified that the Stroop Color–Word Test version was the 10th most used instrument, mentioned by 34 of the 241 studies that made up the review. There are many tasks based on this paradigm, first described in an article published by John Stroop in 1935, in which he presented the classic version of the task (Lezak et al., 2012; Scarpina & Tagini, 2017; Strauss et al., 2006).

The original Color–Word version consists of three parts. First, the participant must read the words 'green,' 'red,' 'blue,' and 'yellow' printed in black on a card. Secondly, the participant says the name of these colors printed in colored circles or rectangles. Finally, in the third, the names of the colors printed incongruously with the written words appear, and the participant has to name the color of the word instead of reading the word. For example, the word "red" colored in blue must be called 'blue.' The "Stroop Effect" occurs in the third part, where there is cognitive interference since reading processing is automatic. It is necessary to inhibit this tendency to name the printed color. The Stroop test is one of the instruments most used for assessing executive functions (EF) (Campanholo, 2018). More specifically, it considers inhibitory control and selective attention (Silva et al., 2017); however, the literature also suggests its application to assess processing speed, cognitive flexibility, and working memory (Strauss et al., 2006).

The Stroop is widely used in research and ANP practice (Ramos & Hamdan, 2016; Santana et al., 2019). Reviews and surveys have been conducted since the 1930s and highlight the increasing use of neuropsychological testing instruments in clinical, hospital, school, and psychiatric contexts (Rabin et al., 2005). The review by Rabin et al. (2005) represented the most significant review of the neuropsychological practice performed so far, examining aspects of the instruments that had not been considered, such as their frequency of use and correction. The

study collected data from 747 North American professionals who performed ANP. The results indicated that the most frequently assessed construct was attention, with Stroop ranking 4th among the instruments most used to assess this construct.

A review of Italian studies also investigated the types of measures (indexes) used in assessing the Stroop Effect (Scarpina & Tagini, 2017). The authors verified various methods for evaluating its effect by analyzing studies that proposed normative data for the tasks based on the Paradigm. However, the review reported that none of the versions presented in the literature represented an effective way of measuring the Stroop Effect, which highlighted the need to strengthen normative references that jointly consider the speed and accuracy of the response in the congruent and incongruent stages of the test (Scarpina & Tagini, 2017).

More recently, the systematic review performed by Santana et al. (2019) investigated which instruments were used to assess EF, in national and international contexts, between 2010 and 2016. The study also outlined the characteristics of the methodological designs of the empirical studies that used these instruments for EF assessment, the frequency of publications in the area, the objectives of the selected studies, the psychometric parameters of the instruments identified, and the populations most frequently investigated in these publications. A total of 35 articles were analyzed. The main instruments used were the Wisconsin Card Sorting Test, the Trail Making Test, the Wechsler Intelligence Scales, and the Stroop Test. The summary of the findings suggested that, although numerous versions of tests and tasks of classical paradigms are commonly used in the national context, there is still a need for the provision of performance standards and a demand for the construction and validation of new instruments in Brazil.

There are several versions of Stroop tasks. One of the most used is the Victoria version, which also uses color and words and has the advantage of being shorter than the original version of the Stroop, with only 24 stimuli in each card/part of the task (Silva et al., 2017; Strauss et al., 2006). There are also versions explicitly developed for preschool children, such as versions with shapes or figures (Carlson, 2005), and others that use numbers, such as the Five Digit Test (Sedó, 2004), which also do not require reading. The literature mentions computerized versions as well (Campanholo, 2018).

Given the wide use in neuropsychological and clinical research and the number of versions available, this review aimed to map the use of tasks based on the Stroop paradigm in Brazilian empirical studies over the last 22 years (from 2000 to the present). Specifically, this review investigated: 1) the contexts in which the instruments based on the Stroop paradigm (age groups and clinical conditions) are used; 2) the skills that are assessed (as reported by the studies selected for the review); 3) the versions used, and 4) the scores and indexes used.

A scoping review is suitable for this purpose. Its objectives include mapping and examining how research is conducted in a determined area and identifying the primary evidence and characteristics associated with a topic, area, or concept (Munn et al., 2018). This study aimed to provide information about the Stroop task versions used and to support the growth of the ANP in the national context.

Method

This scoping review was prepared based on the recommendations of the PRISMA extension for Scoping Reviews—Prisma-ScR (Tricco et al., 2018).

Eligibility Criteria

For the initial survey of studies in this review process, the following inclusion criteria were used: a) being an empirical study; b) being a Brazilian study; c) having used an instrument based on the Stroop Paradigm for evaluation purposes; d) using a typical or atypical sample of participants from any age group of the life cycle (human studies); e) having been published between 2000 and 2022 (see below the topic about updating the search), and f) published in the Portuguese, English or Spanish languages. As exclusion criteria, the following were considered: a) review studies or theoretical studies; b) studies that did not use any assessment instrument based on the Stroop Paradigm; and c) studies conducted with foreign populations.

