Journal of the International Neuropsychological Society (2021), **27**, 835–854
Copyright © INS. Published by Cambridge University Press, 2021.
doi:10.1017/S1355617720001241

CRITICAL REVIEW

Brief, Performance-Based Cognitive Screening in Youth Aged 12–25: A Systematic Review

Shayden D. Bryce[1,2], Stephen C. Bowden[3], Stephen J. Wood[1,2] and Kelly Allott[1,2,*]

[1]Orygen, Parkville, Victoria, Australia
[2]Centre for Youth Mental Health, The University of Melbourne, Parkville, Victoria, Australia
[3]Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria, Australia

(Received March 24, 2020; Final revision October 19, 2020; Accepted October 26, 2020; First Published Online January 19, 2021)

Abstract

Objective: Cognitive screening is an efficient method of detecting cognitive impairment in adults and may signal need for comprehensive assessment. Cognitive screening is not, however, routinely used in youth aged 12–25, limiting clinical recommendations. The aims of this review were to describe performance-based cognitive screening tools used in people aged 12–25 and the contexts of use, review screening accuracy in detecting cognitive impairment relative to an objective reference standard, and evaluate the risk of bias of included studies. **Method:** Electronic databases (Scopus, Medline, PsychINFO, and ERIC) were searched for relevant studies according to pre-determined criteria. Risk of bias was rated using the Quality Assessment of Diagnostic Accuracy Studies-2. Dual screening, extraction, and quality ratings occurred at each review phase. **Results:** Twenty studies met the review inclusion criteria. A diverse range of screening tools (length, format) were used in youth aged 12–25 with or without health conditions. Six studies investigating cognitive screening were conducted as primary accuracy studies and reported some relevant psychometric parameters (e.g., sensitivity and specificity). Fourteen studies presented correlational data to investigate the cognitive measure utility. Studies generally presented limited data on classification accuracy, which impacted full screening tool appraisal. Risk of bias was high (or unclear) in most studies with poor adherence to the Standards for Reporting Diagnostic Accuracy Studies (STARD) criteria. **Conclusions:** Few, high quality studies have investigated the utility of cognitive screening in youth aged 12–25, with no screening measure emerging as superior at detecting cognitive impairment in this age group.

Keywords Screening, Cognition, Youth, Validity, Neuropsychology

INTRODUCTION

Cognitive impairment is a feature of many diseases and injuries affecting the central nervous system. Rapid identification of impairment is often critical for diagnostic formulation as well as to inform treatment and case management. Brief cognitive screening may provide an efficient method of detection and triage to guide referral for more comprehensive investigations, such as neuropsychological assessment, as part of a stepped-care approach (Block, Johnson-Greene, Pliskin, & Boake, 2017; Lanca, 2018). Effective cognitive screening may also be used to document and monitor difficulties when comprehensive assessments are not available or possible, such as in rural settings or when there are competing medical demands (Bryce & Allott, 2019; Sheehan, 2012). The use of cognitive screening is already well-established in primary, tertiary, and community settings in the assessment of health conditions in adults, including mild cognitive impairment, dementia, acquired brain injury or disease, and psychiatric illness (Roebuck-Spencer et al., 2017). Critically, cognitive screening has to have demonstrated high diagnostic sensitivity to identify those in need of further assessment (Summers, Bondi, & Bowden, 2019). Screening sensitivity is enhanced when cognitive impairment is corroborated by observable signs or symptoms, or when family or other caregivers express concern over decline (Laver et al., 2016; Pottie et al., 2016). Classic cognitive screening tools used in older adults with suspected neurological illness include the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and Montreal Cognitive Assessment

*Correspondence and reprint requests to: Kelly Allott, 35 Poplar Road, Parkville, VIC 3052, Australia. E-mail: kelly.allott@orygen.org.au

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

(MoCA; Nasreddine et al., 2005). The use of these measures in adults has been reviewed elsewhere (Arevalo-Rodriguez et al., 2015; Cullen, O'Neill, Evans, Coen, & Lawlor, 2007; Davis et al., 2015; Holsinger, Deveau, Boustani, & Williams Jr, 2007; Tsoi, Chan, Hirai, Wong, & Kwok, 2015). Notably, there is variability in the risk of bias in reviewed studies, indicating the potential for inflated estimates of diagnostic validity (e.g., Tsoi et al., 2015).

In contrast to adults, cognitive screening in youth, aged 12–25, is not routinely conducted in clinical care (Bryce & Allott, 2019). Evidence-based recommendations regarding suitable cognitive screening measures are lacking in this age group (Bryce & Allott, 2019; Roebuck-Spencer et al., 2017; Summers, Bondi, & Bowden, 2019). Cognitive screening tools commonly used in adults may not be sensitive to the nature and extent of cognitive impairment, or trajectories of cognitive development, in people aged 12–25 (e.g., MoCA and MMSE).

The age range of 12–25 is an important age bracket that requires special consideration in relation to cognitive screening for at least four clear reasons. First, this age range is consistent with definitions of young people presented by the World Health Organization and United Nations (www.who.it; www.un.org), which encapsulates both adolescence (i.e., 10–19 years old) and youth (i.e., 15–24 years old). The ages of 12–25 therefore capture most individuals from the start of secondary school into young adulthood. Secondly, this age range is a particularly vulnerable time for cognitive health. Of critical importance for cognitive screening, young people continue a trajectory of dynamic structural and functional cortical and subcortical maturation between the ages of 12–25 (Bava et al., 2010; Gogtay et al., 2004; Foulkes & Blakemore, 2018), with neural development traversing traditional definitions of pediatric (i.e., <18 years old) and adult samples (i.e., ≥18 years old). This period of growth coincides with expansion of cognitive functions throughout this age range, emphasizing the importance of up-to-date normative data for individuals at each stage of development (Arain et al., 2013; Larsen & Luna, 2018; Weintraub et al., 2013). The rate of improvement of some cognitive skills may begin to plateau at ages 12–15 (e.g., reading); however, most fluid and crystallized domains continue to develop throughout early adulthood (Weintraub et al., 2013).

Thirdly, between the ages of 12 and 25, youth are at increased risk of developing mental illness and substance abuse or sustaining brain injury associated with elevated risk-taking behaviors (Allott et al., 2013; Arain et al., 2013). Common health conditions affecting youth may stem from subthreshold or threshold mental illnesses (e.g., mood, substance use, and psychosis-spectrum disorders; Gore et al., 2011). Acquired or congenital neurological or medical conditions [e.g., epilepsy and traumatic brain injury (TBI)] and undetected neurodevelopmental disorders [e.g., intellectual disability and attention-deficit/hyperactivity disorder (ADHD)] are also common in this age group (Gore et al., 2011; Kassebaum et al., 2017; Proffitt, Brewer, Parrish, McGorry & Allott, 2018; Vos et al., 2015). In addition, functional demands are likely to increase as a young person transitions into adulthood (i.e., independent living). These demands may exceed a young person's functional capacity, revealing milder cognitive difficulties that have gone undetected earlier in life (Allott et al., 2013). Accordingly, there has been some shift from traditional pediatric (<18 years) and adult (≥18 years) health services, to youth-focused healthcare services that encapsulate the dynamic period of development and period of increased risk for the onset of brain-related conditions during this phase of life. This provides a fourth reason for focusing on this age range in relation to cognitive screening and has been particularly evident in mental health services, with intake criteria catering for the adolescent to young adult age range. For example, *headspace*, Australia's federally funded primary mental health service (https://headspace.org.au/), *Youth One Stop Shop (YOSS)* in New Zealand (http://www.yoss.org.nz/), *Foundry* in Canada (https://foundrybc.ca/), and *jigsaw* in the Republic of Ireland (https://www.jigsaw.ie), are all accessible for people aged 12–25. *Maisons des Adolescents* in France as well as *The Junction* in Scotland are accessible to people aged 11–21 requiring mental health support (www.monenfant.frl; https://the-junction.org/). Further, EASA, an early psychosis service in the USA is designed for people aged 15–25 (http://www.easacommunity.org). Many of these services are discussed in Hetrick et al., (2015).

Patterns of cognitive impairment are highly heterogeneous in disorders affecting youth, particularly in the context of personal illness and environmental factors (e.g., premorbid learning style, age of onset, or social milieu; Doyle et al., 2018; Kavanaugh, Cancilliere, & Spirito, 2019). Nevertheless, cognitive deficits across various health conditions are generally found to predict real-world functioning including educational and occupational, interpersonal, independent living, and psychological domains (Arnett et al., 2013; Kerr & Fayed, 2017; Lee et al., 2018). Cognitive difficulties in youth may also increase caregiver burden and interfere with the effectiveness of medical and psychological interventions (Gonzalez-Heydrich et al., 2014; Hetrick et al., 2015; Patel, Wong, Cuevas, & Van Horn, 2013). Thus, similar to older adults, cognitive screening of youth with suspected cognitive deficits may provide an efficient means for early identification and intervention. Screening may also mitigate potential functional impairment related to undiagnosed cognitive impairment or decline (e.g., diminishing academic achievement and poor vocational participation). Early detection of cognitive impairment during this phase of life can set in motion personalized interventions and services that optimize a young person's developmental pathway into adulthood. The absence of evidence-based recommendations regarding appropriate cognitive screening tools in people aged 12–25 is therefore an unmet clinical need (Bryce & Allott, 2019).

The American Psychological Association (APA) Practice Organization Working Group on Screening and Assessment (2014; APA-WGSA) provided general guidelines for

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

distinguishing cognitive screening from comprehensive neuropsychological assessment in mental and behavioral health conditions (www.apaservices.org/practice/reimbursement/billing/assessment-screening). These guidelines were produced to address the greater uptake of cognitive screening in general healthcare settings and provide an initial foundation to discuss cognitive screening in youth in the absence of age-specific recommendations (Block et al., 2017; Roebuck-Spencer et al., 2017). With respect to these guidelines, cognitive screening tools in youth should be (1) brief (e.g., <30 min), (2) adequately normed to account for a wide range in cognitive development, (3) sensitive to the diverse range of cognitive difficulties experienced in youth with various health conditions, and (4) administered by health practitioners of various professions (Bryce & Allott, 2019).

To be useful, cognitive screening tools should have high classification accuracy relative to a gold-standard such as a neuropsychological assessment (Bunnage, 2017). That is, for youth with suspected cognitive issues, a demonstrated ability to correctly detect true cases of cognitive impairment (i.e., high sensitivity) and rule out false cases of impairment (i.e., high specificity; Bunnage, 2017). There is a natural trade-off between sensitivity and specificity for any given cut-off score on a cognitive test; that is, specificity decreases as sensitivity increases and *vice versa*. Screening tools in youth that balance sensitivity and specificity may be particularly valuable when comprehensive assessment resources are limited (Bauer et al., 2012; Cullen et al., 2007). For example, elevated rates of psychological distress in vulnerable youth may result from screening measures generating elevated false positive test results, while missed treatment opportunities may occur if false negative results are high (Roebuck-Spencer et al., 2017). Cognitive screening tools should also provide adequate construct coverage, with the tool measuring a representative range of cognitive abilities in youth (Bunnage, 2017).

Objective

The aims of this systematic review were to (1) describe the types of cognitive screening measures used in people aged 12–25, (2) identify the settings in which these tools are utilized, (3) review the diagnostic accuracy of these screening measures in relation to detecting cognitive impairment according to an accepted objective reference standard, and (4) evaluate the methodology of these studies for risk of bias. It was envisaged that this review would inform recommendations for cognitive screening in youth.

