Measurement properties of performance-based instruments to assess mental function during activity and participation in traumatic brain injury

Kristensen, Lola Qvist; Muren, Marie Almkvist; Petersen, Annemette Krintel; van Tulder, Maurits W; Gregersen Oestergaard, Lisa

published in
Scandinavian Journal of Occupational Therapy
2020

DOI (link to publisher)
10.1080/11038128.2019.1689291

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Scandinavian Journal of Occupational Therapy

ISSN: 1103-8128 (Print) 1651-2014 (Online) Journal homepage: https://www.tandfonline.com/loi/iocc20

Measurement properties of performance-based instruments to assess mental function during activity and participation in traumatic brain injury: A systematic review

Lola Qvist Kristensen, Marie Almkvist Muren, Annemette Krintel Petersen, Maurits W. van Tulder & Lisa Gregersen Oestergaard

To cite this article: Lola Qvist Kristensen, Marie Almkvist Muren, Annemette Krintel Petersen, Maurits W. van Tulder & Lisa Gregersen Oestergaard (2020) Measurement properties of performance-based instruments to assess mental function during activity and participation in traumatic brain injury: A systematic review, Scandinavian Journal of Occupational Therapy, 27:3, 168-183, DOI: 10.1080/11038128.2019.1689291

To link to this article: https://doi.org/10.1080/11038128.2019.1689291

SCANDINAVIAN JOURNAL OF OCCUPATIONAL THERAPY
2020, VOL. 27, NO. 3, 168–183
https://doi.org/10.1080/11038128.2019.1689291

Taylor & Francis
Taylor & Francis Group

Check for updates

REVIEW ARTICLE

Measurement properties of performance-based instruments to assess mental function during activity and participation in traumatic brain injury: A systematic review

Lola Qvist Kristensen[a,b], Marie Almkvist Muren[a,c], Annemette Krintel Petersen[a,c,d], Maurits W. van Tulder[a,e] and Lisa Gregersen Oestergaard[a,c,d,f,g]

[a]Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark; [b]Neurorehabilitation Skive, Hammel Neurorehabilitation Centre and University Research Clinic, Skive, Denmark; [c]Centre of Research in Rehabilitation (CORIR), Aarhus University Hospital and Aarhus University, Aarhus, Denmark; [d]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [e]Department of Health Sciences and the EMGO + Institute for Health and Care Research, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, the Netherlands; [f]Department of Public Health, Aarhus University, Aarhus, Denmark; [g]Department of Public Health, The Research Initiative of Activity Studies and Occupational Therapy, Research Unit of General Practice, University of Southern Denmark, Odense, Denmark

ABSTRACT

Background: Performance-based measures that focus primarily on the ability to engage in ADL are routinely used by occupational therapists to assess a client's cognitive abilities.

Objective: To perform a systematic review to investigate measurement properties of performance-based instruments to assess mental function during activity and participation in individuals with traumatic brain injury.

Material and methods: Pubmed, EMBASE, CINAHL, PsycINFO and OTseeker were searched. The Consensus-based Standards for the selection of health measurement instruments checklist was used to evaluate methodological quality of each included study. The quality criteria adapted by Terwee were applied to extract the results of each measurement property followed by a best evidence synthesis.

Results: Twenty-eight articles, including 40 ratings of measurement properties, were included. The combination of the Functional Independence Measure and the Functional Assessment Measure showed moderate evidence of good internal consistency (Cronbach's alpha 0.99), but conflicting evidence of reliability (ICC 0.83) and poor evidence of construct validity. All other instruments showed limited or unknown evidence.

Conclusions: This review provides an overview of measurement properties of performance-based instruments and contributes to such methodological considerations before choosing an instrument. Though, the results reveal a lack of high-quality evidence for any of the measurement properties, it is recommended to use tools with the highest possible evidence for positive ratings.

Significance: This review contributes with psychometric evidence on instruments to use in occupational therapy practice and research.

ARTICLE HISTORY
Received 21 February 2019
Revised 5 October 2019
Accepted 31 October 2019

KEYWORDS
Brain injury; cognitive impairment; psychometric properties; ecological validity; validity; reliability; rehabilitation

Introduction

Traumatic brain injury (TBI) is associated with substantial disability and mortality. In Europe, the incidence of hospital admissions due to TBI is estimated to 262 people per 100,000 [1]. Disabilities associated with TBI have consequences for the individual's performance of activities of daily living (ADL), as well as consequences for society. The economic burden of TBI due to increased costs of health care and loss of productivity is high [2]. Impaired mental function is a frequently reported TBI disability, reducing the individual's performance of ADL and participation in society; this leads to increased rehabilitation needs [3,4]. The International Classification of Functioning, Disability and Health (ICF) is a World Health Organisation framework aiming at classifying health and health-related domains, including mental functions. Specific mental function subdomains in the ICF is divided into the following subdomains: attention, memory, psychomotor function, emotion, perception, thought, higher-level cognitive function (executive function), mental function of language, calculation function,

CONTACT Lola Qvist Kristensen lola.qvist@rm.dk Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark

mental function of sequencing complex movements, and experience of self and time function [5,6]. Especially decreased memory, attention, and executive functions are often reported in TBI individuals [7]. Although mental function may improve over time as a result of both restorative mechanisms in the brain and rehabilitation, one in three TBI survivors still experience impaired mental function three years post-injury [4,8,9]. A meta-analysis showed that impaired mental function following TBI resulted in long-lasting ADL limitations with a need for ADL assistance [7]. It is well known that impaired mental function affects the possibility to be independent in performing and participating in daily routines, employment, mobility, self-care and leisure as well as interpersonal/social interactions [9,10].