Databases and search strategies

The initial searches were conducted in June 2020 jointly by two authors (CMGT and MEOM). The PubMed, SciELO Brazil, LILACS, PePSIC, and INDEX PSI PERIODIC databases were consulted to locate studies published in article format. The INDEX PSI TESES databases, the Brazilian Digital Library of Theses and Dissertations (BDBTD), and the CAPES Theses & Dissertations Portal were consulted to identify studies published as theses and dissertations (T&D). As a search strategy, the following string was used in all databases, filtering for studies published between 2000 and 2020 and searching in all indexes:

Assessment AND (Stroop OR "executive function" OR "inhibitory control" OR inhibition OR "selective attention" OR "interference control" OR "cognitive conflict" OR self-control).

In the PubMed database, another part was inserted in this string to select only studies with Brazilian populations: ...*AND (Brazil OR Brazilian)*. In addition to searching the databases, other studies were included based on checking lists of references and from the authors' previous knowledge (CFC and NMD).

Selection of studies

The first selection stage (screening) took place through reading titles and abstracts. It was conducted independently by two authors (CMGT and MEOM), with 100% agreement regarding the T&D. Concerning the articles, there was only one disagreement, which was resolved through discussion between the authors. After the screening, the selection stage began with the studies being read in full, again performed by two authors (CMGT and MEOM). There were discrepancies in 4.65% of the articles and 7.07% of the T&D, which were later resolved through the analysis of a third judge (role played by NMD and CFC). All reasons for exclusion were recorded at this stage.

Data extraction

From the reading of the studies in full, the authors extracted the following information from the included studies: a) General information (i.e., title, authors' names, year of publication, journal in which it was published, or university in which it was defended); b) main objective; c) the total number of people who composed the sample analyzed (typical or atypical); d) the age group/mean age of the participants; and e) other sample data (e.g., whether or not the study used a clinical sample and what type of clinical condition made up the sample studied).

Regarding the tasks based on the Stroop Paradigm, the authors extracted the following information from the studies: f) the name of the test used; g) the version the test used was based on (if the information was available); h) whether or not the version used in the study was computerized; i) the references of the Brazilian version and the original version (on which the test was based); j) description of the instrument, in terms of (j1) the type of stimulus and (j2) the scores used; and k) the skill assessed by the instrument, according to the authors of the studies.

The studies were also classified as to whether or not they were 'psychometric,' that is, whether their objective was to provide data on reliability, validity, or normative task data. Data extraction was performed by four authors (CMGT, MEOM, BPL, and LHT), with supervision and verification by other team members (CMGT, MEOM, and NMD), and was carried out using an Excel spreadsheet.

The search update and study selection stage

The search was updated at two moments, using the same procedure and search strategy in each database: 1) in August 2021 to identify new studies published during the year 2020; and 2) in May 2022, covering the entire period from 2021 to the date of the search update (May 11th, 2022). The selection from this new material followed the same procedures described above.

Results

Selection of studies

After performing searches in the databases as mentioned earlier using the strings presented, 868 articles were identified (PubMed = 184; SciELO Brasil = 182; LILACS = 459; PePSIC = 25; Index Psi Periodicals = 18) and 576 T&D (Index Psi Theses = 5; BDBTD = 530; CAPES Theses & Dissertations Portal = 41); in addition to these, one thesis and three articles were included based on the authors' indication (prior knowledge/checking of reference lists). After removing duplicate studies, the sample included 745 articles and 544 T&D, totaling 1289 studies that went on to the selection stage with reading the titles and abstracts. After this screening step, 1088 studies (649 articles and 439 T&D) were excluded from the analysis as they did not fulfill the inclusion criteria of this review.

Two hundred-one studies (96 articles and 105 T&D) were then assessed for eligibility. After reading the studies, 54 were excluded (39 papers and 15 T&D) due to fulfilling any exclusion criteria, as identified in Figure 1. In studies without access to the full text, attempts were made

to contact the authors to request the respective materials. There was no response in any of the cases. At the end of the eligibility stage, 147 studies were included in the review (90 T&D and 57 articles). Figure 1 presents the process flowchart.