METHOD

Search Strategy

The reporting standards of this review were guided by the PRISMA statement (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). Searches of systematic review registries (e.g., Cochrane Database of Systematic Reviews and PROSPERO) did not reveal any similar reviews underway as of November 2018. The protocol for this review was prospectively registered with PROSPERO on August 27, 2019 (ID: CRD42019133918; www.crd.york.ac.uk/prospero). Article searches were conducted using the online databases OvidMEDLINE, PsychINFO, Scopus, and ERIC. Title, abstract, and keywords were examined in each database using combinations of the following terms: *adolescen**, *teen**, *young adult**, *young people*, *youth*, *cognit**, *neurocognit**, *MMSE**, *MoCA*, and *screen**. The *MMSE** and *MoCA* terms were included as (1) these are two of the most commonly used cognitive screening tools across health populations and (2) articles containing these terms were commonly retrieved during preliminary searches. Final database searching was conducted on December 3, 2018 (see Supplementary Material for final combinations for each database). No search limits were applied in any database. Eligibility was assessed manually according to specified inclusion criteria. Reference lists of all included studies and selected reviews on cognitive screening and computerized assessment were also screened for relevant articles (e.g., Arrieux, Cole, & Ahrens, 2017; Cernich, Reeves, Sun, & Bleiberg, 2007).

Inclusion Criteria

Studies were included for review if they met the following criteria (1) original peer-reviewed research article published, or *in press*, in English, (2) participants were youth, aged between 12 and 25 years inclusive, as indicated by sample mean, (3) the study reported on a tool designed to screen for cognitive functioning, or the tool was used for this purpose, either as an isolated measure or embedded within a broader screening process, and (4) studies examined, or provided data that supported inferences regarding, the criterion validity of the specified screening tool(s) relative to an accepted standardized criterion of cognitive functioning. Guided by the 2014 APA-WGSA statement, the studies were also included if the screening tool being investigated (1) was brief to administer (i.e., ≤30 min), (2) collected objective test data, (3) required minimal, non-specialist training to utilize, (4) was not integrated with historical or functional information, and (5) was not used for diagnostic purposes (e.g., concussion status). Definitions of cognitive impairment were not used as an inclusion criterion due to potential differences between studies (e.g., 1 standard deviation below the general population on two domains, 2 standard deviations below the general population on one domain, etc.). Further, no search terms related to impairment domains were specified due to the broad range of difficulties anticipated within this age group across various health conditions (i.e., intelligence, memory, etc.). No restriction on the type of analysis was applied [e.g., correlations, regression, sensitivity or specificity analysis, receiver operating characteristic (ROC) curve analysis, etc.] to capture a range of study types.

So as to increase the representativeness of the studies included in the review, additional qualifying criteria were applied in view of the number of studies with imprecise

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

reporting of participant age and variable assessment tools used as cognitive screeners. Specifically, where the mean age was not reported, studies were still considered for inclusion if greater than 50% of the sample were aged 12–25 years inclusive. Further, when only age bands were presented (e.g., 18–26), studies were still considered for inclusion if the midpoint of the age band was less than 26. Studies that did not specifically report any objective indication of participant age were not considered for inclusion, even if age might be assumed (e.g., college samples). Studies using collections of established neuropsychological tasks as screening tools were considered for inclusion if (1) at least two tasks measuring different domain-specific Cattell-Horn-Cattell constructs were used (for an overview, see Jewsbury, Bowden, & Duff, 2017; Jewsbury, Bowden, & Strauss, 2016; McGrew, 2009), (2) minimal training was required, and (3) clear guidelines for interpretation were provided. Finally, given the exploratory nature of the review, studies were still considered for inclusion if they examined measures that were not originally designed for cognitive screening or clinical purposes, or required some psychometric training. Screening measures based on self or informant report were not included.

Screening, Extraction, and Risk of Bias

Two authors, SDB and KA, independently screened titles and abstracts for articles meeting inclusion criteria for full-text eligibility (Phase 1). Full-text articles were retrieved when eligibility was uncertain. The same two authors independently reviewed each full-text article retrieved in Phase 1 using the same review criteria (Phase 2). Screening discussions were held after piloting the review criteria in Phases 1 and 2 to mitigate potential for bias early in the screening process. When disagreements regarding article suitability in Phase 1 and 2 could not be resolved between the two screening authors; a third author, SCB, was consulted and disagreement were resolved by consensus. Additional studies cited in the reference lists of included studies, or in reference lists of studies excluded according to the design criterion only (criterion 4 above), were considered for inclusion if they met the above inclusion criteria. This additional process was implemented during screening to increase the likelihood that studies relevant to the current review were detected.

Dual data extraction from articles meeting criteria was conducted by authors SDB and KA. Variables of interest included year of publication, country, study design, recruitment method, sample size, participant demographics (e.g., age, sex, and education), index test, reference standard, and any relevant reported psychometric data [e.g., sensitivity and specificity, area under the curve (AUC), likelihood ratios, correlation coefficients, etc.]. Included studies were also assessed for risk of bias and applicability by the same two authors using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool according to scoring guidelines presented in Whiting et al. (2011). Data extraction and risk of bias and applicability ratings both followed the same dual, independent examination and consensus procedures as described for article screening.

Risk of Bias Assessment Using QUADAS-2

The QUADAS-2 assessment tool was designed to assist the evaluation of study quality in diagnostic accuracy studies (www.quadas.org; Whiting et al., 2011). Scale development was guided by STARD (Standards for Reporting Diagnostic Accuracy Studies) criteria, which were developed to improve the completeness and transparency of reporting classification accuracy studies (Bossuyt et al., 2003, 2015; Cohen et al., 2016). The QUADAS-2 has been recommended in reviews of diagnostic accuracy by the Cochrane Collaboration (www.cochrane.org) and the National Institute for Health and Clinical Excellence (NICE; www.nice.org.uk). This scale is used in four structured phases (1) summarize the review question, (2) tailor the tool to the review, (3) construct a flow diagram for the primary study, and (4) assess risk of bias and concerns regarding applicability. Risk of bias is assessed in four domains (1) participant selection, (2) index tests, (3) reference standard, and (4) flow and timing (i.e., flow of participants through the study and timing of administration of index and reference tests). Each domain contains signaling questions that highlight elements of study design that may introduce bias. Signaling questions deemed applicable to this review are rated as *"Yes"*, "*No*", or "*Unclear*". Risk of bias may be judged as "*Low*" if all relevant signaling questions are rated "*Yes*", yet may exist if these questions are rated as "*No*" or "*Unclear*". Concerns regarding applicability may be present if related questions are rated as "*High*" or "*Unclear*".

Statistical Interpretation

The sensitivity of a cognitive screening tool reflects its ability to correctly detect individuals with cognitive impairment, while specificity is its ability to correctly exclude those without cognitive impairment (Bunnage, 2017; Lalkhen & McCluskey, 2008; Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008). The rate of true positives (sensitivity) can be plotted against the false positive rate (1-specificity) for given cut-off scores (or decision threshold) of a screening measure to generate a ROC curve. The AUC can be used as an interpretive guide to overall classification accuracy (Mandrekar, 2010).

Sensitivity and specificity are also used to calculate predictive power and likelihood ratios. Positive predictive power indicates the probability of an individual with a positive test result (i.e., score suggestive of cognitive impairment) actually being impaired, whereas negative predictive power indicates the probability that an individual with a negative test result (i.e., score not suggestive of cognitive impairment) is not impaired. Predictive values are dependent on the prevalence of impairment in the target population (Bunnage, 2017). Positive likelihood ratios (i.e., sensitivity/100-specifity) are another way of interpreting the diagnostic implications of a

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

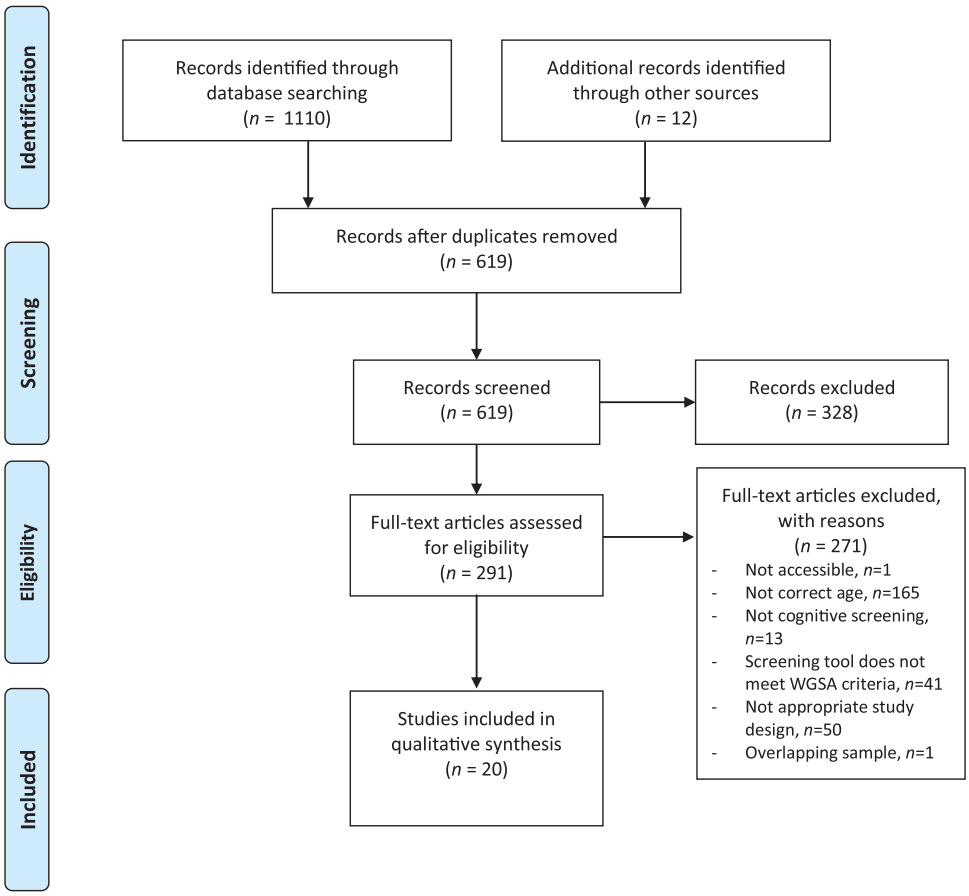


Fig. 1. PRISMA flow diagram.

score or score range and make use of both sensitivity and specificity information. A positive likelihood ratio greater than 10 is used as strong evidence to "rule in" impairment when the test result is positive. In contrast, negative likelihood ratios (i.e., 100-sensitivty/specificity) indicate how much less likely it is that the person has cognitive impairment at a given score or score range. A negative likelihood ratio below .1 is used as strong evidence to "rule out" impairment when the test result is negative (Bunnage, 2017). For studies examining correlations between cognitive screening measures and reference standards, correlation coefficients reaching at least .1, .3, and .5 can be described as small, medium, and large, respectively (Cohen, 1988).

RESULTS

Search Results

Searches of electronic databases and reviews of additional sources retrieved 1122 records, with 619 studies remaining after removing duplicates. After title and abstract screening (Phase 1), 291 articles were subject to full-text review. Of the full-text articles reviewed, 271 were excluded (Phase 2), primarily for not meeting the age criterion (61%). Twenty articles met review criteria and were included in the qualitative synthesis (see Figure 1).

Study Characteristics

The characteristics of the 20 included studies and an overview of the screening measures retrieved are presented in Tables 1 and 2, respectively. Most studies were conducted in North America (85%, $n = 17$). The average sample size was 116 ($SD = 77$; range: 28–279) participants. The median of the average ages of participants across each of the included studies was 17 years old. Fourteen different cognitive screening measures were described in the reviewed studies. These measures varied in assessment length (5–30 min), format (pen-and-paper or computerized) and cognitive domains evaluated, as well as initially intended populations (e.g., military- or sport-related concussion or TBI), neuropsychiatry, neurology, geriatrics, oncology.