There are different available approaches to assess impaired mental function with no consensus on which outcome measures are the most appropriate [11,12]. Pen and paper tests, like the Montreal Cognitive Assessment or Star Cancellation test, are often used when screening for mental impairments [13,14]. Although useful to perform a quick screening to quantify the extent of impairment, these tools do not assess mental function in relation to the individual's ability to engage in ADL, which is an important focus in occupational therapy [15–17]. Performance based measures that focus primarily on the ability to engage in ADL are routinely used by occupational therapists to assess a client's cognitive abilities and inform occupational therapy practice [15–17]. A cross-sectional survey found that the use of performance-based instruments not addressing activity and participation made it difficult to link results to ADL performance and rehabilitation strategies [17]. According to the ICF, activity and participation relate to broader aspects of functioning, including communication, mobility, interpersonal interactions, self-care, learning, and application of knowledge [18]. Performance-based instruments for assessment of mental function during activity and participation are used in occupational therapy to reflect the impact on performance of and engagement in ADL. Examples of instruments are the Functional Independence Measure (FIM) and The Perceive, Recall, Plan and Perform system of Task analysis (PRPP) [19,20]. International guidelines emphasise the importance of using instruments with adequate measurement properties for assessment of mental function in TBI individuals [21,22].

The objective of this systematic review was to investigate measurement properties of performance-based instruments to assess mental function during activity and participation in individuals with TBI.

This with the purpose to increase occupational therapists' knowledge of available instruments and help to understand, which instruments are best to support occupational therapy interventions.

Material and methods

Study design and registration

This manuscript was reported according to the Preferred Reporting Items for Systematic Review and Meta-Analysis [23]. A protocol was prepared for this systematic review and registered in the International Prospective Register of systematic reviews, PROSPERO, with the registration number CRD42017053881. A study protocol has been published [24]. The review was conducted using the Consensus-based Standards for the selection of Health Measurements Instruments (COSMIN) methodology [25].

Data sources and searches

The search strategy described was used for a series of systematic reviews in preparation on different types of acquired brain injury. This systematic review only includes studies of TBI and is the first in the series.

A computer-based literature search in the databases PubMed, EMBASE, CINAHL, PsycINFO and OTseeker was performed from the inception of each database to 17 May 2018. The search strategy consisted of three search blocks: 'Acquired brain injury', 'mental function' and 'method of assessment'. The term 'Acquired brain injury' was used because this review was part of a series of systematic reviews on acquired brain injury also including, for example, stroke. The search blocks were combined with the Boolean operator 'AND' and further with a published search filter to identify articles with studies on measurement properties related to reliability, validity and responsiveness in PubMed, CINAHL and EMBASE [26]. A translation of the published search filter was applied to the literature search in PsycINFO and OTseeker [26]. The search strategy used Medical Subjects Headings (MeSH), and controlled vocabulary, i.e. EMTREE, Cinahl headings and Thesaurus in addition to free-text terms derived from the search blocks. Citation search and reference lists of relevant articles were screened for additional studies to meet the inclusion criteria and to detect potential grey literature. Potentially relevant articles were retrieved and assessed. The online bibliographic programme RefWorks (https://www.refworks.com) was used to manage all identified materials. The full search strategy is available upon request from the corresponding author.

Study selection

Inclusion and exclusion criteria

A study was included if it was a full-text original study of measurement properties on performance-based instruments to assess mental function during activity and participation in individuals with a TBI. All TBI categories were included, i.e. mild, moderate and severe as well as unknown severity, regardless of time since injury and presence of any comorbidity. Only studies in which the measurement properties were assessed disease-specific for adult TBI individuals were included. When a study did not include a separate analysis of TBI, it was excluded.

To be included, a study needed to report on performance-based instruments with at least one measurement property and at least one sub-domain of mental function performed within at least one subdomain of activity and participation. Performance-based instruments are included, which are direct measures of actual performance of everyday activities that have high potential to engage clients in a relatively brief amount of time [27]. Paper and pencil tests are excluded, because within occupational therapy these are typically not considered performance-based instruments to assess mental function during activity and participation [17].

Measurement properties were defined using the COSMIN checklist, which includes reliability (internal consistency, reliability, measurement error), validity (content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity) and responsiveness [25,28].

The ICF Framework definition of specific mental functions was used, which relates to attention, memory, psychomotor function, emotion, perception, thought, higher-level cognitive function (executive function), mental function of language, calculation function, mental function of sequencing complex movements and experience of self and time function [18]. Likewise, the ICF definition of activity and participation was used, which relates to communication, mobility, interpersonal interactions, self-care, learning and applying knowledge [18]. To be included, the instrument needed a separate scale for mental function.

Selection procedure

One review author (MAM) performed the literature search in collaboration with a health science librarian and removed duplicates with the Refworks tools. Two review authors (MAM and LQK) independently assessed all titles and abstracts for relevance and screened the reference lists. There was no language restriction. Articles in other languages than the authors mastered (Danish, English, Norwegian, Swedish, Dutch, German, French and Spanish), were translated in the authors' network or by an interpreter. Both review authors (MAM and LQK) applied the eligibility criteria to each included study. The review authors were blinded to each other's scores until meeting to achieve consensus on each study, after scoring all studies. In case of disagreement, a third author (LGO) was consulted. All authors participating in the selection procedure were clinical and research content area experts.

Appraisal of the methodological quality of included studies

In pairs (LQK and MAM or LGO), two authors independently evaluated the methodological quality of the included studies using the COSMIN checklist [25]. The COSMIN checklist is a critical appraisal tool with standards on how to evaluate methodological quality of studies on measurement properties of health-related instruments [25,28]. Reliability concerns the degree to which an instrument is free from measurement error and the extent to which scores are the same for repeated measurements under different conditions for patient, who have not changed. This includes three properties: internal consistency, measurement error and reliability.