Figure 1

Flowchart of study selection steps (PRISMA-ScR)
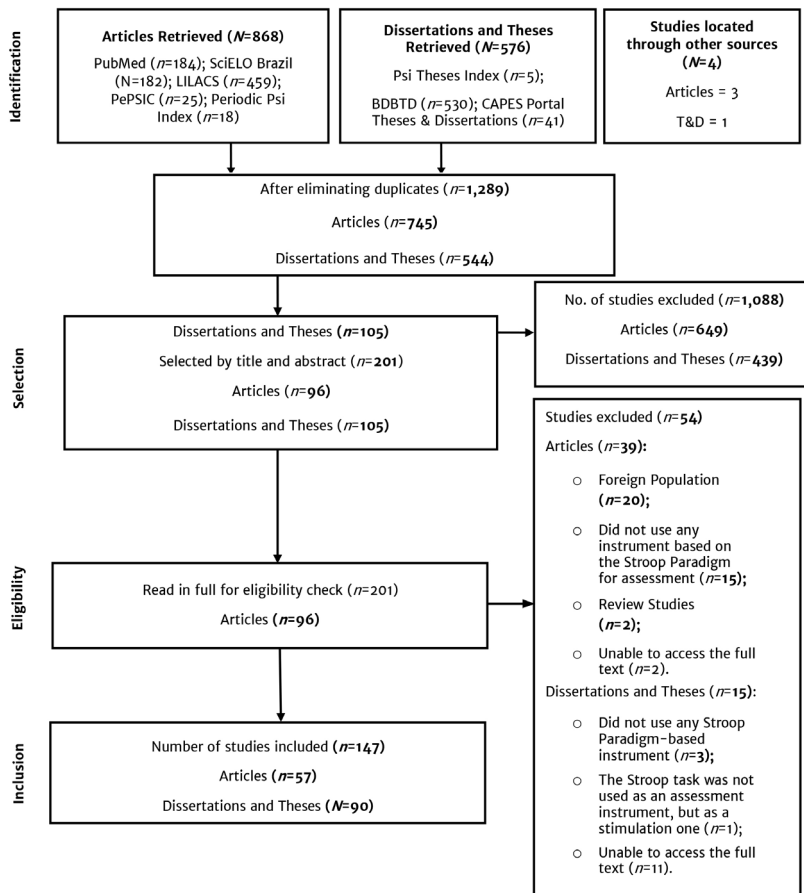
Characteristics of the studies

Due to the inclusion criteria adopted, all studies selected for this manuscript were published between 2000 and 2021. A greater concentration of studies was published between 2011 and 2016 (*n* = 83), with 17 studies published in 2015 and 18 published in 2016. In subsequent years, there was a decrease in the number of productions (29 between 2017 and 2019), with 7 in 2020, 2 in 2021 and no study published in 2022. Among the articles analyzed, 28 articles were published in psychology journals, 15 in psychiatry journals, and 14 in journals from other areas.

The T&D was conducted in graduate programs in Medical Sciences and subfields (44), Psychology and subfields (26), Biological Sciences and subfields (5), Developmental Disorders (5), Health Sciences (3), Physical Education (3), Speech Therapy (1), Pharmacology (1), Education (1) and Community Development (1). Regarding location, the T&D was conducted at universities located in the states of São Paulo (57), Rio Grande do Sul (7), Minas Gerais (6), Rio de Janeiro (4), Rio Grande do Norte (3), Alagoas (3), Ceará (2), Pernambuco (1), Paraná (1), Paraíba (1), Mato Grosso do Sul (1), Goiás (1), Bahia (1), and Distrito Federal (1).

Of the total number of studies, 15 (10.2%) were characterized as psychometric, as they described among their objectives the adaptation, construction, and survey of validity evidence, reliability, or provision of normative data from instruments to assess aspects of executive functions. Three (2.04%) studies provided normative data.

Summary of studies

Regarding the contexts of the Stroop tasks used, the 147 studies were conducted with varied age groups, with emphasis on the 40-59 age group, which was present in 42.18% of the studies, followed by almost 40% with the age range of 18-39 years and 38.78% with the 60-69 age group. One study included a three-year or less sample (Table 1). In 14 of the studies, there was no information on the age range. In these cases, the mean age reported was considered (e.g., in one study, the mean age was 36.55 ± 11.34. Therefore, this study was computed to cover the age group of 18 to 39 years).

Table 1

The Age Range of the Samples of the Studies Analyzed

Age Range	No.	%
≤3	1	0.68
4-5	12	8.16
6-10	40	27.21
11-14	28	19.05
15-17	17	11.56
18-39	58	39.46
40-59	62	42.18
60-79	57	38.78
≥80	17	11.56

Note: The sum and % exceed the total number of the reviewed studies, as some included more than one of the delimited age groups.

Individuals with typical development composed the sample in 40% of the studies (Table 2), which refers to studies that did not investigate clinical samples. Most of the studies surveyed individuals with some clinical condition, with emphasis on people with some general physical

health conditions (22 studies), dementia (17 studies), ADHD (10 studies), and abusive users of alcohol and other drugs (8 studies).

Table 2

Characterization of the Samples of the Studies Reviewed according to the Clinical Conditions

Samples	No.	%
No clinical condition (typical sample)	59	40.14
General physical health conditions*	22	14.97
Dementia*	17	11.56
ADHD	10	6.80
Abusive users of alcohol and other drugs	8	5.44
Depression	6	4.08
Neurological injuries and accidents*	5	3.40
Bipolar disorder	5	3.40
Schizophrenia	4	2.72
Anxiety Disorders*	3	2.04
Epilepsy	3	2.04
Learning difficulties*	2	1.36
Migraine	2	1.36
Eating disorders*	2	1.36
Self-mutilation	1	0.68
Autism	1	0.68
Dyscalculia and dyslexia	1	0.68
Multiple sclerosis	1	0.68
Narcolepsy	1	0.68
Cerebral palsy	1	0.68
Restless legs syndrome	1	0.68
Williams syndrome	1	0.68
Premenstrual dysphoric syndrome	1	0.68
Clinically isolated syndrome	1	0.68
Disruptive behavior disorder	1	0.68
Conduct disorder	1	0.68