Cognitive screening tools were used across multiple target populations including healthy people ($k = 7$, 35%), concussion or mild TBI ($k = 3$, 15%), neurological conditions (including epilepsy, TBI, brain tumor; $k = 3$, 15%), other oncology ($k = 2$, 10%), autoimmune conditions including multiple sclerosis and systemic lupus erythematosus ($k = 2$, 10%), congenital heart disease ($k = 1$, 5%), mental illness ($k = 1$, 5%), and ADHD ($k = 1$, 5%). The most commonly used screening tools in the included studies were the computerized Immediate Post-Concussion Assessment and Testing (ImPACT; $k = 5$) and Automated Neuropsychological

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 1. Study characteristics of articles included for review

First author, year (country)	Target population (clinical setting, if applicable)	Study design	Recruitment method	Sample size†	Mean age, years (*SD*; range)	Sex, % male	Education level, years (*SD*; range)
Allen, 2011 (USA)	Undergraduate university students	Cross-sectional	Convenience	100	19.69 (1.19; NR)	44	13.21 (1.09; NR)
Bleiberg, 2000 (USA)	High school/college students and student athletes	Cross-sectional	NR	122	17.2 (2.78; 15–27)	64	11.8 (2.13; 9–18)
Boulet-Craig, 2018 (Canada)	Oncology, Acute Lymphoblastic Leukemia (outpatient)	Longitudinal (3 year follow-up)	Convenience	247	21.6 (6.3; 8–38)	49	11.2 (2.5; NR)
Gualtieri, 2006 (USA)	ADHD (NR)	Cross-sectional	NR	272	ADHD-treated: 13.08 (NR; NR); ADHD-untreated: 13.54 (NR; NR)	NR	NR
Heaton, 2014 (USA)¶	Community-based adults	Cross-sectional	Convenience	92	22.4 (1.48; 20–24)	36	12.9 (2.42; NR)
Holzer, 2007 (Switzerland)	Mental illness (psychosis and other illness; outpatient)	Cross-sectional	Convenience	80	Psychosis: 16.2 (1.4; 13–18); Other: 15.7 (1.4; 13–18)	Psychosis: 72; Other: 46	NR
Hoppe, 2009 (Germany)	Epilepsy (outpatient)	Cross-sectional	Consecutive	212	NR (NR; 87 people aged 16–29)	NR	NR
Iverson, 2005 (USA/Canada)	Sport-related concussion (outpatient)	Cross-sectional	NR	72	17.1 (1.9; NR)	83	10.5 (1.8; NR)
Jones, 2008 (USA)	Undergraduate university students	Cross-sectional	Convenience	77	24.38 (6.77; 19–56)	23	NR
Kahn, 2015 (USA)	TBI (inpatient)	Cross-sectional	Convenience (retrospective)	56	15.95 (1.63; 12–17)	63	NR
Kelly, 2012 (USA/Iraq)	Combat-related concussion (outpatient)	Cross-sectional	Convenience	66	25 (NR; IQR: 22–30)§	100	NR
Krull, 2008 (USA)	Oncology, mixed (outpatient)	Longitudinal (1–2 years follow up)	Consecutive	240	12.6 (3.72; 6–18)	60	NR
Maerlender, 2010 (USA)	College athletes	Cross-sectional	NR	54	19.1 (NR; 17–22)	100	NR
McLeod, 2006 (USA)	Community athletes	Cross-sectional	Convenience	28	Males: 13.24 (0.87; 12–14); Females: 13.29 (0.83; 12–14)	39	NR
Pike, 2017 (USA)	Congenital heart disease (outpatient)	Cross-sectional	Convenience	80	16.8 (2.1; 14–21)	59	NR (68% in Grade 9–11)
Portaccio, 2009 (Italy)	Multiple sclerosis (outpatient)	Cross-sectional	Consecutive	61	15.4 (2.4; 8.8–17.9)	49	9.3 (2.5; 2–13)
Raiker, 2015 (USA)	TBI and brain tumor (outpatient)	Longitudinal (up to 15 months follow up)	Convenience (chart review)	77	11.4‡ (3.5; NR)	53	NR
Schatz, 2010 (USA)	Undergraduate university students	Cross-sectional	Convenience	30	21 (NR; 18–23)	47	NR
Solomon, 2014 (USA)	Sport-related concussion history (outpatient)	Cross-sectional	Convenience	219	22.36 (1.18; NR)	100	15.59 (1.78; NR)
Vega-Fernandez, 2015 (USA)	Systemic lupus erythematosus (outpatient)	Longitudinal (up to 18 months follow up)	NR	126	16.17 (2.3; NR)	18	NR (80% in "high school")

Note: Age, percentage male and education reported from primary analysis only; ADHD = Attention-Deficit/Hyperactivity Disorder; NR = Not reported; IQR = Interquartile range; *SD* = Standard Deviation. ‡ Total sample below 12 years of age, yet TBI sample (66% of sample) mean age = 12.2 (4.1) and authors report child and adolescent data (mean age = NR).§ Median. † Sample size of population group of interest (not comparison group, if applicable). ¶ Demographic information presented in this table was obtained from the corresponding author and pertains to youth aged 20–24. Normative data in this paper was calculated based on the total sample [i.e., *N* = 268, Mean age = 52.3 (*SD* = 21.0; range: 20–85)].

Table 2. Overview of cognitive screening measures used in youth

Screening measure	Overview	Administration time (minutes)	Status in youth
Automated Neuropsychological Assessment Metrics (ANAM)	Collection of computerized cognitive assessment tasks measuring attention, reaction time, memory, processing speed, and decision making.	20–25	Normative data from people aged 17–65. Not developed for youth specifically.
Brief Neuropsychology Battery for Children (BNBC)	Cognitive assessment tool derived from traditional neuropsychological measures (e.g., vocabulary, symbol digit modalities test, trail making test, and the selective reminding test).	30	Specific to youth.
Central Nervous System – Vital Signs (CNS-VS)	Computerized cognitive assessment battery measuring attention, memory, psychomotor speed, information processing speed, and executive function (i.e., cognitive flexibility).	30	Normative data from people aged 8–90 inclusive. Not developed for youth specifically.
CogSport	Collection of computerized cognitive assessment tasks measuring attention, working memory, psychomotor speed, reaction time, problem-solving, and memory function.	15–20	Normative data from people aged 4–99. Not developed for youth specifically.
DIVERGT	Series of traditional neuropsychological measures (i.e., digit span, grooved pegboard, trail making test, and verbal fluency) measuring attention, working memory, motor dexterity, speed and executive function.	20–30	Specific to youth.
HeadMinder's Concussion Resolution Index (CRI)	Online computerized cognitive assessment battery measuring simple reaction time, complex reaction time, processing speed, and memory. Self-reported neurophysiologic symptoms, medical background, and concussion history are incorporated into assessment.	20–25	Normative data from people aged 12–40. Not developed for youth specifically.
Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT)	Computerized cognitive assessment battery measuring attention, memory, processing speed, reaction time, and impulse control.	20–25	Normative data from people aged 12–59 inclusive. Not developed for youth specifically.
Lebby-Asbell Neurocognitive Screening Examination-Adolescents (LANSE-A)	Series of brief non-computerized tasks measuring orientation, attention, working memory, visuospatial function, language, memory, and reasoning/abstraction.	20–30	Developed from people aged 11–17 inclusive.
Montreal Cognitive Assessment (MoCA)	Series of brief non-computerized cognitive tasks measuring attention, working memory, visuospatial function, language, memory, and executive functions (e.g., reasoning/abstraction, fluency, and inhibition).	10–15	Developed for adults, with normative data for people aged 55–85 inclusive. Not developed for youth specifically.
National Institute of Health Toolbox – Cognition Battery (NIHTB-CB)	Brief, computerized cognitive screening battery comprised of seven assessment tasks that measure five cognitive domains including processing speed, episodic memory, language, working memory, and attention/executive function (i.e., switching/set shifting and inhibitory control).	30	Normative data from people aged 3–85 inclusive.
NeurocognitiveEffects (NeuroCog FX)	Collection of computerized cognitive assessment tasks, designed to be similar to traditional neuropsychology measures (e.g., digit span and phonemic fluency), and experimental cognitive tasks (e.g., Go/No-Go, Reaction speed) measuring attention, working memory, verbal and figural learning, and executive functioning (i.e., fluency and flexibility).	30	Specific to youth.

(*Continued*)

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 2. (*Continued*)

Screening measure	Overview	Administration time (minutes)	Status in youth
Pediatric Automated Neuropsychological Assessment Metrics (Ped-ANAM)	Collection of computerized cognitive assessment tasks (adapted from the adult ANAM for children aged 9+) measuring attention, working memory, processing speed, visuospatial functioning, and memory.	20–30	Adapted from ANAM specifically for children aged 9 and older.
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)	Brief, pen-and-paper cognitive screening battery containing of 12 subtests that are used to measure 5 cognitive domains, including immediate memory, visuospatial/construction, language, attention, and delayed memory.	30	Normative data from people aged 20–89. Original battery not developed for youth specifically.
Standardized Assessment of Concussion (SAC)	Brief, non-computerized cognitive screen tool measuring orientation, attention, working memory, and memory. Results are interpreted alongside neurological screening (e.g., loss of consciousness) and exertional mauvers (e.g., push-ups).	5–10	Developed for post-concussion assessment. Not developed for youth specifically.

Assessment Metrics (ANAM; $k = 4$). These tools were used primarily in healthy people ($k = 6$) and those who had sustained concussion ($k = 3$).

Screening Measure Utility

Diagnostic accuracy

Each included study compared performance on at least one cognitive screening measure (the index test) and one standardized measure or battery of cognitive function (i.e., often a global composite; the reference standard). The index tests and reference standards are displayed in Table 3. Approximately one-third (6/20) of studies presented some classification accuracy data as shown in Table 4 (Boulet-Craig et al., 2018; Hoppe, Fliessbach, Schlegel, Elger, & Helmstaedter, 2009; Krull et al., 2008; Pike, Poulsen, & Woo, 2017; Portaccio et al., 2009; Vega-Fernandez et al., 2015). In general, evaluation of screening measure utility was limited by the small number of studies retrieved alongside the varied screening measures, wide range of specific populations examined, or minimal diagnostic accuracy data reported. Only two studies provided AUC data (Pike et al., 2017; Vega-Fernandez et al., 2015). Based on these data, the MoCA demonstrated "good" classification accuracy (AUC = .84) for detecting cognitive impairment when used in youth with congenital heart disease (Pike et al., 2017). In contrast, the pediatric ANAM (pedANAM) demonstrated "good" to "poor/very poor" accuracy (AUC = .60–.89) in youth with systemic erythematosus lupus depending on the composite scores used (Vega-Fernandez et al., 2015). Only one study, which examined the Neurocog FX in youth with epilepsy, reported positive and negative likelihood ratios (Hoppe et al., 2009). Of all the Neurocog FX tasks, only one (Figural Memory) provided a positive likelihood ratio above 10 when compared to the reference standard (i.e., scores lower than 2 standard deviations below mean). However, while specificity was high (i.e., range: .57–.99), the sensitivity of this measure for the purposes of cognitive screening was lower (i.e., range: .14–.76).