Validity relates to the extent to which an instrument measures the construct it is supposed to measure. The item encompasses content validity, construct validity (including hypothesis testing and cross-cultural validity) and criterion validity. Responsiveness relates to the ability of an instrument to detect change in the underlying construct. In total, nine measurement properties are evaluated on the COSMIN checklist [25,28,29]. A four-point scale was used to rate each measurement property study as either excellent, good, fair or poor based on worst score counts, meaning that the lowest score within a box determined the overall study quality. A study can thus be downgraded based on standards for the quality of the study according to the COSMIN checklist. Detailed information on these standards can be found elsewhere [29].

In pairs (LQK and MAM or LGO), two authors assessed the quality of the studies separately and blinded to each other's scores. Disagreements were resolved through consensus discussion; if no consensus was reached, a third author (LGO or AKP) was consulted.

Evidence synthesis

Evidence synthesis included results of the measurement properties and evidence of each included instrument. In step one, results of each measurement property were extracted from the original studies using the Terwee quality criteria [29]. A criterion was defined for each measurement property for a positive (+), negative (–) or indeterminate (?) rating, depending on the design, methodology and outcomes [30].

Step two accompanied results from step one by a best evidence synthesis using a 'levels of evidence' rating to summarise all evidence for each instrument [31].

The possible levels of evidence for a measurement property for each instrument were (I) strong (+++ or –) if findings were consistent in multiple studies with good quality or one study with excellent quality, (II) moderate (++ or –) if findings were consistent in multiple studies of fair quality or one study with good quality, (III) limited (+ or –) if only one study with fair quality was available, (IIII) conflicting (±) if findings were not consistent but conflicting or (V) unknown (?) if only studies of poor quality or no studies at all were available.

Results

Study selection

The search resulted in a total of 1,354 unique articles after removing duplicates. After screening titles and abstracts, 1,239 articles were excluded. A total of 115 articles were selected for full-text inclusion. The full-text process resulted in further exclusion of 96 articles with no separate analysis of TBI ($n = 91$) or not reporting on performance-based instruments ($n = 5$). Screening of reference lists resulted in inclusion of nine additional articles; thus, 28 articles were included in this review (Figure 1).

Study characteristics

Study characteristics are described in Table 1. Of the 28 articles included in the review, 16 were from the USA [32–47], four from Australia [48–51] and one from UK [52], Brazil [53], Canada [54], New Zealand [55], Chile [56], Netherlands [57], Italy [58] and Spain [59], respectively. Sample sizes varied in the included studies from 2 [48] to 8,136 [39] participants and ages ranged from 14 to 86 years [47]. TBI severity involved mild, moderate and severe TBI.

Instruments

In total 24 instruments were evaluated in the 28 included articles. The instruments were the Baycrest Multiple Errands Test (BMET) [32], the Behavioural Dysregulation Rating Scale (BDRS) [33], a cognitive test battery [57], Disability Rating scale (DRS) [34,35], the Executive Function Route-finding Task (EFRT) [36], the Functional Assessment Measurement (FAM) [58], the Functional Cognition Index (FCI) [37], the Functional Cognitive task (FCT) [38], the FIM [39], a combination of FIM + FAM [52], the Test of Functional Executive Abilities (TOFEA) [40], the Hotel Task [53], Levels of Cognitive Functioning Scale (LCFS) [59], the Mayo-Portland Adaptability Inventory–4 (MPAI-4) [41,42], the Measure of Executive Functioning [55], the Neurobehavioural Cognitive Status Examination (NCSE) [43,47], the Neuropsychological Assessment Battery daily living tests (NAB-DL) [44], an Occupation-based Online Awareness Assessment [48], the PRPP [49], the Screen-Based Simulated Cup Of Tea (SBS-COT) [56], the Virtual Library Task (VLT) [50], the Virtual Planning Test (VIP) [45], a Virtual Reality Kitchen Task [46] and the Virtual Reality Shopping Task (VRST) [51].

Some articles included studies that combined several instruments to create a test battery, which thus became a separate instrument, not comparable to the original instruments. This was the case with the FAM, which appeared separately [58], in a combination with the FIM [52,54] and in combination with both DRS and LCFS [57].

Mental function domains

The subdomains higher level cognitive functions (executive functions) ($n = 27$) [32–50,52–59], memory ($n = 17$) [37–39,41–44,47,49–52,54,55,57–59], mental function of language ($n = 12$) [37–39,41–44,47, 52,54,57,58] and attention ($n = 14$) [37–39,41–44,47, 52,54,55,57–59] were the most reported. Other reported subdomains included emotion ($n = 8$) [38,39, 41,42,47,52,57,58], perception ($n = 6$) [41–44,49], experience of self ($n = 4$) [41,42,46,48] and psychomotor function ($n = 1$) [38]. No studies reported on the subdomains thought function, calculation function or mental function of sequencing complex movements.