*General health conditions: diabetics, HIV, rheumatoid arthritis, bariatric patients, writer's cramp, cirrhosis, myotonic dystrophy type 1, peripheral arterial disease, Wilson's disease, basilar artery occlusive disease, chronic kidney disease, chronic pain, chronic hepatitis c, hyperphenylalaninemia, hypothyroidism, chronic occupational mercurialism, obesity, perimenopausal women, toxoplasmosis and breast cancer.
*Dementia: Alzheimer's disease, Parkinson's disease, frontotemporal lumbar degeneration, dementia, infectious parasitic dementia, and Machado-Joseph disease (ataxia).
*Learning difficulties: attention, language, reading, writing, psychomotor aspects difficulties.
*Neurological injuries and accidents: delayed diffuse axonal injury, hemorrhagic cerebrovascular disease, central neuropathic pain, and HTLV-1-associated myelopathy
*Anxiety Disorders: agoraphobia, anxiety disorder, and panic disorder.
*Eating disorders: anorexia and binge eating disorder.
The sum and % exceed the total number of the reviewed studies, as some included more than one of the delimited samples.

Table 3 details the number and percentage of studies that reported the cognitive constructs measured using paradigm-based instruments. Concerning the cognitive skills assessed by the tasks, inhibitory control was the most mentioned (70.74%), followed by attention (65.98%). One study did not report which skill was assessed by the task based on the Stroop paradigm.

Table 3

Skills/Constructs Assessed Using the Stroop Paradigm-Based Instruments in the Studies Selected

Skills	No.	%
Inhibitory control	104	70.74
Attention and subcomponents	97	65.98
Cognitive flexibility	35	23.80
Executive functions	26	17.68
Processing speed	16	10.88
Working memory	5	3.4
Emotional measures	3	2.04
Impulsiveness	2	1.36
Not included	1	0.68

Note: The sum and % exceed the total number of the reviewed studies, as some listed more than one construct assessed by the measure.

Table 4 categorizes the instruments surveyed by type/version, describes the names used by the authors, the type of stimulus, and whether the format was conventional (generally using printed cards) or computerized. Among the versions used, the most frequent was the Stroop Color-Word Test version (124 studies; 125 tasks), followed by the Stroop Figures Test (14), the Stroop Numerical Test (7), and the Stroop Emotional Test (2). One study did not provide enough information to allow the proper identification of the instrument, version or stimuli used. Regarding the number of versions/tasks, 22.14% were computerized (used in 22.44% of the studies), emphasizing the Stroop Figures Test version in which all tasks were computerized.

Table 4

Instruments/Versions Used in the Studies Reviewed

Type/Version	No	%*	Mentioned Names	Stimulus	Format	No.
The Stroop Color–Word Test	125	85.03	Stroop Color–Word, Stroop Classic, Stroop Victoria, Stroop Computerized, Stroop Test, Stroop Neuropsychological Screening Test	Colors and words	Conventional / Computerized	107/ 18
The Stroop Figures Test	14	9.52	Semantic Stroop, M–Stroop, Stroop Day and Night	Figures: day, night, boy, girl, big, small; green and red arrows	Computerized	14
The Stroop Numerical	7	4.76	Five Digit Test (FDT) and Stroop Numerical	Numbers from 1 to 5	Conventional / Computerized	6/ 1
The Stroop Emotional Test	2	1.36	Stroop Emotional and Emotional Regulation Test	Anxiogenic words and neutral words	Conventional	2
Not Enough Information	1	0.68	Stroop "Dodrill" version	Not included	Not included	1

Note: The sum and % exceed the total number of the reviewed studies, as two used more than one instrument based on the Stroop paradigm.
* Percentage calculated based on the number of studies (n=147) – ex: 85.03% of the studies used the Stroop Color–Word version.

Of the 124 studies that specified the type as the "traditional" Stroop Color–Word Test, 90 indicated the version with references and description, and the remaining 34 did not specify the version. Among those that defined the version that served as the basis for adapting the task used in their study, 55 reported using the Victoria version and 9 using the Golden version. It should be noted that even considering the tasks based on the same version, there was heterogeneity among the descriptions. For example, considering the Victoria version, nine different descriptions were identified in terms of the number of stimuli (generally 24, however, some studies did not provide this information), types of stimuli (rectangles or circles), colors used, types of words used (among names of colors or neutral words or both) and even the order of the steps.

This lack of uniformity is even more evident when considering the measures and indexes used, with 17 different performance measures identified and studies that did not specify the index used. Approximately 60% of the studies recorded the completion time per stage. The number of errors per stage varied in 42.86% of the studies, therefore being the most commonly observed measure (Table 5). Indexes with reaction times were relatively infrequent since the measure's specificity requires computerized versions.