Correlational data

The remaining two-thirds of studies ($k = 14$) were not designed as primary classification accuracy studies, as they presented correlation data between index measures and reference standards only as part of broader investigations of cognitive screening tool utility. These data were reported in various ways, as presented in Table 5. Most of these studies ($k = 10$, 71%) used computerized measures (i.e., ImPACT, ANAM, NIH Toolbox, CNS-VS, CogState, and CRI) and recruited healthy people ($k = 8$) or those who had sustained concussion ($k = 2$). As shown in Table 5, the strongest correlations were between LANSE-C/A subtests (excluding the visuospatial domain) and traditional neuropsychological reference standards, with 11/21 significant large correlations (i.e., rs = .6–.9), followed by RBANS total scores and WAIS/WISC Intelligence Quotients (rs = .61–.57). Most correlations between studies comparing computerized index tests and neuropsychological reference standards with available correlation coefficient data in youth specifically ($k = 7$) were small-to-moderate (i.e., rs = .20–.49; see Table 5). Of these studies, only 11.9% (i.e., 19/160) of significant correlations between specific variations of computerized screening indices and individual neuropsychological tests or composites were large (i.e., $r \geq .50$). Most of these large correlations (i.e., 13/19, 68%) were between processing speed and reaction time indices and related reference tasks [i.e., trail making test (TMT) and digit symbol substitution test (DSST)].

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 3. Overview of screening measures and reference standards for included studies

First author (year)	Target population	Index test(s), version (if reported)	Reference standard
Allen 2011	Undergraduate university students	ImPACT, v3.0	NFL Neuropsychological battery: HVLT-R; BVMT-R; TMT; COWAT; and WAIS-III Symbol Search, Digit Symbol-Coding and Digit Span.
Bleiberg 2000	High school/college students and student athletes	ANAM, v1.0 (tasks: SRT, STN, MTH, SPT, and MSP)	Traditional neuropsychological tests: WAIS-R Block Design; TMT; Finger Tapping Test; Auditory Consonant Trigrams Test; PASAT; HVLT-R; COWAT; and Stroop Color-Word Test.
Boulet-Craig, 2018	Oncology, Acute Lymphoblastic Leukemia	DIVERGT	Traditional neuropsychological tests: WAIS-IV Vocabulary, Block Design, Matrix Reasoning, Coding and Symbol Search; WIAT-III Math Problem Solving and Math Fluency; WMS-IV Spatial Addition; and CVLT-II.
Gualtieri, 2006	ADHD	CNS-VS	Traditional neuropsychological tests: RAVLT; WMS-III Logical Memory and Facial Recognition; Finger Tapping Test; Stroop Color-Word Test; TMT (Part B only) and Verbal Fluency Test; as well as NES2 computerized tests (i.e., Finger Tapping, Switching Attention and CPT)
Heaton, 2014	Community population	NIHTB-CB	Traditional neuropsychological tests: WRAT-4; PPVT-IV; WAIS-IV Coding and Symbol Search; D-KEFS Color-Word Interference Test; WCST; BVMT-R; RAVLT, and WAIS-IV Letter Number Sequencing. Tests were used to create total, crystallized and fluid ability composite scores.
Holzer, 2007	Mental illness (psychosis and other disorders)	RBANS	WISC-III or WAIS-III (Full Scale, Verbal and Perceptual IQ; dependant on participant age).
Hoppe, 2009	Epilepsy	NeuroCog FX, v2.0	German equivalents of traditional neuropsychological tests: TMT; WMS-III Digit Span and Spatial Span; RAVLT and Verbal Fluency; as well as Test for Cerebral Insufficiency, Chapuis Maze Test and DCS (i.e., visual learning and memory test).
Iverson, 2005	Sport-related concussion	ImPACT, v2.0	SDMT
Jones, 2008	Undergraduate university students	ANAM, version 1.0 (tasks: 2CH, CDS, LGR, MSP, MTH, and STN)	WJ-III: General Intellectual Ability Index and Fluid Reasoning, Comprehension-Knowledge and Cognitive Efficiency clusters.
Kahn, 2015	TBI	LANSE-A	WeeFIM Total Cognition
Kelly, 2012	Combat-related concussion	ANAM-IV-TBI (tasks: SRT, PRT, CDS, CDD, MTH, and MSP)	Traditional neuropsychological tests: HVLT-R, SDMT, Stroop, TMT, COWAT.
Krull, 2008	Oncology, mixed	DIVERGT	Traditional neuropsychological tests: WASI/WISC-III subtests (e.g., vocabulary, verbal reasoning, visuospatial construction, processing speed, and working memory) and WRAT-III subtests (e.g., word reading and mathematical calculation).
Maerlender, 2010	College athletes	ImPACT, v2.0	Traditional neuropsychological tests: CVLT; BVMT-R; CPT; D-KEFS Trail Making, Verbal Fluency and Color Word Interference; Grooved Pegboard and PASAT; and experimental tests (i.e., Verbal Continuous Memory Task and N-back).
McLeod, 2006	Community athletes	SAC	Traditional neuropsychological tests: Buschke Selective Reminding Test; TMT (B only); and WISC-III Symbol Search and Coding.
Pike, 2017	Congenital heart disease	MoCA	WRAML2: General Memory Index

(*Continued*)

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 3. (*Continued*)

First author (year)	Target population	Index test(s), version (if reported)	Reference standard
Portaccio, 2009	Multiple sclerosis	BNBC	Traditional neuropsychological tests: WISC-R; Selective Reminding Test (from BRB); Spatial Recall Test (from BRB); SDMT; TMT; MCST; Verbal Fluency; Oral Denomination Test (from the Aachener Aphasie Test); Token Test; Indication of Pictures (from Neuropsychological Examination for Aphasia); and Phrase Comprehension test (from Analysis of Aphasic Deficits).
Raiker, 2015	TBI and brain tumor	LANSE-C/A	Traditional neuropsychological tests: WISC-IV; CVLT-C, and CMS.
Schatz, 2010	Undergraduate university students	ImPACT, CRI, and CogSport (versions NR)	Traditional and computerized neuropsychological tests: WAIS-R: Digit Symbol, TMT, and d2 test of attention.
Solomon, 2014	Sport-related concussion history	ImPACT, version NR	Wonderlic Classic Cognitive Ability Test
Vega-Fernandez, 2015	Systemic lupus erythematosus	PedANAM, version NR (tasks: CDS, CDD, SPT, MTH, MSP, STN, LGR, MGT, and CPT)	Neuropsychological assessment measures. Details were not described. Tasks used clustered into four domains: working memory, processing speed, visuospatial construction ability, and attention/executive function.

Note: 2CH = 2-Choice Reaction Time; ANAM = Automated Neuropsychological Assessment Metrics; BNBC = Brief Neuropsychological Battery for Children; BRB = Rao's Brief Repeatable Battery; BVMT-R = Brief Visuospatial Memory Test-Revised; CDD = Code Substitution Delay; CDS = Code Substitution; CMS = Children's Memory Scale; CNS-VS = Central Nervous System – Vital Signs; COWAT = Controlled Oral Word Association Test; RAVLT = Rey Auditory Verbal Learning Test; CPT = Continuous Performance Test; CRI = HeadMinder's Concussion Resolution Index; CVLT = California Verbal Learning Test; CVLT-II = California Verbal Learning Test-Second Edition; CVLT-C = CVLT = California Verbal Learning Test for Children; D-KEFS = Delis-Kaplan Executive Function System; DCS = Diagnosticum für Cerebralschädigung – revidiert; HVLT-R = Hopkins Verbal Learning Test-Revised; ImPACT = Immediate Post-Concussion Assessment and Cognitive Testing; LANSE-C/A = Lebby-Asbell Neurocognitive Screening Examination-Child/Adolescents; LGR = Logical Reasoning; MCST = Modified Card Sorting Test; MGT = Matching Grid Test; MoCA = Montreal Cognitive Assessment; MSP = Matching to Sample; MTH = Math Processing; NeuroCog FX = Neurocognitive Effects; NIHTB-CB = National Institute of Health Toolbox-Cognition Battery; NR = Not reported; PASAT = Paced Auditory Serial Addition Test; PedAMAM = Pediatric Automated Neuropsychological Assessment Metrics; PPVT-4 = Picture Peabody Vocabulary Test-Fourth Edition; PRT = Procedural Reaction Time; RBANS = Repeated Battery for the Assessment of Neuropsychological Status; SAC = Standardized Assessment of Concussion; SDMT = Symbol Digit Modalities Test; SPT = Spatial Processing; SRT = Simple Reaction Time; STN = Sternberg Memory; TMT = Trial Making Test; WAIS-R = Wechsler Adult Intelligence Scale-Revised; WAIS-III = Wechsler Adult Intelligence Scale-Third Edition; WAIS-IV = Wechsler Adult Intelligence Scale-Fourth Edition; WASI = Wechsler Abbreviated Scale of Intelligence; WCST = Wisconsin Card Sorting Test; WeeFIM = Functional Independence Measure; WISC-R = Wechsler Intelligence Scale for Children-Revised; WISC-III = Wechsler Intelligence Scale for Children-Third Edition; WISC-IV = Wechsler Intelligence Scale for Children-Fourth Edition; WJ-III = Woodcock–Johnson Tests of Cognitive Ability; WMS-III = Wechsler Memory Scale-Third Edition; WRAML2 = Wide Range Assessment of Memory and Learning Second Edition; WRAT(III/IV) = Wide Range Achievement Test (Third/Fourth Edition).

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 4. Psychometric data retrieved from primary classification accuracy studies