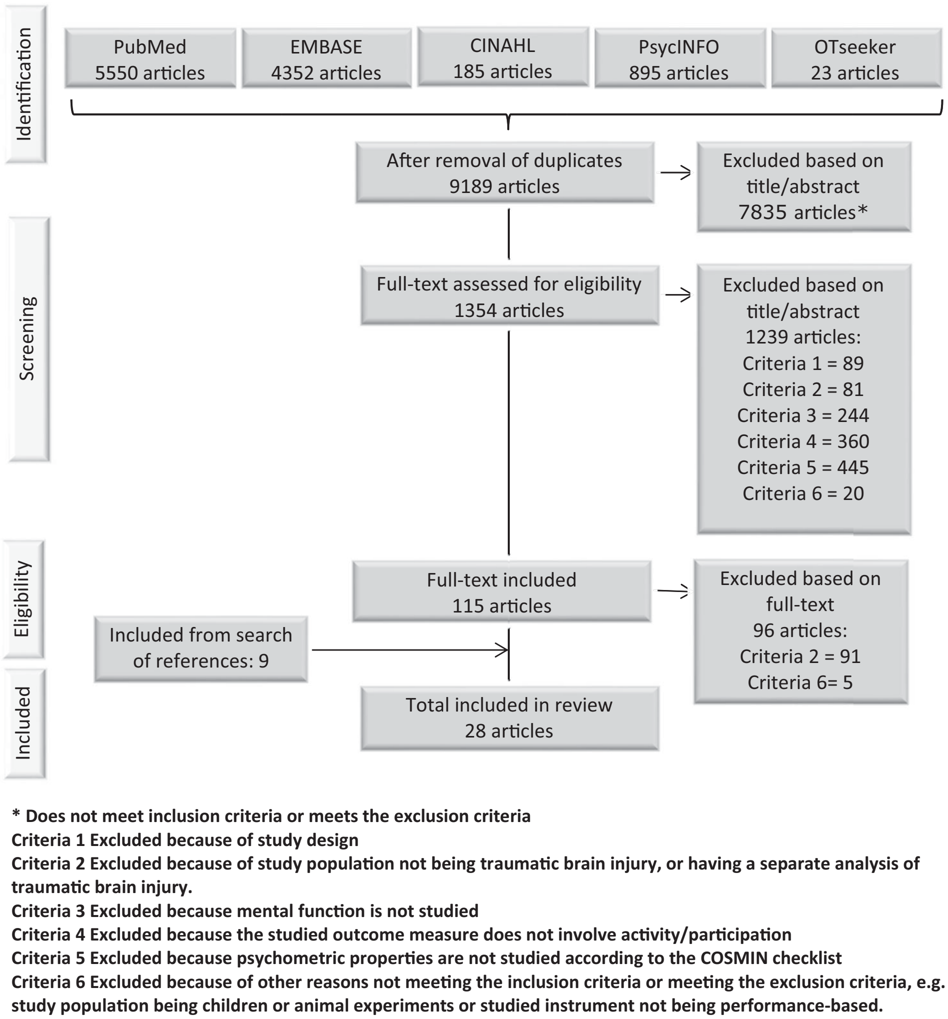

Figure 1. Flowchart. *Does not meet inclusion criteria or meets the exclusion criteria. Criteria 1: Excluded because of study design. Criteria 2: Excluded because of study population not being TBI, or having a separate analysis of TBI. Criteria 3: Excluded because mental function is not studied. Criteria 4: Excluded because the studied outcome measure does not involve activity/participation. Criteria 5: Excluded because psychometric properties are not studied according to the COSMIN checklist. Criteria 6: Excluded because of other reasons not meeting the inclusion criteria or meeting the exclusion criteria, e.g. study population being children or animal experiments or studied instrument not being performance-based.

Sixteen out of 24 instruments assessed more than one subdomain (range 2–10 subdomains) of mental function during activity and participation. The following instruments all included assessment of the five most reported subdomains; attention, memory, emotion, executive functions and mental function of language: the FIM [39,52,54], the FAM [52,54,57,58], the FCT [38], the MPAI-4 [41,42] and the NAB-DL [44]. Performed activities included self-care (FIM, FAM, MPAI-4) and domestic activities (NAB-DL, FCT). The FCT was the only instrument identified in this review that also assessed psychomotor function. The NCSE assessed several mental function subdomains, although only the judgement item of executive function was assessed during activity and participation.

Of the instruments assessing one or two subdomains of mental function, assessment of executive functions was the most reported. It was reported in the BMET [32], the EFRT [36], the TOFEA [40], the Hotel Task [53], Occupation-based Online Awareness Assessment [48], the SBS-COT [56], the VIP [45], the VLT [51] and the Virtual Reality Kitchen Task [46]. These instruments assessed executive function during performance of domestic activities such as shopping, cooking and route-finding along with major life areas such as work-tasks. While some tasks were performed in naturalistic environments, others were performed by computer or virtual reality. Two instruments (an Occupation-based Online Awareness Assessment and the Virtual Reality Kitchen Task) assessed mental function subdomains of experience of self through

Table 1. Study characteristics.

Author (year), Country (language)	Instrument	Sample size (*n*)	TBI severity	Age (year), mean (SD), range	Time post TBI, mean (SD), range	Mental function domain	Activity/ participation	COSMIN domain
Bamdad et al. (2003), USA (English) [40]	The TOEFA	340	Mild - Severe	25.71 (6.7)	136.89 (429.4) d	Executive functions	Activity and participation	Hypothesis testing Criterion validity
Boyd et al. (1993), USA (English) [36]	The EFRT	31	Mild-Moderate	25.4, 18–42	5 mo to 18 yrs	Executive functions	Activity	Reliability Criterion validity
Canty et al. (2014), Australia (English) [51]	The VRST	30	Severe	31.68 (11.77)	138 (81.94) d	Memory	Activity	Hypothesis testing Responsiveness
Cardoso et al. (2015), Brazil (Brazilian) [53]	The Hotel Task	10	N/A	37.09 (10.29)	N/A	Executive functions	Activity	Content validity Cross cultural validity
Dawson et al. (2009), USA (English) [32]	The BMET	13	Moderate - Severe	43.3 (9.3)	10.9 (8.2) yrs	Executive functions	Activity and participation	Reliability
Doig et al. (2017), Australia (English) [48]	An Occupation-based Online Awareness Assessment	2	Severe	22, 23 yrs	55, 63 mo	Executive functions Experience of self	Activity	Reliability
Donaghy et al. (1998) Canada (N/A) [54]	FIM + FAM, combined	53	Severe	38.2 (13.7), 19–65	N/A (measures every week during postacute rehabilitation)	Attention Memory Emotion Executive functions Mental functions of language	Activity and participation	Reliability
Doninger et al. (2000) USA (English) [47]	NCSE	186	N/A	34, 18–60	Median 32 mo, 1–336 mo	Attention Memory Perception Executive functions Mental functions of language	Activity and participation in judgement item	Internal consistency
Gouvier (1987) Spain (Spanish) [59]	LCFS	40	N/A	Median 24.8 (13.1), 5–69	N/A (measures four days a week during inpatient rehabilitation)	Attention Memory Executive functions	Activity	Reliability Hypothesis testing Criterion validity
Hawley et al. (1999), UK (English) [52]	FIM + FAM, combined	652	N/A	34.4 (13), 16–72	N/A (measures at admission and discharge from acute inpatient rehabilitation)	Attention Memory Emotion Executive functions Mental functions of language	Activity and participation	Internal consistency Hypothesis testing
Kean et al. (2011), USA (English) [41]	The MPAI-4	603	Mild - Severe	40.2 (14.8), 14–86	81 (1418.9) days	Attention Memory Emotion Perception Executive functions Mental functions of language Experience of self	Activity and participation	Internal consistency
Labi et al. (1998) USA (English) [37]	FCI	Doubtful. Stated 35 subjects, but described as 12 females and 22 males.	Moderate-Severe	Male: 38.8 (N/A) Female: 52.2 (N/A)	N/A	Attention Memory Executive functions Mental functions of language	Activity and participation	Reliability Criterion validity
LeBlanc et al. (2012), USA (English) [38]	The FCT	46	Severe	40.7 (17.3), 17–80	>2 yrs	Attention Memory Psychomotor function Emotion Executive functions Mental functions of language	Activity and participation	Internal consistency Criterion validity