Table 5

Scores Used, as Reported by the Reviewed Studies

Index/Measure Type	No.	%
Completion time per step	88	59.86
Number of errors per step	63	42.86
Interference time from completion time(s)*	18	12.24
Total correct responses per step	17	11.56
Total completion time	16	10.88
Reaction time per step	15	10.20
Number of errors in the incongruous step	11	7.48
Interference time minus reaction time (ms)**	11	7.48
Total number of errors	11	7.48
Interference score of correct responses***	7	4.76
Total correct responses	5	3.40
Completion time in the incongruous step	3	2.04
Total correct responses in the incongruous step	3	2.04
Number of items correctly named in 45s in each step	3	2.04
Reaction time in the incongruous step	2	1.36
Omissions per step	1	0.68
Self-corrections per step	1	0.68
Did not specify the score used	2	1.36

Note: The sum and % exceed the total number of the reviewed studies, as they may use more than one type of index or measure (score or time) of performance.

*(*Completion* time in seconds in the incongruent step) – (Completion time in seconds in the congruent step).

**(*Reaction* Time in milliseconds in the incongruent step) – (Reaction Time in milliseconds in the congruent step).

***(*Total* number of correct responses in the incongruous step) – (Total number of correct responses in the congruent step).

Discussion

This study aimed to survey the use of tasks based on the Stroop paradigm in Brazilian empirical studies from 2000 to 2022. The scoping review identified that the paradigm had been used with different age groups, with typical development and various clinical conditions, with the predominance of the classic Stroop Color-Word Test version, not computerized, to assess inhibitory control and attention.

Regarding the age groups of the studies reviewed, there was a higher concentration of Stroop use among older adults and adults, with less use in children and adolescents. Although the development and maturation of executive functions have their highest peak in childhood and adolescence, few studies with this age group were found. Among these, the study by Duncan (2006) assessed 132 adolescents aged between 12 and 14 years and provided norms for that age for the Victoria version. With children, studies with semantic Stroop paradigms stand out, such as the one by Elage (2016), which evaluated 51 children using the day-night/boy-girl conflict, and the one by Martins (2020), with 780 children using animals (body/head conflict), both in the

age group between 4 and 10 years. One aspect that could explain the smaller number of studies with children may be that the most used version is based on the automation of the reading process (Silva et al., 2017) and that clinical studies are also more present in adulthood (substance use, dementia, neurological damage).

It was found that the traditional Stroop Color-Word Test version was the most used (in 85.03% of the studies), followed by the Stroop Figures, the Stroop Numerical, and the Stroop Emotional Tests versions. This data corroborates the literature that indicates a wide use of the original Color-Word version (Campanholo, 2018; Ramos & Hamdan, 2016; Santana et al., 2019; Silva et al., 2017) while highlighting the heterogeneity of the versions derived from it. The presentation of the items in the studies was quite varied, even in the Stroop Color-Word Test version. The items presented rectangles, circles, and words (neutral or color names) in black, different colors, and varied configurations and orders. This is an aspect to be considered since using other versions, or stimulus configurations/presentations of the same version may compromise the comparability among findings from different studies, with some standardization being desirable.

The availability of a standardized version (and its respective psychometric studies and standards) remains a requirement for researchers working with the instrument. Other versions are also popular in the literature and seek to exclude stimuli with reading demands, using mainly numbers (De Paula et al., 2017) or semantic knowledge conflicts, such as the Stroop Day and Night Test (Carlson, 2005), also identified in the review. These versions offer alternatives for the assessment of illiterate children or even adults.

There was significant heterogeneity in the skills assessed, despite inhibitory control and attention and its subcomponents being the constructs most mentioned, as found by other authors (Silva et al., 2017). This heterogeneity is predictable, considering the range of cognitive processes involved in performing the Stroop and even the different versions that may require specific cognitive demands. Although widely described for the assessment of "aspects" of executive functions (Campanholo, 2018; Cazassa et al., 2020; Guerra et al., 2021; Klein et al., 2010), there are also descriptions of the paradigm, including the assessment of executive function components, such as working memory, cognitive flexibility and even processing speed (Strauss et al., 2006). For example, there are versions, such as the Five Digit Test, a type of Stroop Numerical Test (De Paula et al., 2017; Sedó, 2004), in which there is an alternation of the rule and the demand when faced with the specificities of the stimulus presentation (e.g., count the digits presented in the box, but, in those with a border, change the rule and read the digits), engaging processes of flexibility beyond inhibition.

Heterogeneity was also observed in the number of clinical conditions composing the samples, suggesting that the Stroop is versatile and can be applied in research and professional practice and different contexts with different types of demands (developmental disorders, general physical health conditions, chronic diseases, dementia, substance use, neurological injuries, etc.). These aspects show that the instrument continues to be widely used in the ANP (Rabin et al., 2005; Ramos & Hamdan, 2016; Santana et al., 2019), including in samples of people

with typical development, among them samples of children and adolescents. This shows that, although in smaller numbers, there is an engagement of research groups in expanding studies on the NEPSY of children, a requirement highlighted in previous reviews.