First author (year)	Comparison parameters for impairment	Sensitivity	Specificity	AUC	PPV	NPV	LR+	LR−
Boulet-Craig, 2018	Impaired DIVERGT performance (i.e., either one scaled score of 4 or less or two scaled scores of 6 or less) against GAI impairment	.80	.70	NR	.24	.97	NR	NR
	Impaired DIVERGT performance against mathematical reasoning impairment	.58	.69	NR	.24	.91	NR	NR
	Impaired DIVERGT performance against verbal memory impairment	.75	.69	NR	.18	.97	NR	NR
Hoppe, 2009	NeuroCogFX verbal memory <80 (-2 *SD*) against Verbal Learning scores	.49	.89	NR	.58	.84	4.3	.58
	NeuroCogFX verbal memory <85 (-1.5 *SD*) against Verbal Learning scores	.65	.74	NR	.44	.87	2.49	.48
	NeuroCogFX verbal memory <90 (-1 *SD*) against Verbal Learning scores	.76	.63	NR	.4	.89	2.07	.38
	NeuroCogFX verbal memory <80 (-2 *SD*) against Verbal Final scores	.39	.88	NR	.59	.75	3.1	.7
	NeuroCogFX verbal memory <85 (-1.5 *SD*) against Verbal Final scores	.55	.74	NR	.5	.78	2.12	.61
	NeuroCogFX verbal memory <90 (-1 *SD*) against Verbal Final scores	.69	.64	NR	.48	.82	1.95	.48
	NeuroCogFX verbal memory <80 (-2 *SD*) against Verbal Retention scores	.39	.84	NR	.34	.86	2.4	.73
	NeuroCogFX verbal memory <85 (-1.5 *SD*) against Verbal Retention scores	.54	.69	NR	.27	.87	1.71	.68
	NeuroCogFX verbal memory <90 (-1 *SD*) against Verbal Retention scores	.64	.57	NR	.25	.88	1.5	.63
	NeuroCogFX verbal memory <80 (-2 *SD*) against Verbal Recognition scores	.45	.89	NR	.63	.8	4.21	.61
	NeuroCogFX verbal memory <85 (-1.5 *SD*) against Verbal Recognition scores	.59	.75	NR	.48	.82	2.34	.55
	NeuroCogFX verbal memory <90 (-1 *SD*) against Verbal Recognition scores	.73	.64	NR	.44	.86	2.02	.43
	NeuroCogFX figural memory <80 (-2 *SD*) against Figural Learning scores	.14	.99	NR	.88	.68	12.71	.87
	NeuroCogFX figural memory <85 (-1.5 *SD*) against Figural Learning scores	.59	.89	NR	.74	.8	5.27	.46
	NeuroCogFX figural memory <90 (-1 *SD*) against Figural Learning scores	.71	.87	NR	.74	.85	5.3	.33
	NeuroCogFX figural memory <80 (-2 *SD*) against Figural Final scores	.12	.98	NR	.75	.66	5.2	.9
	NeuroCogFX figural memory <85 (-1.5 *SD*) against Figural Final scores	.55	.87	NR	.71	.77	4.26	.52
	NeuroCogFX figural memory <90 (-1 *SD*) against Figural Final scores	.67	.85	NR	.72	.82	4.4	.39
	NeuroCogFX figural memory <80 (-2 *SD*) against Nonmemory test scores	.17	.95	NR	.75	.56	3.3	.88
	NeuroCogFX figural memory <85 (-1.5 *SD*) against Nonmemory test scores	.31	.91	NR	.76	.59	3.45	.76
	NeuroCogFX figural memory <90 (-1 *SD*) against Nonmemory test scores	.54	.87	NR	.79	.67	4.17	.53
	NeuroCogFX figural memory <95 (-0.5 *SD*) against Nonmemory test scores	.69	.72	NR	.69	.72	2.45	.43
Krull, 2008	Impaired DIVERGT performance (i.e., either one scaled score of 4 or less or two scaled scores of 6 or less) against global IQ impairment	.94	.63	NR	NR	NR	NR	NR
Pike, 2017	MoCA cut-off of 26 against GMI impairment (i.e., <85) in congenital heart disease	.96 (95% CI [.85, .99])	.67 (95% CI [.19, .79])	.84	NR	NR	NR	NR
	MoCA cut-off of 26 against GMI impairment (i.e., <85) in healthy control	.75 (95% CI [.19, .99])	.9 (95% CI [.7, 1.0])	NR	NR	NR	NR	NR
Portaccio, 2009	One or more tests failed (defined as <5th percentile); BNBC tests included in criterion	.96	.81	NR	NR	NR	NR	NR
	One or more tests failed (defined as <5th percentile); BNBC tests excluded from criterion	.96	.76	NR	NR	NR	NR	NR
	Two or more tests failed (defined as <5th percentile); BNBC tests included in criterion	.64	.94	NR	NR	NR	NR	NR
	Two or more tests failed (defined as <5th percentile); BNBC tests excluded from criterion	.65	.92	NR	NR	NR	NR	NR
Vega-Fernandez, 2015	PedANAM-CPS-uwa against cognitive impairment (defined as either 1 domain z-score <-2 *SD* or two domain z-score <-1 *SD*) using a cut-off of 0.25 (validation sample).	.8	.48	.78	NR	NR	NR	NR

(*Continued*)

Table 4. (*Continued*)

First author (year)	Comparison parameters for impairment	Sensitivity	Specificity	AUC	PPV	NPV	LR+	LR−
	PedANAM-CPS-pca against cognitive impairment using a cut-off of 0.25 (validation sample).	.9	.42	.8	NR	NR	NR	NR
	PedANAM-CPS-logit against cognitive impairment using a cut-off of -2.70 (validation sample).	.9	.32	.8	NR	NR	NR	NR
	PedANAM-CPS-multiscore against cognitive impairment using a cut-off of 0.09 (validation sample).	.89	.55	.89	NR	NR	NR	NR
	PedANAM-CPS-uwa against cognitive impairment using a cut-off of 0.25 (development sample).	.83	.37	.6	NR	NR	NR	NR
	PedANAM-CPS-pca against cognitive impairment using a cut-off of 0.25 (development sample).	.83	.42	.6	NR	NR	NR	NR
	PedANAM-CPS-logit against cognitive impairment using a cut-off of -2.70 (development sample).	.92	.31	.77	NR	NR	NR	NR
	PedANAM-CPS-multiscore against cognitive impairment using a cut-off of 0.09 (development sample).	.92	.66	.85	NR	NR	NR	NR

Note: AUC = Area Under the Curve; BNBC = Brief Neuropsychological Battery for Children; CI = Confidence interval; CPS = Cognitive Performance Score; GMI = General Memory Index; IQ = Intellectual Quotient; LR− = Negative Likelihood Ratio; LR+ = Positive Likelihood Ratio; MoCA = Montreal Cognitive Assessment; NeuroCog FX = Neurocognitive Effects; NPC = Negative Predictive Value; NR = Not Reported; PPV = Positive Predictive Value; SD = Standard Deviation.

Quality

Ratings of risk of bias and quality of study design are presented in Table 6. The majority of studies were deemed of high or unclear risk of bias in terms of patient selection (QUADAS-2 Domain 1; 70%) due to convenience or unclear sampling methods. Further, most studies were judged to be high or unclear risk of bias for the index test (Domain 2; 85%) and reference standard (Domain 3; 90%). Commonly, the extent to which cognitive screening and reference standard neuropsychological assessment were completed and scored independently (i.e., blind rating) was not documented. In addition, roughly half of the studies were rated as high or unclear risk of bias for participant flow and timing (Domain 4; 50%), with participants either not all receiving the reference standard, not all receiving the same reference standard or not all being included in the analysis. Study status as a primary classification accuracy report did not consistently result in lowered risk of bias ratings across the four QUADAS-2 domains, highlighting poor adherence to STARD criteria across most of the included studies.

DISCUSSION

This review had four primary aims. These were to (1) describe the cognitive screening measures used for people aged 12–25, (2) describe the settings in which the screening measures were used, (3) review the accuracy of these index measures in detecting cognitive impairment relative to an objective reference standard (e.g., neuropsychological assessment), and (4) evaluate the risk of bias of the included studies. To our knowledge, this is the first systematic review of cognitive screening in youth aged between 12 and 25 years.

Overview of Screening Measures Used in Youth

This review highlighted, for the first time systematically, that there exists a diverse range of cognitive screening tools that have been utilized in people aged 12–25, across many clinical and non-clinical populations. These groups included healthy individuals, people with concussion or mild TBI, neurological conditions, oncological diseases, autoimmune disorders, congenital heart disease, mental illness, and neurodevelopmental conditions. Screening approaches differed widely according to assessment length (i.e., 5–30 min), format (i.e., pen-and-paper or computerized), and the range of cognitive domains evaluated. While the types of screening tools were variable, the most commonly used screening tools retrieved in this review were computerized, particularly the ImPACT and ANAM. These measures were used in 45% of the included studies and were mostly administered in youth who were healthy or those who had sustained concussion.

Classification Accuracy

This review found that only six of the 20 studies retrieved (30%) were designed as classification accuracy studies

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 5. Summary of studies examining correlations between cognitive screening tools and comprehensive cognitive assessment

First author, year	Comparisons	Summary of main findings
Allen, 2011	Pearson correlations between ImPACT domains and traditional neuropsychological tests.	ImPACT *Verbal Memory* domain was significantly correlated with traditional neuropsychological tasks measuring verbal learning and memory (*r*s = .27–.31) as well as those measuring visual learning memory and executive (*r*s = .31–.43). ImPACT *Visual Memory* domain was significantly correlated with traditional neuropsychological tasks measuring visual learning and memory (*r*s = .35–.38) as well as those measuring attention/working memory, verbal memor and executive function (*r*s = .20–.29). ImPACT *Visual Motor* domain was significantly correlated with traditional neuropsychological tasks measuring processing speed (*r*s = .39–.43) as well as those measuring verbal learning and memory, visual learning, attention/working memor and executive function (*r*s = .26–.38). ImPACT *Reaction Time* domain was only significantly correlated with traditional neuropsychological tasks measuring processing speed (*r*s = .27–.28). ImPACT *Impulse Control* was not significantly correlated with any traditional neuropsychological tasks. TMT Part A and COWAT were not related to any ImPACT domain.
Bleiberg, 2000	Pearson correlations between accuracy, throughput or reaction time on ANAM subscales and traditional neuropsychological tests. Stepwise regression using ANAM data to predict traditional neuropsychological tests.	All ANAM subtests including *MTH* (accuracy, throughout and reaction time), *SPT* (throughput and reaction time), *STN* (throughput and reaction time), and *MSP* (throughput and reaction time) were significantly correlated with traditional measures of executive functioning, including TMT (Part B only; *r*s = .18–.41) and Stroop CWI (*r*s = .22–.54), Auditory Consonant Trigrams total (*r* = .19–.36), PASAT total (*r*s = .21–.66) and HVLT T3 total (*r*s = .20–.34). *MTH* throughput was most strongly and consistently correlated with traditional neuropsychological measures. In addition, a series of stepwise regression analyses revealed that many ANAM subtests (particularly *MTH*) accounted for significant amounts of variance in these measures. Data was used in support of concurrent validity.
Gualtieri, 2006	Correlations between CNS-VS subtests and traditional neuropsychological tests.	Moderate correlations were reported between CNS-VS domains and traditional neuropsychological tasks measuring memory, perceptual motor speed, and executive function as well as tasks measuring psychomotor speed and executive function on the NES2. Correlations between CNS-VS *CPT* and NES2 were low. No correlation coefficients were reported. Data was used in support of concurrent validity.
Heaton, 2014	Correlations between NIHTB-CB and "gold-standard" measures of total, crystallized and fluid cognition composites.	Significant correlations were observed between NIH Toolbox and gold-standard crystallized cognition composites (*r* = .90), NIH Toolbox and gold-standard fluid cognition composites (*r* = .78) and NIH Toolbox and gold-standard total cognition composites (*r* = .89) in the total sample. NIH Toolbox and gold-standard crystallized cognition composites were positively correlated with age for the total sample (*r*s = .18 and .14, respectively). Fluid cognition composites were negatively correlated with age for the total sample (*r*s = -.68 and -.55). Graphical data showed overlapping age trajectories for NIH Toolbox and gold-standard crystallized and fluid cognition composites in each five-year age band of the total sample (i.e., 20–85), including age 20–24. No correlation data was presented for the 20–24 age group only.
Holzer, 2007	Correlations between RBANS total/subtests and WAIS/WISC-III IQ scales.	RBANS total score was significantly correlated with WISC/WAIS-III Full Scale IQ (*r* = .61), Verbal IQ (*r* = .57) and Performance IQ (*r* = .60). RBANS index scores including *Immediate Memory, Delayed Memory, Visuospatial, and Language* were significantly correlated with Full Scale, Verbal and Performance IQ score (*r*s: .30–.51). No correlational data between RBANS *Attention* and IQ was presented.
Iverson, 2005	Pearson correlations between ImPACT domains and SDMT completion time	All ImPACT domains examining, including *Verbal Memory*, *Visual Memory*, *Processing Speed*, and *Reaction Time* were significantly correlated with SDMT completion time (*r* = .37–.70).