(*continued*)

Table 1. Continued.

Author (year), Country (language)	Instrument	Sample size (*n*)	TBI severity	Age (year), mean (SD), range	Time post TBI, mean (SD), range	Mental function domain	Activity/ participation	COSMIN domain
Lewis et al. (2017) New Zealand (English) [55]	Measure of Executive Functioning	Study 1: 10 Study 2: 27 with TBI 16 normative	Most cases severe	Study 1: 46.4 (14.5) Study 2: Normative 48.5 (15.7), TBI 49.1 (18.5)	Study 1: 5.8 (4.8) yrs Study 2: 5.9 (7.4) yrs	Attention Memory Executive functions	Activity	Reliability Criterion validity
Martínez-Pernía et al. (2017) Chile (N/A) [56]	SBS-COT	8 dyads (patient/ professional)	N/A	N/A	N/A	Executive functions	Activity	Content validity
McKeon et al. (2017) USA (English) [33]	BDRS	14	N/A	40.5 (3.28)	152.2 mo (36.0)	Executive functions	Activity	Reliability Content validity Hypothesis testing
Malec et al (2012) USA (English) [34]	DRS	406 (287 TBI subjects /119 caregivers)		38.35 (18.07)	Follow-up between 1 and 20 yrs	Executive functions	Activity and participation	Internal consistency Criterion validity
Murrey et al. (2005) USA (English) [42]	MPAI-4	43 with TBI + frontal lobe damage 69 with TBI – frontal lobe damage	Mild-severe	TBI + frontal lobe damage: 34.1 (11.0) TBI – frontal lobe damage: 34.3 (13.5)	TBI + frontal lobe damage: 5.2 (5.4) yrs TBI - frontal lobe damage: 4.8 (6.4) yrs	Attention Memory Emotion Perception Executive functions Mental functions of language Experience of self	Activity and participation	Hypothesis testing
Nott MT et al. (2009), Australia (English) [49]	PRPP	5	Moderate	N/A, 19–58	N/A	Memory Perception Executive functions	Activity and participation	Reliability
O'Neil-Pirozzi et al. (2010), USA (English) [45]	The VIP	75	Mild - Severe	46.5 (10.5)	11.8 (9.6) yrs	Executive functions	Activity	Reliability
Pretz et al. (2016) USA (English) [39] Memory Emotion Executive functions Mental functions of language	FIM	8136	N/A	39.6 (18.6)		Admission + discharge: 27.3 (25.3) d + 1 yr follow-up	Attention Internal consistency	Activity and participation
Rappaport et al. (1982) USA (English) [35]) Reliability	DRS	88	Severe	N/A	Admission within	90 days + follow-up at 12 mo	Executive functions	Activity and participation
Renison et al. (2012), Australia (English) [50]	The VLT	30	Moderate - Severe	N/A	Median: 5, 1–28 yrs	Memory Executive functions	Activity	Reliability Hypothesis testing Criterion validity
Tesio et al. (1998), Italy (Italian) [58]	The FAM	60	Moderate – Severe	30.7 (13.3), 16–65	Median: 16, 2–88 mo	Attention Memor y Emotion Executive functions Mental functions of language	Activity and participation	Hypothesis testing
		25	Moderate – Severe	35.3 (12.8), 18–50		Attention (FAM, LCFS) Memory (FAM, LCFS)	Activity and participation	Reliability Responsiveness

(*continued*)

Table 1. Continued.

Author (year), Country (language)	Instrument	Sample size (n)	TBI severity	Age (year), mean (SD), range	Time post TBI, mean (SD), range	Mental function domain	Activity/ participation	COSMIN domain
van Baalen et al. (2006), Netherlands (N/A) [57]	A cognitive test battery; FAM, DRS, LCFS				Discharge: 35 (19) days + 1 yr post TBI	Emotion (FAM) Executi ve functions (FAM, DRS, LCFS) Mental functions of language (FAM)		
Wallace et al. (2000), USA (English) [43]	NCSE	48	Severe	29.8 (11.3)	113.3 (82.6) d	Attention Memory Perception Executive functions Mental functions of language	Activity and participation in judgement item	Criterion validity
Zgaljardic et al. (2011), USA (English) [44]	NAB-DL	47	Moderate – Severe	31.7 (11.4)	16 (26.6) mo	Attention Memory Perception Executi ve functions Mental functions of language	Activity and participation	Criterion validity
Zhang et al. (2003), USA (English) [46]	A Virtual Reality Kitchen Task	54	Mild - moderate	31.8 (12.7), 18–69	5 we – 240 mo	Executive functions Experience of self	Activity and participation	Reliability Criterion validity

computerised activities and one instrument (the MPAI-4) assessed the subdomain in a naturalistic environment.