Considering technological advances and the tendency to use computerized tests (Casaletto & Heaton, 2017) and despite the advantages that this type of version can provide to the assessment process, such as better control in task administration, easier and more accurate collection of scores and reaction and execution time measurements (Casaletto & Heaton, 2017; Parsey & Schmitter-Edgecombe, 2013) and easier data export for analysis (Schatz & Browndyke, 2002), computerized versions were used in a limited percentage (22.44%) of the selected studies. However, this application seems to be increasing, especially in assessing children using versions of the Stroop Figures Test (e.g., Elage, 2016; Martins, 2020). Years earlier, this trend was also observed in a study with North American psychologists, with the participation of 512 neuropsychologists who collected information on the use of 693 instruments, with only 40 being computerized. Although computerized instruments are less used, the study showed that younger psychologists with fewer years of experience are likelier to use them (Rabin et al., 2014). Although there has been increased use of computerized instruments in recent decades in the international scenario, especially in the North American context, conventional tests continue to be used more (Hewitt et al., 2022).

The still-nascent use of computerized versions in Brazil can be associated with numerous factors, including the lack of resources for programming tasks. However, a possible resistance to incorporating technology into the neuropsychological practice must also be considered, with this being a phenomenon identified internationally (Miller & Barr, 2017). There are initiatives in Brazil to develop platforms for online assessment (e.g., Serpa et al., 2019), a practice that would make teleneuropsychology possible, which would be very useful in times of restricted social contact, as is still the case due to the pandemic. The psychometric properties of these computerized measures should be investigated, and equivalence to the pencil and paper versions should not be inferred a priori (Bauer et al., 2012). This review is a relevant finding, highlighting the need for investments in studies with computerized instruments, which are still scarce in the national scenario.

A point to emphasize is that there were few studies with psychometric objectives (10.20%), that is, studies in which the authors adapted, constructed, or investigated evidence of validity, reliability, or provided norms for the instrument (e.g., Albertini, 2015; Brandelero & de Toni, 2015; Bueno, 2013; Campos et al., 2016; Canali et al., 2011; De Paula et al., 2017; Dias, 2009; Duncan, 2006; Elage, 2016; Fonseca et al., 2015; Guerra et al., 2021; Martins, 2020; Natale et al., 2008; Oliveira et al., 2016; Zimmermann et al., 2015). Of these, only three studies provided norms for elementary school children/adolescents (Duncan, 2006), for the 7 to 14 years age group (Oliveira et al., 2016) and the 19 to 75 years age group (Zimmermann et al., 2015).

There was also variability in the types of scores recorded by the studies. The most recorded were the completion time per step (59.86%) and the number of errors (42.86%). Only about 10% of the studies recorded reaction time associated with the limited use of computerized versions. Similar results were reported in the review by Scapina and Tangini (2017), showing that

the Stroop Color-Word Test mainly considered the parameters of speed and accuracy. However, the form of measurement varied greatly among Italian studies. The authors suggest that effective measures of the Stroop Color-Word Test should include: 1) the time and accuracy (correct responses) in each condition and 2) a global index calculated to relate the performance in the incongruent condition with word reading and color naming skills. The second aspect highlighted by the authors was not considered in the indexes found.

As a limitation of this study, it is possible that the survey of articles and T&D did not cover all the national scientific production since some journals are not indexed in the databases, and T&D may not be included in the used digital repositories. Furthermore, it was impossible to access some of the studies, despite attempts to obtain them from the authors (with no responses, as mentioned earlier in the Method section). Finally, the arbitrariness of the time frame should be noted, even though this criterion is necessary for the viability of the work. On the other hand, including T&D in the review increased the number of studies and the possibility of a more reliable mapping of the instrument usage in national research.

The findings of this review enabled the use of the Stroop paradigm in Brazilian research to be mapped. However, in line with the objective of a scoping review, these results do not provide indicators of the quality or sensitivity of the measure and its versions (Munn et al., 2018). Considering the versions identified, the indexes used, and the populations included in these studies as a starting point, future systematic reviews may deepen the understanding of more specific issues, such as the usefulness of the measure in the assessment of inhibitory control/attention in certain clinical conditions or the sensitivity of different types of indexes derived from their scores or time measures.

Final considerations

The studies in this review show significant heterogeneity in using the Stroop instruments in national research. There are different versions, with more studies published concerning the 18 to 79 age group, mainly assessing inhibitory control and attention and its subcomponents. The scores especially involve time and correct responses/errors per stage, and the studies focus on typical development and various clinical conditions. The computerized paradigm is recent and still little used in Brazil, with this being a vast field for development and research in the area. Furthermore, few psychometric studies have provided data on the validity and reliability of the tasks. The gap is even more significant concerning the availability of norms, virtually non-existent for some age groups, posing a pressing demand for future ANP investigations.

The results of this review may guide future studies into the use of the Stroop instruments and ANP-related research and practice. It is necessary to deepen the research on the usefulness of the instrument's different versions and measures/indexes in the different contexts in which the paradigm is used. There is scope to advance in the development and use of computerized versions. Finally, it is necessary to address the lack of psychometric studies, especially regarding the normative data for the Brazilian population, which limit the clinical use of the task.