(*Continued*)

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 5. (*Continued*)

First author, year	Comparisons	Summary of main findings
Jones, 2008	Pearson correlations between performance accuracy and throughput (i.e., efficiency, calculated as or accuracy/speed) on ANAM subscales and WJ-III cluster scales. Stepwise regression using ANAM data to predict WJ-III cluster scales and individual subtests.	ANAM subtests including *2CH* (throughput), *CDS* (throughput), *LGR* (accuracy and throughput), *MSP* (accuracy), and *STM* (accuracy) were significantly correlated with WJ-III General Intellectual Ability (*r*s = .24–.55). ANAM subtests including *CDS* (accuracy and throughput), *CDD* (accuracy and throughput), *LGR* (accuracy), *MSP* (accuracy) and *STN* (accuracy) were significantly correlated with WJ-III Fluid Reasoning (*r*s = .24–.42). ANAM subtests including *CDD* (accuracy), *LGR* (accuracy), *MSP* (accuracy), and *STN* (accuracy) were significantly correlated with WJ-III Comprehension-Knowledge (*r*s = .24–.36). ANAM subtests including *2CH* (throughput), *CDS* (throughput), *LGR* (throughput), *MSP* (accuracy and throughput), *MTH* (accuracy and throughput), and *STN* (accuracy and throughput) were significantly correlated with WJ-III Fluid Reasoning (*r*s = .27–.52). Various significant correlations were also revealed between ANAM subtests (particularly throughput) and individual WJ-III subtests. In addition, a series of stepwise regression analyses revealed that many ANAM subtests accounted for significant amounts of variance in WJ-III cluster scales and individual subtests. Data was used in support of concurrent validity.
Kahn, 2015	Multiple regression using LANSE-A subtest data to predict WeeFIM Total Cognition at discharge	After controlling for premorbid complicating factors, parental education, time to follow commands and baseline WeeFIM Total Cognition, LANSE-A subtests, *Verbal Association* (B = 0.99) and *Visual Memory* (B = 0.47), uniquely predicted WeeFIM *Total Cognition* at discharge, ps = 0.02 and 0.05, respectively. Other subtests not entered into the model.
Kelly, 2012	Correlations between ANAM subtests and traditional neuropsychological tests.	Correlation between ANAM subtests and individual or combination of traditional neuropsychological test was poor. No correlation coefficients were reported.
Maelender, 2010	Pearson correlations between ImPACT domains and traditional and experimental neuropsychological tests.	ImPACT *Verbal Memory* domain was significantly correlated with traditional neuropsychological tasks measuring verbal memory (r = .40) and visual memory (r = .44). ImPACT *Visual Memory* domain was correlated with traditional neuropsychological tasks measuring verbal memory (r = .59). ImPACT *Visual Motor* domain was significantly correlated with traditional neuropsychological tasks measuring working memory (r = .39), processing speed (r = .41), and reaction time (r = .34). ImPACT *Reaction Time* domain was significantly correlated with traditional neuropsychological tasks measuring working memory (r = −.31), processing speed (r = −.37), and reaction time (r = −.39). ImPACT *Impulse Control* domain was not significantly correlated with any traditional neuropsychological tasks. Various significant correlations were also revealed between individual ImPACT domains and experimental cognitive measures purporting to tap similar domains. Data was used in support of concurrent validity.
McLeod, 2006	Pearson correlations between SAC total score and traditional neuropsychological tests.	SAC total score was significantly correlated with WISC-III Symbol Search (r = .32), TMT Part B (r = .29) and SRT Sum Recall (r = .28) and Delayed Recall (r = .36), yet not with WISC-III Coding (r = .27) or SRT Continuous Long-Term Recall (r = .22). Data was acknowledged as indicating a weak relationship between SAC and traditional neuropsychological tests.
Raiker, 2015	Zero-order correlations between LANSE-C/A subtests and traditional neuropsychological tasks.	Most LANSE attention and working subtests were significantly correlated with WISC-IV Digit Span and Letter Number Sequencing (all r > .5). The LANSE executive function task *Verbal Associations*, yet not *Judgment*, was significantly correlated with WISC-IV Similarities and Comprehension (*r*s = .7–.9). LANSE *Verbal/Auditory Memory* was significantly correlated with most CVLT-C recall (*r*s = .6–.7) tasks, while *Visual Memory* was significantly correlated with CMS Faces Delay (r = .9) yet not CMS Dot Locations Delay. LANSE Visual Motor, yet not *Visual Spatial*, was significantly correlated with WISC-IV Block Design or Matrix Reasoning. Data was used in support of concurrent validity.

(*Continued*)

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Table 5. (*Continued*)

First author, year	Comparisons	Summary of main findings
Schatz, 2006	Pearson correlations between ImPACT, HeadMinder's CRI, and CogSport domains and traditional neuropsychological tests (i.e., TMT, WAIS-R Digit Symbol).	CRI *Simple Reaction Time* was correlated with TMT Part A and Digit Symbol (*r*s = .43–.53), yet not TMT Part B. Cogstate *Simple Reaction Time* was not correlated with any measure. ImPACT *Reaction Time* domain was significantly correlated with TMT (Part A and B) and Digit Symbol (*r*s = .44–.64). CogState *Complex Reaction Time* was correlated with TMT (Part A and B: *r*s = .54) yet not Digit Symbol. CRI *Complex Reaction Time* was not correlated with any measure. ImPACT *Processing Speed* was correlated with TMT Part B and Digit Symbol (*r*s = .51–.54). CRI *Processing Speed* was correlated with TMT Part B and Digit Symbol (*r*s = .60–.61). Three screening measures (ImPACT, CRI, and CogSport) share common variance on processing speed and reaction time domains.
Solomon, 2014	Correlations between ImPACT domains and Wonderlic scores.	ImPACT *Verbal Memory*, *Visual Memory*, *Visual Motor Speed*, and *Reaction Time* were significantly correlated with Wonderlic total scores (*r*s = .23–.49).

Note: Cognitive screening subtests have been italicized. Correlations in the expected direction (e.g., TMT) are written as positive. 2CH = 2-Choice Reaction Time; ANAM = Automated Neuropsychological Assessment Metrics; BNBC = Brief Neuropsychological Battery for Children; CDD = Code Substitution Delay; CDS = Code Substitution; CMS = Children's Memory Scale; CNS-VS = Central Nervous System – Vital Signs; COWAT = Controlled Oral Word Association Test; CPT = Continuous Performance Test; CRI = HeadMinder's Concussion Resolution Index; CVLT-C = CVLT = California Verbal Learning Test for Children; HVLT-R = Hopkins Verbal Learning Test-Revised; ImPACT = Immediate Post-Concussion Assessment and Cognitive Testing; IQ = Intellectual Quotient; LANSE-C/A = Lebby-Asbell Neurocognitive Screening Examination-Child/Adolescents; LGR = Logical Reasoning; MSP = Matching to Sample; MTH = Math Processing; NeuroCog FX = Neurocognitive Effects; NIHTB-CB = National Institute of Health Toolbox-Cognition Battery; NR = Not reported; PedAMAM = Pediatric Automated Neuropsychological Assessment Metrics; PASAT = Paced Auditory Serial Addition Test; RBANS = Repeated Battery for the Assessment of Neuropsychological Status; SAC = Standardized Assessment of Concussion; SDMT = Symbol Digit Modalities Test; SPT = Spatial Processing; SRT = Simple Reaction Time; STN = Sternberg Memory; TMT = Trial Making Test; WAIS-R = Wechsler Adult Intelligence Scale-Revised; WeeFIM = Functional Independence Measure; WISC-R = Wechsler Intelligence Scale for Children-Revised; WISC-III = Wechsler Intelligence Scale for Children-Third Edition; WISC-IV = Wechsler Intelligence Scale for Children-Fourth Edition; WJ-III = Woodcock–Johnson Tests of Cognitive Ability.

and provided explicit classification accuracy data. Of these studies, few reported sufficient psychometric data to enable a full appraisal of the evaluated screening tools. Some screening tools included in this review, such as the DIVERGT (oncology), MoCA (congenital heart disease), BNBC (multiple sclerosis), and pedANAM (systemic lupus erythematosus) yielded encouraging sensitivity and specificity ranges (i.e., greater than .80 and .60, respectively) in specific groups of people aged 12–25 (Boulet-Craig et al., 2018; Krull et al., 2008; Pike et al., 2017; Portaccio et al., 2009; Vega-Fernandez et al., 2015). The MoCA may have pragmatic utility in youth in primary or tertiary care when compared to other measures given its brevity (i.e., less than 10–12 min to administer), wide availability, ease of administration (i.e., pen-and-paper), and quick interpretation (i.e., single cut-off score). Nevertheless, Pike et al. (2017) used memory impairment as the reference standard rather than global cognitive functioning in their evaluation of the MoCA in congenital heart disease. Therefore, the extent to which this screening tool may accurately classify general cognitive impairment in youth is unclear. Only one tool, NeuroCogFX (epilepsy) presented a range of psychometric parameters (i.e., sensitivity, specificity, predictive values, and likelihood ratios); however, the balance of sensitivity (i.e., .12-.76) and specificity (i.e., .57-.99) fluctuated significantly in range for each individual screening task, which may question its clinical usefulness (Hoppe et al., 2009). Taken together, while some screening measures demonstrate some promise, there is generally very limited data on classification accuracy of cognitive screening in people aged 12–25. Based on the retrieved articles, there are no instances where test accuracy has been replicated in the target clinical populations. Screening tools could therefore not be compared within or between different youth populations. Independent replication of cognitive screening tests is essential to establishing the validity of cognitive screening tools in youth, or any age-group (Straus, Richardson, Glasziou, & Haynes, 2010).

The majority of studies used correlational data to support the utility of cognitive screening in youth, which provides only *indirect* evidence of validity. Correlations were primarily reported from studies evaluating computerized tools. Most of the significant correlations between index tests and reference standards were small-to-medium; however, with less than 12% of correlation coefficients in studies with available data deemed to be large (i.e., >.50). The strongest relationships appeared to be between computerized processing speed and reaction time indices and speed-based reference standards (e.g., TMT and DSST). This pattern of results may suggest that some specific computerized screening tasks included in this review may be better suited to measuring lower-level abilities including attention and processing speed rather than a broader range complex cognitive domains. Nevertheless, the assessment of whether computerized assessment tasks most closely overlap with established neuropsychological measures of processing speed requires further research.

Table 6. QUADAS-2 ratings for studies included in the review

	Domain 1: Patient Selection		Domain 2: Index Test(s)		Domain 3: Reference Standard		Domain 4: Flow and Timing
First author (year)	Risk of Bias	Applicability	Risk of Bias	Applicability	Risk of Bias	Applicability	Risk of Bias
Allen 2011	HR	LR	UR	LR	UR	LR	LR
Bleiberg 2000	UR	LR	UR	LR	UR	LR	LR
Boulet-Craig, 2018	HR	LR	UR	LR	UR	LR	HR
Gualtieri, 2006	UR	LR	UR	LR	UR	LR	UR
Heaton, 2014	HR	LR	UR	LR	UR	LR	LR
Holzer, 2007	HR	UR	UR	LR	UR	LR	HR
Hoppe, 2009	LR	LR	UR	LR	UR	LR	HR
Iverson, 2005	UR	LR	UR	LR	UR	LR	UR
Jones, 2008	HR	LR	UR	LR	UR	LR	UR
Kahn, 2015	LR	UR	LR	UR	LR	LR	LR
Kelly, 2012	LR	LR	UR	LR	UR	LR	LR
Krull, 2008	LR	LR	LR	LR	LR	LR	HR
Maerlender, 2010	UR	LR	UR	LR	UR	LR	UR
McLeod, 2006	HR	LR	LR	LR	HR	LR	LR
Pike, 2017	LR	LR	UR	LR	UR	UR	LR
Portaccio, 2009	LR	LR	HR	LR	HR	LR	LR
Raiker, 2015	HR	UR	UR	LR	UR	LR	HR
Schatz, 2010	HR	LR	UR	LR	UR	LR	LR
Solomon, 2014	HR	LR	UR	LR	UR	LR	LR
Vega-Fernandez, 2015	UR	LR	UR	LR	UR	LR	HR

Note: HR = High Risk; LR = Low Risk; QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies – Version 2; UR = Unclear Risk.