One of the instruments (the VLT) assessed memory beside executive functions. Another virtual reality instrument identified to assess memory was the VRST simulating shopping situation to remember items in a pre-specified order. Memory was also assessed in the PRPP in a naturalistic environment during activities that were meaningful to the patient [49]. The PRPP was the only instrument identified that assessed perception.

Methodological quality

The methodological quality of the studies as well as quality of each measurement property is presented in Table 2. Some studies reported more than one measurement property, but no studies addressed all the properties included in the COSMIN checklist. Levels of evidence of the overall quality of each instrument's measurement property are presented in Table 3. The evidence level for most measurement properties of each instrument is based on only one study.

Reliability

Quality and results. A total of 20 studies reported on reliability, divided into internal consistency ($n = 6$) and reliability ($n = 14$). The latter included studies on one or more of the following: inter-rater reliability ($n = 11$), test-retest ($n = 3$) and intra-rater reliability ($n = 2$). None of the included studies reported on measurement error (Table 2). The quality of each study reporting on reliability was mostly poor ($n = 10$) or fair ($n = 9$); one study of the internal consistency on the combined FIM + FAM achieved a good score [52]. The quality of the studies on reliability was downgraded due to the risk of selection bias, such as no description of how missing items were handled, and no description of patients only being included if data were complete (Table 2).

Evidence synthesis. Levels of evidence related to reliability were either positive with moderate ($n = 1$) to limited ($n = 6$) evidence or of unknown rating ($n = 12$) due to poor study quality (Table 3). The instrument with a moderate positive evidence was the combined FIM + FAM on internal consistency, whereas instruments with limited positive evidence included internal consistency of the DRS, the FIM, the LCFS and reliability of the FCI, the VLT and a Virtual Reality Kitchen Task. Instruments with poor unknown rating

Table 2. Methodological quality of each study on each COSMIN measurement property/Terwee quality criteria of each measurement property.

Instrument	Study author	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross cultural validity	Criterion validity	Responsiveness
1. The BMET	Dawson et al. (2009) [32]		Poor/NA							
2. BDRS	McKeon et al (2017) [33]		Poor/?		Fair/?		Poor/+/–*			
3. A cognitive test battery; FAM, DRS, The Rancho Los Amigos LCFS	van Baalen et al. (2006) [57]		Poor/+ ICC on FIM/FAM 0.92/0.70							Poor/NA
4. DRS	Malec et al. (2012) [34] Rappaport et al. (1982) [35]	Fair/+ Cronbach's alpha 0.84	Poor/+ Correlation 0.97–0.98						Fair/?	
5. The EFRT	Boyd et al. (1993) [36]		Poor/+ Pearsons correlations 0.94						Fair/?	
6. The FAM	Tesio et al. (1998) [58]						Fair/?			
7. FCI	Labi et al (1998) [37]		Fair/+ ICC 0.88						Fair/?	
8. The FCT	LeBlanc et al. (2012) [38]	Poor/+ Cronbach's alpha .878							Fair/+/–**	
9. FIM + FAM, combined	Hawley et al. (1999) [52] Donaghy et al. (1998) [54]	Good/+ Cronbach's alpha 0.99	Fair/+/–*** ICC 0.83				Poor/+			
10. FIM	Pretz et al (2016) [39]	Fair/+ Cronbach's alpha 0.98								
11. The TOEFA	Bamdad et al. (2003) [40]						Fair/?		Fair/– Correlations between –0.259; 0.167 compared to standard executive measures	
12. The Hotel Task	Cardoso et al. (2015) [53]				Fair/+			Poor/?		
13. LCFS	Gouvier et al. (1987) [59]		Fair/+ Spearmans correlation 0.82; 0.89				Fair/?		Fair/?	
14. The MPAI-4	Kean et al. (2011) [41] Murrey et al. (2005) [42]	Fair/? Person/item separation 3.03/14.5					Fair/?			
15. Measure of Executive Functioning	Lewis et al. (2017) [55]		Poor/+/–**** ICC 0.0–1.00						Poor/?	
16. NCSE	Wallace et al. (2000) [43] Doninger et al. (2000) [47]	Fair/?							Fair/– Pearsons correlation 0.56 compared to neuropsychological tests	
17. NAB-DL	Zgaljardic et al. (2011) [44]								Fair/NA	
18. An Occupation-based Online Awareness Assessment	Doig et al. (2017) [48]		Poor/?							
19. The PRPP	Nott MT et al. (2009) [49]		Poor/–/ NA******							
20. SBS-COT	Martínez-Pernía et al. (2017) [56]				Poor/+					
21. The VLT	Renison et al. (2012) [50]		Fair/+ Pearsons correlation 1.0				Fair/+		Fair/– Pearsons correlation 0.68 compared to Real Library Task	
22. The VIP	O'Neil-Pirozzi et al. (2010) [45]		Poor/+/– ********							

(*continued*)

Table 2. Continued.