References

Albertini, K. M. (2015). Desenvolvendo um protocolo "Stroop" para avaliar as habilidades de controle inibitório e flexibilidade cognitiva das funções executivas. *Dementia & Neuropsychologia*, *9*(2). https://doi.org/10.1590/1980-57642015dn92000006

Bauer, R., Iverson, G., Cernich, A., Binder, L., Ruff, R., & Naugle, R. (2012). Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology*, *27*(3), 362–373. https://doi.org/10.1093/arclin/acs027

Bertrand, E., Mograbi, D. C., Brown, R. G., Landeira-Fernandez, J., & Morris, R. G. (2019). Heterogeneity of anosognosia in Alzheimer's disease according to the object of awareness. *Psychology & Neuroscience*, *12*(2), 282–290. https://doi.org/10.1037/pne0000164

Bilder, R., & Reise, S. (2019). Neuropsychological tests of the future: How do we get there from here? *Clinical Neuropsychologist*, *33*(2), 220–245. https://doi.org/10.1080/13854046.2018.1521993

Brandelero, V. & Toni, P. (2015). Estudo de validade do teste Stroop de cores e palavras para controle inibitório. *Psicologia Argumento*, *33*(80), 282–297. https://doi.org/10.7213/psicol.argum.33.080.AO05

Bueno, J. M. H. (2013). Construção e validação de um instrumento para avaliação da regulação emocional. *Estudos Interdisciplinares em Psicologia*, *4*(2), 186–200. https://doi.org/10.5433/2236-6407.2013v4n2p186

Campanholo, K., (2018). Teste Stroop. In E. Miotto, K. Campanholo, V. Serrão, & B. Trevisan (Eds.), *Manual de Avaliação Neuropsicológica: A prática da testagem cognitiva* (pp. 87–94). Memnon.

Campos, M. C., Silva, M. L., Florêncio, N. C., & Paula, J. J. de. (2016). Confiabilidade do Teste dos Cinco Dígitos em adultos brasileiros. *Jornal Brasileiro de Psiquiatria*, *65*(2), 135–139. https://doi.org/10.1590/0047-2085000000114

Canali, F., Brucki, S. M. D., Bertolucci, P. H. F., & Bueno, O. F. A. (2011). Reliability study of the Behavioral Assessment of the Dysexecutive Syndrome adapted for a Brazilian sample of older-adult controls and probable early Alzheimer's disease patients. *Revista Brasileira de Psiquiatria*, *33*(4), 338–346. https://doi.org/10.1590/s1516-44462011005000015

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Casaletto, K. B. & Heaton, R. K. (2017). Neuropsychological Assessment: Past and Future. *Journal of the International Neuropsychological Society*, *23*(9/10), 778–790. doi:10.1017/S1355617717001060

Cazassa, M. J., Oliveira, M. D. S., Spahr, C. M., Shields, G. S., & Slavich, G. M. (2020). The Stress and Adversity Inventory for Adults (Adult STRAIN) in Brazilian Portuguese: Initial validation and links with executive function, sleep, and mental and physical health. *Frontiers in Psychology*, *10*, 3083. https://doi.org/10.3389/fpsyg.2019.03083

De Paula, J. J., Oliveira, T. D., Querino, E. H. G., & Malloy-Diniz, L. F. (2017). The Five Digits Test in the assessment of older adults with low formal education: construct validity and reliability in a Brazilian clinical sample. *Trends in Psychiatry and Psychotherapy*, *39*(3), 173–179. https://doi.org/10.1590/2237-6089-2016-0060

Dias, N. M. (2009). *Avaliação neuropsicológica das funções executivas: tendências desenvolvimentais e evidências de validade de instrumentos* [Master's Thesis]. Universidade Presbiteriana Mackenzie, São Paulo, Brasil.

Dias, N. M. & Seabra, A. G. (2019). Avaliação neuropsicológica e seu papel no direcionamento da intervenção. In C. O. Cardoso & N. M. Dias (Eds.), *Intervenção neuropsicológica infantil: Da estimulação precoce-preventiva à reabilitação* (pp. 29–70). Pearson.

Duncan, M. T. (2006). Obtenção de dados normativos para desempenho no teste de Stroop num grupo de estudantes do ensino fundamental em Niterói. *Jornal Brasileiro de Psiquiatria*, 55(1), 42–48. https://doi.org/10.1590/S0047-20852006000100006

Elage, G. K. C. F. (2016). *Análise das propriedades psicométricas de uma bateria de testes informatizados para avaliação das funções executivas em crianças de 4 a 10 anos* [Master's. Thesis]. Universidade Presbiteriana Mackenzie, São Paulo, Brazil.

Fonseca, G. U., Lima, R. F. de, Ims, R. E., Coelho, D. G., & Ciasca, S. M. (2015). Evidências de validade para instrumentos de atenção e funções executivas e relação com desempenho escolar. *Temas em Psicologia*, *23*(4), 843–858. https://doi.org/10.9788/TP2015.4-04

Goldstein, L. H., & Mcneil, J. E. (2013). *Clinical Neuropsychology: A practical guide to assessment and management for clinicians* (2th ed.). Wiley-Blackwell.