Risk of Bias

Most studies were deemed to be at high or unclear risk of bias using the QUADAS-2. This major and novel finding in this review highlights the significant methodological limitations of the published studies, to date. The most problematic domains were related to patient selection, use of index tests and use of reference standards. More specifically, ratings of unclear or higher risk of bias were primarily due to convenience or unclear sampling methods as well as uncertainty regarding whether index tests or reference standards were completed and scored independently. A lack of blinding is well known to inflate correct classification rates (Blacker et al., 1994; Schmidt & Factor, 2013; Straus et al., 2010). Sources of bias related to sampling and assessment blinding were present in studies that were and were not designed as primary classification accuracy studies. More stringent adherence to STARD criteria in future studies would therefore appear to be essential to reduce the risk of bias when evaluating cognitive screening in youth (Bossuyt et al., 2003; 2015; Cohen et al., 2016). The potential for publication bias must also be considered. While unable to be formally assessed in this type of review, it is possible that authors may have been less likely to publish negative screening papers.

Implications

Based on our search strategy and inclusion criteria, studies examining cognitive screening in people aged 12–25 were limited in number and quality. Given the small number of studies, diverse range of cognitive domains and population groups assessed, and an elevated risk of bias among included studies, not even one screening measure can be currently recommended for use in people aged 12–25. Future studies must focus on evaluating cognitive screening tools in youth in accordance with STARD criteria (Bossuyt et al., 2003, 2015; Cohen et al., 2016), including brief assessment measures with extensive normative data (e.g., NIH Toolkit).

The lack of consensus regarding cognitive screening measures in youth may contribute to the limited accessibility of neuropsychological assessment in primary care settings (Delagneau et al., 2020). A validated tool that could assist clinical teams with the process of identifying cognitive impairment and help guide referrals for comprehensive assessment may support more timely access to neuropsychological evaluation (Allott et al., 2019; Delagneau et al., 2020). This would have particular utility in busy clinical settings if various healthcare clinicians, including those without specialist training in cognitive assessment, could administer the tool and interpret it easily [e.g., general practitioners (GPs), mental health clinicians, nurses, social workers, etc.].

With consideration to the review findings, there is a significant need to establish validated cognitive screening measures for people aged 12–25. This may involve examining whether measures used in older adults are appropriate or need to be adapted, or whether new screening tools need to be developed based on the unique trajectories of cognitive

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

development in this younger age group (Bryce & Allott, 2019). Given the large number of screening measures already developed, the first step may be to explore the utility of existing tools used in adults as a "downward extension" (e.g., MoCA), with consideration to methodological limitations identified in past studies. It is plasuable, however, that measures developed for adults may only be appropriate for downward extension without significant adaption to people aged 16 and older, when the rate of development of some cognitve skills may start to flatten (Weintraub et al., 2013). Similar evaluations are needed to determine whether any screening tool is effective at detecting severe or milder impairment in youth, and whether new theoretically-driven test development relevant to younger adolescents may be required. Particular screening measures may be more appropriate for specific groups of youth (e.g., mental health, TBI, etc.). The needs of individual settings will also likely influence what cognitive screening tools and parameters are acceptable, such as what rates of false positives are acceptable if weighting sensitivity over specificity.

Limitations

Reviewing available data on cognitive screening in youth for the first time, with consideration of contemporary definitions, neurobiological development between ages 12 and 25 and service delivery, is a major strength of this review. However, the findings must be considered in light of some limitations. First, the search term "*screen**" was used as one of key criterion given the focus of this review. Consequently, studies that did not use this term may have been missed in the electronic search. Reference lists of included articles and reviews that focused on cognitive screening in youth were searched in an attempt to mitigate this possibility, yet inconsistent use of terminology is a limitation in this field (Roebuck-Spencer et al., 2017). We also acknowledge that, given our age range of interest, we did not include the search terms *pediatric/pediatric*, *child**, or similar. Inclusion of these search terms would have made our otherwise focused review untenable; however, their absence may have resulted in some missed articles. Second, some flexibility was incorporated into age-related inclusion criterion given the wide range of reporting standards. That is, studies that did not report the mean age of the sample could still be considered for inclusion if either (1) more than 50% of the sample were aged 12–25 or (2) the midpoint of an age bracket containing relevant data fell within 12–25 (e.g., subset of participants aged 18–26). Consequently, some of the included studies inevitably included some participants over 25, which may have influenced the generalizability of the presented diagnostic accuracy data. Thirdly, this systematic review was guided by screening definitions proposed by Roebuck-Spencer et al. (2017) presented by the National Academy of Neuropsychologists. We acknowledge, however, that aspects of these criteria (e.g., up to 30-min screening duration) may not be appropriate in vulnerable youth and may require further adaptation (Rossell, 2019).

CONCLUSIONS

Cognitive screening in people, aged 12–25, at risk of cognitive problems represents an important method of triage guiding comprehensive assessment. Screening also may provide a technique to document and monitor cognitive concerns. Nevertheless, despite widespread use in adults, few studies have examined the utility of cognitive screening measures in youth, defined in this review as 12–25 years. Studies that have evaluated the usefulness of cognitive screening in youth are limited and generally have not presented strong psychometric validity or are at high risk of bias with limited adherence to the STARD criteria. Studies investigating whether current screening tools used in adults, such as the MoCA, or another measure, are suitable for downward extension for subsets of youth (e.g., 16 or older), or need to be adapted across the 12–25 age bracket, would appear to be essential. Alternatively, novel measures that are uniquely sensitive to trajectories of cognitive development in groups of youth, such as those aged 12–14 or 15–17, may need to be developed. Critically, future studies must address the methodical limitations of past research.

ACKNOWLEDGMENTS

We thank the authors from the reviewed studies for their contribution to the field.

FINANCIAL SUPPORT

KA is supported by a Career Development Fellowship from the National Health and Medical Research Council (NHMRC; APP1141207).

CONFLICTS OF INTEREST

The authors have nothing to disclose.

REFERENCES

Allen, B. J., & Gfeller, J. D. (2011). The immediate post-concussion assessment and cognitive testing battery and traditional neuropsychological measures: A construct and concurrent validity study. *Brain Injury*, *25*(2), 179-191. doi: 10.3109/02699052.2010.541897

Allott, K., Proffitt, T.-M., McGorry, P. D., Pantelis, C., Wood, S. J., Cumner, M., & Brewer, W. J. (2013). Clinical neuropsychology within adolescent and young-adult psychiatry: Conceptualizing theory and practice. *Applied Neuropsychology: Child*, *2*(1), 47-63. doi: 10.1080/08841233.2012.670566

Allott, K., van-der-EL, K., Bryce, S., Hamilton, M., Adams, S., Burgat, L., & Rickwood, D. (2019). Need for clinical neuropsychological assessment in headspace youth mental health services: A national survey of providers. *Australian Journal of Psychology*, *71*(2), 108–116. doi:https://doi.org/10.1111/ajpy.12225

American Psychological Association Practice Organization. (2014). Distinguishing between screening and assessment for mental and

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

behavioral health problems: A statement from an American Psychological Association Practice Organization work group on screening and psychological assessment. Retrieved from: https://www.apaservices.org/practice/reimbursement/billing/assessment-screening

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., … Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment*, *9*, 449–460. doi: 10.2147/ndt.s39776

Arevalo-Rodriguez, I., Smailagic, N., Roquéi Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., … Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, *2015*(3). doi: 10.1002/14651858.CD010783.pub2

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

Arrieux, J., Cole, W., & Ahrens, A. (2017). A review of the validity of computerized neurocognitive assessment tools in mild traumatic brain injury assessment. *Concussion*, *30*(2), CNC31. doi: 10.2217/cnc-2016-0021

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (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. doi: 10.1093/arclin/acs027

Bava, S., Thayer, R., Jacobus, J., Ward, M., Jernigan, T., & Tapert, S. (2010). Longitudinal characterisation of white matter maturation during adolescence. *Brain Research*, *1327*, 38–46.

Blacker, D., Albert, M. S., Bassett, S. S., Rodney, C. P., Harrell, L. E., & Folstein, M. F. (1994). Reliability and Validity of Nincds-Adrda Criteria for Alzheimer's Disease: The National Institute of Mental Health Genetics Initiative. *Archives of Neurology*, *51*(12), 1198-1204. doi: 10.1001/archneur.1994.00540240042014

Bleiberg, J., Kane, R. L., Reeves, D. L., Garmoe, W. S., & Halpern, E. (2000). Factor analysis of computerized and traditional tests used in mild brain injury research. *Clinical Neuropsychologist*, *14*(3), 287-294. doi: 10.1076/1385-4046(200008)14:3;1-P;FT287

Block, C. K., Johnson-Greene, D., Pliskin, N., & Boake, C. (2017). Discriminating cognitive screening and cognitive testing from neuropsychological assessment: implications for professional practice. *Clinical Neuropsychologist*, *31*(3), 487-500. doi: 10.1080/13854046.2016.1267803

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., … Cohen, J. F. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ : British Medical Journal*, *351*, h5527. doi: 10.1136/bmj.h5527

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., … Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine*, *138*(1), W1–12.

Boulet-Craig, A., Robaey, P., Laniel, J., Bertout, L., Drouin, S., Krajinovic, M., … Lippé, S. (2018). DIVERGT screening procedure predicts general cognitive functioning in adult long-term survivors of pediatric acute lymphoblastic leukemia: A PETALE study. *Pediatric Blood and Cancer*, *65*(9). doi: 10.1002/pbc.27259

Bryce, S., & Allott, K. (2019). Cognitive screening: A significant unmet need in youth mental health. *Australian And New Zealand Journal Of Psychiatry*.

Bunnage, M. (2017). How do I know when a diagnostic test works? In S. Bowden (Ed.), *Neuropsychological assessment in the age of evidence-based practice: Diagnostic and treatment evaluation* (pp. 223–238). Oxford University Press.

Cernich, A., Reeves, D., Sun, W., & Bleiberg, J. (2007). Automated Neuropsychological Assessment Metrics sports medicine battery. *Archives of Clinical Neuropsychology*, *22*(Suppl. 1), 101-114. doi: 10.1016/j.acn.2006.10.008

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., … Bossuyt, P. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*, *6*(11), e012799. doi: 10.1136/bmjopen-2016-012799

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F., & Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery and Psychiatry*, *78*(8), 790-799. doi: 10.1136/jnnp.2006.095414

Davis, D. H. J., Creavin, S. T., Yip, J. L. Y., Noel-Storr, A. H., Brayne, C., & Cullum, S. (2015). Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database of Systematic Reviews*, *2015*(10). doi: 10.1002/14651858.CD010775.pub2

Delagneau, G., Bowden, S. C., Bryce, S., van-der-El, K., Hamilton, M., Adams, S., … Allott, K. (2020). Thematic analysis of youth mental health providers' perceptions of neuropsychological assessment services. *Early Intervention in Psychiatry*, *14*(2), 220–227. doi: 10.1111/eip.12876

Doyle, A. E., Vuijk, P. J., Doty, N. D., McGrath, L. M., Willoughby, B. L., O'Donnell, E. H., … Braaten, E. B. (2018). Cross-disorder cognitive impairments in youth referred for neuropsychiatric evaluation. *Journal of the International Neuropsychological Society*, *24*(1), 91–103. doi: 10.1017/S1355617717000601

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Foulkes, L., & Blakemore, S-J. (2018). Studying individual differences in human adolescence breain development. *Nature Neuroscience*, *21*, 315–323.