Instrument	Study author	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross cultural validity	Criterion validity	Responsiveness
23. A Virtual Reality Kitchen Task	Zhang et al. (2003) [46]		Fair/+ ICC 0.76						Fair/– Pearsons correlation 0.62 compared to actual kitchen performance	
24. The VRST	Canty et al. (2014) [51]						Fair/+			Poor/+

Results of each instrument based on rating from the COSMIN checklist and Terwee quality criteria. Methodological quality of each study on each COSMIN measurement property is indicated by poor, fair good or excellent. The Terwee quality criteria of each measurement property is indicated by +, –, ? or N/A; +: positive; –: negative; ?: indeterminate; NA: not applicable. Statistical information are applied where this is of importance for the rating on the Terwee quality criteria.
*Positive Pearsons correlation compared to DEX (0.535) and Go-NO-GO (0.564), negative Pearsons correlation compared to MAAS (0.469) and SLS (–0.047).
**Positive in items of attention, memory, organisation, problem solving and judgement scores related to FIM cognitive scale (Spearmans correlation between –0.795; –0.702), negative in other item scores related to FIM and item scores related to NBRS-R (Spearmans correlation < 0.70).
***Positive in total FIM + FAM and cognitive subscale, negative in separate analysis of FIM cognitive subscale.
**** Positive ICC in Planning and Organisation (1.00), Time Management (1.00), Executive Memory (0.78), Error Detection (0.77), Impulsivity (0.73) and Sustained and Directed Attention (0.71), negative ICC in Error Correction (0.35), Cognitive Shifting (0.23) and Initiation (0.00).
*****Only mental domains according to ICF are included in the score.
******Negative ICC on Interrater reliability (0.60). Intrarater-reliability and test-retest score = N/A.
*******Positive Pearsons correlation in two item scores: Correct/accuracy (0.825) and absence (0.855). Negative Pearsons correlation in other item scores (0.725; 0.341).

due to poor study quality included internal consistency of the FCT, the MPAI-4 and the NCSE and reliability of the BMET, the BDRS, a Cognitive Test battery (FAM/DRS/LCFS, DRS, EFRT), the Measure of Executive Functioning, an Occupation-based Online Awareness Assessment, the PRPP and the VIP.

One study on the interrater reliability of the combined FIM + FAM reported both positive and negative limited ratings depending on the instrument's subscales. Positive ratings were related to the total FIM + FAM including the cognitive subscale, but with negative ratings on a separate analysis of the cognitive subscale of FIM [54] (Table 2).

Validity

Quality and results. Most studies reporting measurement properties were related to validity ($n = 23$). Of these, ten studies reported on criterion validity, eight studies on hypothesis testing, three studies on content validity and one study reported on cross-cultural validity. None of the included studies reported on structural validity (Table 2). Study quality of each measurement property ranged from poor ($n = 5$) to fair ($n = 18$) (Table 2). Reasons for downgrading study validity were small sample sizes, missing data about the target group (e.g. age or gender), no description of how missing data was handled, vague hypotheses, or because it was doubtful whether an appropriate gold standard was selected (Table 2).

Evidence synthesis. There was limited evidence for positive validity of the Hotel Task, the VRST and the VIP and for negative validity of the TOFEA, the NCSE, the VLT and a Virtual Reality Kitchen Task. Criterion validity of the FCT included both limited positive and negative results, depending on the comparator instrument. Fifteen studies on instrument validity were rated as unknown evidence because of poor study quality, one because of inconsistent results (Table 3). Instruments of unknown rating included the BDRS for both content validity and hypotheses testing, the DRS, the EFRT, the FAM, the FCI, the combined FIM + FAM, TOFEA, the Hotel Task, the LCFS for both hypotheses testing and criterion validity, the MPAI-4, the Measure of Executive Functioning, the NAB-DL and the SBS-COT.

Responsiveness

Quality and results. Two studies, one using a cognitive test battery (FAM/DRS/LCFS) and one using the VRST, reported on responsiveness. They were both

Table 3. Best evidence synthesis of measurement properties.

Instrument	Study author	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross cultural validity	Criterion validity	Responsiveness
1. The BMET	Dawson et al. (2009) [32]		?							
2. BDRS	McKeon et al. (2017) [33]		?		?		?			
3. A cognitive test battery; FAM, DRS, LCFS	van Baalen et al. (2006) [57]		?							?
4. DRS	Malec et al. (2012) [34]; Rappaport et al. (1982) [35]	+	?						?	
5. The EFRT	Boyd et al. (1993) [36]		?						?	
6. The FAM	Tesio et al. (1998) [58]						?			
7. The FCI	Labi et al. (1998) [37]		+						?	
8. The FCT	LeBlanc et al. (2012) [38]	?							+ and –*	
9. FIM + FAM, combined	Hawley et al. (1999) [52]; Donaghy et al. (1998) [54]	++	+/–				?			
10. FIM	Pretz et al. (2016) [39]	+								
11. The TOEFA	Bamdad et al. (2003) [40]						?		–	
12. The Hotel Task	Cardoso et al. (2015) [53]				+			?		
13. LCFS	Gouvier et al. (1987) [59]	+					?		?	
14. The MPAI-4	Kean et al. (2011) [41]; Murrey et al. (2006) [42]	?					?			
15. Measure of Executive Functioning	Lewis et al (2017) [55]		?						?	
16. NCSE	Wallace et al. (2000) [43]; Doninger et al. (2000) [47]	?							–	
17. NAB-DL	Zgaljardic et al. (2011) [44]								?	
18. An Occupation-based Online Awareness Assessment	Doig et al. (2017) [48]		?							
19. The PRPP	Nott MT et al. (2009) [49]		?							
20. SBS-COT	Martínez-Pernía et al. (2017) [56]				?					
21. The VLT	Renison et al. (2012) [50]		+				+***		–****	
22. The VIP	O'Neil-Pirozzi et al. (2010) [45]		?							
23. A Virtual Reality Kitchen Task	Zhang et al. (2003) [46]		+						–	
24. The VRST	Canty et al. (2014) [51]						+			?