Guerra, A., Hazin, I., Roulin, J. L., Gall, D. L., & Roy, A. (2021). Pieces of evidence of reliability of the Brazilian version of the Child Executive Functions Battery (CEF-B). *Psicologia: Reflexão e Crítica*, *34*(6),1–8. https://doi.org/10.1186/s41155-021-00171-2

Hewitt, K. C., Block, C., Bellone, J. A., Dawson, E. L., Garcia, P., Gerstenecker, Gerstenecker, A., Grabyan, J. M., Howard, C., Kamath, V., LeMonda, B. C., Margolis, S. A., McBride, W. F., Salinas, C. M., Tam, D. M., Walker, K. A., & Del Bene, V. A. (2022). Diverse experiences and approaches to tele neuropsychology: Commentary and reflections over the past year of COVID-19. *The Clinical Neuropsychologist*, *36*(4). https://doi.org/10.1080/13854046.2022.2027022

Klein, M., Adda, C., Miotto, E. C., Lucia, M. C., & Scaff, M. (2010). O paradigma Stroop em uma amostra de idosos brasileiros. *Psicologia Hospitalar*, *8*(1), 93–112.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.

Malloy-Diniz, L. F., Mattos, P., Abreu, N., & Fuentes, D. (2016). O exame neuropsicológico: o que é e para que serve? In L. Malloy-Diniz, P. Mattos, N. Abreu & D. Fuentes (Eds.), *Neuropsicologia: Aplicações clínicas* (pp. 21–34). Artmed.

Martins, C. R. (2020). *Evidências de validade do teste de inibição e flexibilidade cognitiva – TIF: Analisando o desenvolvimento do controle inibitório e da flexibilidade cognitiva na infância*. Dissertação de Mestrado, Universidade Presbiteriana Mackenzie, São Paulo, Brasil.

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology*, *32*(5), 541–554. https://doi.org/10.1093/arclin/acx050

Munn, Z., Peters, M., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*, 143. https://doi.org/10.1186/s12874-018-0611-x

Natale, L. L., Teodoro, M. L. M., Barreto, G. D. V., & Haase, V. G. (2008). Propriedades psicométricas de tarefas para avaliar funções executivas em pré-escolares. *Psicologia em Pesquisa*, *2*(2), 23–35.

Oliveira, R. M., Mograbi, D. C., Gabrig, I. A., & Charchat-Fichman, H. (2016). Normative data and evidence of validity for the Rey Auditory Verbal Learning Test, Verbal Fluency Test, and Stroop Test with Brazilian children. *Psychology and Neuroscience*, *9*(1), 54–67. https://doi.org/10.1037/pne0000041

Parsey, C. M. & Schmitter-Edgecombe, M. (2013). Applications of Technology in Neuropsychological Assessment. *The Clinical Neuropsychologist*, *27*(8), 1328–1361. https://doi.org/10.1080/13854046 .2013.834971

Rabin, L. A., Barr, W. B. & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65. https://doi.org/10.1016/j.acn.2004.02.005

Rabin, L. A., Spadaccini, A. T., Brodale, D. L., Grant, K. S., Elbulok-Charcape, M. M., & Barr, W. B. (2014). Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. *Professional Psychology: Research and Practice*, *45*(5), 368–377. https://doi.org/10.1037/a0037987

Ramos, A. A., & Hamdan, A. C. (2016). O crescimento da avaliação neuropsicológica no Brasil: Uma revisão sistemática. *Psicologia: Ciência e Profissão*, *36*(2), 471–485. https://doi.org/10.1590/1982-3703001792013

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. *Revista Avaliação Psicológica*, *18*(1), 96–107. https://doi.org/10.15689/ap.2019.1801.14668.11

Scarpina, F., & Tagini, S. (2017). The stroop color and word test. *Frontiers in Psychology*, *8*, 557. https://doi.org/10.3389/fpsyg.2017.00557

Schatz, P. & Browndyke, J. (2002). Applications of Computer-based Neuropsychological Assessment. *Journal of Head Trauma Rehabilitation*, *17*(5), 395–410.

Sedó, M. A. (2004). '5 digit test': A multilinguistic non-reading alternative to the Stroop test. Revista de Neurologia, 38(9), 824–828. https://doi.org/10.33588/rn.3809.2003545

Serpa, A. L., Timóteo, A. P. P., Querino, E. H. G., & Malloy-Diniz, L. F. (2019). Desenvolvimento do teste de planejamento Torre de Londres–versão brasileira (TOL-BR). Debates em Psiquiatria, 9(4), 10–19. https://doi.org/10.25118/2763-9037.2019.v9.45

Silva, J., Starling-Alves, I., Moura, R., & Haase, V. G. (2017). Teste Stroop Victoria. In A. Julio-Costa, R. Moura, & V. Haase (Eds.), *Compêndio de Testes neuropsicológicos: Atenção, funções executivas e memória* (pp. 163-174). Hogrefe.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*, 467–473. https://doi.org/10.7326/M18-0850

Zimmermann, N., Cardoso, C., Trentini, C. M., Grassi-Oliveira, R., & Fonseca, R. P. (2015). Brazilian preliminary norms and investigation of age and education effects on the Modified Wisconsin Card Sorting Test, Stroop Color and Word test and Digit Span test in adults. *Dementia & Neuropsychologia*, *9*(2), 120-127. https://doi.org/10.1590/1980-57642015DN92000006