Gogtay, N., Giedd, J., Lusk, L., Hayashi, K., Greenstein, D., Vaituzis, C., … Thompson, P. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179.

Gonzalez-Heydrich, J., Hsin, O., Gumlak, S., Kimball, K., Rober, A., Azeem, M. W., … Biederman, J. (2014). Comparing stimulant effects in youth with ADHD symptoms and epilepsy. *Epilepsy and Behavior*, *36*, 102–107. doi: 10.1016/j.yebeh.2014.04.026

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., … Mathers, C. D. (2011). Global burden of disease in young people aged 10–24 years: a systematic analysis. *Lancet*, *377*(9783), 2093–2102. doi: 10.1016/s0140-6736(11)60512-6

Gualtieri, C. T., & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs.

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Archives of Clinical Neuropsychology, *21*(7), 623-643. doi: 10.1016/j.acn.2006.05.007

Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., … Gershon, R. (2014). Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. *Journal of the International Neuropsychological Society*, *20*(6), 588–598. doi: 10.1017/S1355617714000241

Hetrick, S. E., Cox, G. R., Fisher, C. A., Bhar, S. S., Rice, S. M., Davey, C. G., & Parker, A. G. (2015). Back to basics: could behavioural therapy be a good treatment option for youth depression? A critical review. *Early Intervention in Psychiatry*, *9*(2), 93-99. doi: 10.1111/eip.12142

Holsinger, T., Deveau, J., Boustani, M., & Williams Jr, J. W. (2007). Does this patient have dementia? *Journal of the American Medical Association*, *297*(21), 2391-2404. doi: 10.1001/jama.297.21.2391

Holzer, L., Chinet, L., Jaugey, L., Plancherel, B., Sofia, C., Halfon, O., & Randolph, C. (2007). Detection of cognitive impairment with the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in adolescents with psychotic symptomatology. *Schizophrenia Research*, *95*(1-3), 48-53. doi: 10.1016/j.schres.2007.06.013

Hoppe, C., Fliessbach, K., Schlegel, U., Elger, C. E., & Helmstaedter, C. (2009). NeuroCog FX: Computerized screening of cognitive functions in patients with epilepsy. *Epilepsy and Behavior*, *16*(2), 298-310. doi: 10.1016/j.yebeh.2009.07.041

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology*, *27*(6), 683-689. doi: 10.1081/13803390490918435

Jewsbury, P. A., Bowden, S. C., & Duff, K. (2017). The Cattell-Horn-Carroll Model of Cognition for Clinical Assessment. *Journal of Psychoeducational Assessment*, *35*(6), 547-567. doi: 10.1177/0734282916651360

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the cattell-horn-carroll model. *Journal of Experimental Psychology: General*, *145*(2), 220-245. doi: 10.1037/xge0000119

Jones, W. P., Loe, S. A., Krach, S. K., Rager, R. Y., & Jones, H. M. (2008). Automated Neuropsychological Assessment Metrics (ANAM) and Woodcock-Johnson III tests of cognitive ability: A concurrent validity study. *Clinical Neuropsychologist*, *22*(2), 305-320. doi: 10.1080/13854040701281483

Kahn, D. A., Asbell, S. J., & Donders, J. (2015). Clinical utility of the LANSE-A in adolescents with traumatic brain injury. *Rehabilitation Psychology*, *60*(2), 187-192. doi: 10.1037/rep0000034

Kassebaum, N., Kyu, H. H., Zoeckler, L., Olsen, H. E., Thomas, K., Pinho, C., … Vos, T. (2017). Child and adolescent health from 1990 to 2015: Findings from the global burden of diseases, injuries, and risk factors 2015 study. *JAMA Pediatrics*, *171*(6), 573–592. doi: 10.1001/jamapediatrics.2017.0250

Kavanaugh, B. C., Cancilliere, M. K., & Spirito, A. (2019). Neurocognitive heterogeneity across the spectrum of psychopathology: Need for improved approaches to deficit detection and intervention. *CNS Spectrums*. doi: 10.1017/S1092852919001081

Kelly, M. P., Coldren, R. L., Parish, R. V., Dretsch, M. N., & Russell, M. L. (2012). Assessment of acute concussion in the combat environment. *Archives of Clinical Neuropsychology*, *27*(4), 375-388. doi: 10.1093/arclin/acs036

Kerr, E. N., & Fayed, N. (2017). Cognitive predictors of adaptive functioning in children with symptomatic epilepsy. *Epilepsy Research*, *136*, 67-76. doi: 10.1016/j.eplepsyres.2017.07.015

Krull, K. R., Okcu, M. F., Potter, B., Jain, N., Dreyer, Z., Kamdar, K., & Brouwers, P. (2008). Screening for neurocognitive impairment in pediatric cancer long-term survivors. *Journal of Clinical Oncology*, *26*(25), 4138-4143. doi: 10.1200/JCO.2008.16.8864

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: Sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care and Pain*, *8*(6), 221-223. doi: 10.1093/bjaceaccp/mkn041

Lanca, M. (2018). Integration of neuropsychology in primary care. *Archives of Clinical Neuropsychology*, *33*(3), 269–279. doi: 10.1093/arclin/acx135

Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience And Biobehavioral Reviews*, *94*, 179-195. doi: 10.1016/j.neubiorev.2018.09.005

Laver, K., Cumming, R. G., Dyer, S. M., Agar, M. R., Anstey, K. J., Beattie, E., … Yates, M. W. (2016). Clinical practice guidelines for dementia in Australia. *Medical Journal of Australia*, *204*(5), 1–3.e2. doi: 10.5694/mja15.01339

Lee, R. S. C., Hermens, D. F., Naismith, S. L., Kaur, M., Guastella, A. J., Glozier, N., … Hickie, I. B. (2018). Clinical, neurocognitive and demographic factors associated with functional impairment in the Australian Brain and Mind Youth Cohort Study (2008–2016). *BMJ Open*, *8*(12). doi: 10.1136/bmjopen-2018-022659

Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., & McAllister, T. (2010). Examination of the construct validity of impact™ computerized test, traditional, and experimental neuropsychological measures. *Clinical Neuropsychologist*, *24*(8), 1309-1325. doi: 10.1080/13854046.2010.516072

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. doi: 10.1097/JTO.0b013e3181ec173d

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. doi: 10.1016/j.intell.2008.08.004

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Plos Medicine*, *6*(6), e1000097. doi: 10.1371/journal.pmed1000097

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, *56*(1), 45-50. doi: 10.4103/0301-4738.37595

Patel, S. K., Wong, A. L., Cuevas, M., & Van Horn, H. (2013). Parenting stress and neurocognitive late effects in childhood cancer survivors. *Psycho-Oncology*, *22*(8), 1774-1782. doi: 10.1002/pon.3213

Pike, N. A., Poulsen, M. K., & Woo, M. A. (2017). Validity of the montreal cognitive assessment screener in adolescents and young adults with and without congenital heart disease. *Nursing Research*, *66*(3), 222-230. doi: 10.1097/nnr.0000000000000192

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press

Portaccio, E., Goretti, B., Lori, S., Zipoli, V., Centorrino, S., Ghezzi, A., … Gallo, P. (2009). The brief neuropsychological battery for children: A screening tool for cognitive impairment in childhood and juvenile multiple sclerosis. *Multiple Sclerosis*, *15*(5), 620–626. doi: 10.1177/1352458508101950

Pottie, K., Rahal, R., Jaramillo, A., Birtwhistle, R., Thombs, B. D., Singh, H., … Tonelli, M. (2016). Recommendations on screening for cognitive impairment in older adults. *CMAJ Canadian Medical Association Journal*, *188*(1), 37–46. doi: 10.1503/cmaj.141165

Proffitt, T. M., Brewer, W. J., Parrish, E. M., McGorry, P. D., & Allott, K. A. (2018). Reasons for referral and findings of clinical neuropsychological assessment in youth with mental illness: A clinical file audit. *Applied Neuropsychology: Child*, *7*(2), 164–174.

Raiker, J. S., Manning, E., Herrington, B., May, A. C., Haynes, S., Graves, P. E., & Karlson, C. W. (2015). Brief neurocognitive screening in youth with brain tumours: A preliminary investigation of the Lebby-Asbell Neurocognitive Screening Examination (LANSE). *Brain Injury*, *29*(10), 1192-1198. doi: 10.3109/02699052.2015.1035331

Roebuck-Spencer, T. M., Glen, T., Puente, A. E., Denney, R. L., Ruff, R. M., Hostetter, G., & Bianchini, K. J. (2017). Cognitive screening tests versus comprehensive neuropsychological test batteries: A national academy of neuropsychology education paper. *Archives of Clinical Neuropsychology*, *32*(4), 491-498. doi: 10.1093/arclin/acx021

Rossell, S. L. (2019). Response to: Bryce and Allott – Cognitive screening: A significant unmet need in youth mental health. 30 minutes is too long! *Australian And New Zealand Journal Of Psychiatry*, *53*(8), 814. doi: 10.1177/0004867419832652

Schatz, P., & Putz, B. O. (2006). Cross-validation of measures used for computer-based assessment of concussion. *Applied Neuropsychology*, *13*(3), 151-159. doi: 10.1207/s15324826an1303_2

Schmidt, R. L., & Factor, R. E. (2013). Understanding sources of bias in diagnostic accuracy studies. *Archives of Pathology and Laboratory Medicine*, *137*(4), 558-565. doi: 10.5858/arpa.2012-0198-RA

Sheehan, B. (2012). Assessment scales in dementia. *Therapeutic Advances in Neurological Disorders*, *5*(6), 349–358. doi: 10.1177/1756285612455733

Solomon, G. S., & Kuhn, A. (2014). Relationship between concussion history and neurocognitive test performance in national football league draft picks. *American Journal of Sports Medicine*, *42*(4), 934-939. doi: 10.1177/0363546513518742

Straus, S. W., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2010). *Evidence-Based Medicine: How to Practice and Teach It*. (4th ed.). New York, NY: Churchill Livingstone.

Summers, M. J., Bondi, M. W., & Bowden, S. C. (2019). Brief cognitive screening versus comprehensive neuropsychological assessment - Is one better than the other? *National Academy of Neuropsychology Bulletin*, *32*(1), 8–12.

Tsoi, K. K. F., Chan, J. Y. C., Hirai, H. W., Wong, S. Y. S., & Kwok, T. C. Y. (2015). Cognitive tests to detect dementia a systematic review and meta-analysis. *JAMA Internal Medicine*, *175*(9), 1450-1458. doi: 10.1001/jamainternmed.2015.2152

Valovich McLeod, T. C., Barr, W. B., McCrea, M., & Guskiewicz, K. M. (2006). Psychometric and measurement properties of concussion assessment tools in youth sports. *Journal of Athletic Training*, *41*(4), 399–408.

Vega-Fernandez, P., Vanderburgh White, S., Zelko, F., Ruth, N. M., Levy, D. M., Muscal, E., … Brunner, H. I. (2015). Cognitive performance scores for the pediatric automated neuropsychological assessment metrics in childhood-onset systemic lupus erythematosus. *Arthritis Care and Research*, *67*(8), 1119–1127. doi: 10.1002/acr.22571

Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., … Murray, C. J. L. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *386*(9995), 743–800. doi: 10.1016/S0140-6736(15)60692-4

Weintraub, S., Dikmen, S., Heaton, R., Tulsky, D., Zelazo, P., Bauer, P., … Gershon, R. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *12*(80), s54–64.

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., … Bossuyt, P. M. M. (2011). Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536. doi: 10.7326/0003-4819-155-8-201110180-00009

https://doi.org/10.1017/S1355617720001241 Published online by Cambridge University Press