+++ or – = strong evidence. ++ or – = moderate evidence. + or – = limited evidence. ±: conflict findings? = studies of poor methodologic quality or studies with indeterminate property quality.
*Positive in item scores related to FIM, negative in item scores related to NBRS-R.
** Only mental domains according to ICF are included in the score.
***Positive in four item scores: Strategy generation and regulation, prospective working memory, response Inhibition and event-based prospective memory.
****Positive in one item score: Strategy generation and regulation.

rated to have poor quality [43,49]. The study using a cognitive test battery (FAM/DRS/LCFS) was downgraded in quality because of a small sample size and a poorly described method section, where information about the comparator instrument was lacking [57]. The study using the VRST was downgraded due to not being a longitudinal design and not including a description of the time interval [51].

Evidence synthesis. Because of the poor quality, the evidence level for both studied instruments, one using a cognitive test battery (FAM/DRS/LCFS) and one using the VRST and the VRST, were rated as unknown (Table 3).

Discussion

Impaired mental function following TBI has a negative impact on the performance of ADL [3,7]. Occupational therapists use various instruments to assess mental function during activity and participation and use this information to plan relevant rehabilitation strategies. Yet, there is no consensuses on which instruments are the most appropriate to assess mental function during activity and participation.

The aim of this systematic review was to assess which instruments are best to support occupational therapy interventions. We did not know at the start that we would identify only few studies on each instrument and poor quality evidence. However, this is an important finding as these instruments are commonly used in occupational therapy practice. Ranking of the instruments based on the measurement properties would be useful. However, based on the results of the review, ranking does not seem opportune. This review showed best evidence for the combined FIM + FAM instrument with moderate evidence on internal consistency. Separately, the two instruments showed limited evidence on internal consistency and unknown evidence on hypotheses testing, respectively.

Most studies showed limited or unknown evidence. The most frequent reason for reducing quality was the risk of selection bias. Risk of selection bias was caused by small sample sizes, missing data about the target group (e.g. age or gender), a poorly described method section (e.g. no description of the management of missing responses), or only including patients if data were complete. Sample sizes can be a challenge in studies targeting TBI because of narrow inclusion criteria without considering the heterogeneity of TBI individuals. It has been suggested to accept the heterogeneity and use less restrictive enrolment criteria to achieve adequate sample sizes, which may increase generalizability and study quality [60].

Even though the search in this systematic review was performed from the establishment of the included databases, only few studies reported on the development of an instrument, i.e. content validity. This is surprising, since content validity is considered an important measurement property encompassing the relevance, comprehensiveness and comprehensibility of the target population [61]. More instruments were originally developed for the broader term of acquired brain injury, and information about content validity may have been excluded in this review due to the lack of separate TBI sub-analyses. For instance, the FIM has shown good face validity in inpatient rehabilitation without distinguishing between diagnostic groups [62]. However, face validity of the FIM among TBI individuals has been questioned due to low sensitivity [63]. Limitations of the mental functions in FIM have also been suggested in a review on older adults, where scores of executive functions showed significant floor and ceiling effects [64].

Not distinguishing between diagnostic groups eliminates the inclusion of several measurement properties of instruments included in this review. Instruments like the FIM, the FAM and the MPAI-4 have shown a satisfactory reliability and validity among people with acquired brain injury, but with no separate sub-analysis of TBI [65–67]. It is possible that if a TBI sub-analysis had been available in these studies, it would have impacted positively or negatively on the quality of the instruments in this review.

Although there are many instruments to assess mental function, only few are performance-based, which complicates the definition of a proper gold standard when measuring the construct of an instrument. Not all included studies encompass a total mental function score; sub-scores exist, e.g. in the MPAI-4 where a total ability score includes domains of mental function and mobility, but these subdomains are also scored separately.

Methodological considerations

A strength of this review is that it was completed according to the COSMIN checklist, which includes a tested and validated manual and search filters to enhance transparency [25,28]. The use of a standardised evaluation of measurement properties is important to reach agreement on the quality of studies.

The search strategy in this systematic review included extensive search terms to ensure a comprehensive and exhaustive search in several electronic databases. However, there is a risk of missing studies during the selection procedure, since the selection procedure was done manually. To reduce study selection bias, two review authors performed the selection procedure independently and blinded to each other's scores.

The current systematic review aimed to include an analysis of adult TBI individuals, although four studies included young adults. In the protocol of the systematic review, we planned to include adult TBI individuals. However, in the final review we included four studies that had included participants from a younger age than 18. By including these studies in the review, we did not completely comply with our inclusion criteria. However, the mean age and SD in the studies show that there were only very few people under the age of 18 included. It is unlikely that this affected the results.

A total of 28 articles with 40 studies of measurement properties of 24 performance-based instruments were identified. The limited number of studies on each instrument indicates lack of psychometric data to achieve excellent quality or to reveal possibly conflicting results. Hence, results of this review may not necessarily indicate poor measurement properties of the instruments reported rather than poor evidence.

The lack of information on the instruments impacts on our knowledge about an instrument's reliability, validity and responsiveness. Consequently, this review cannot make recommendations on which instrument to use. The incomplete picture of instruments to assess mental function during activity and participation among TBI individuals may impact on the interpretation of outcomes in clinical practice, since clinicians may select instruments based on incomplete psychometric evidence. Before choosing an instrument, it is important to be sure that the instrument measures what you intend to measure [68]. Hence, there is a need for more psychometric studies on existing instruments.

The use of standardised instruments has been reported time consuming and not routinely used among occupational therapists. However, the use of standardised instruments was preferable due to increased quality and consistency when assessing mental functions [69].

The review shows that there is no instrument that is clearly better than others, so recommending a specific instrument does not seem opportune. What does this mean for clinical practice? If evidence on measurement properties do not inform the choice of using an instrument, other factors may play a role. This could be availability of the instrument (in your language), time for performing the assessment, the impact on patients, being skilled in and feeling comfortable performing the assessment, feasibility of performing the test in a specific setting, etc. These aspects have been found to be barriers to a consistent use of standardised assessments in occupational therapy, beside the lack of measurement properties [70,71].

This systematics review provides a structured and comprehensive overview of measurement properties of available performance-based instruments to assess mental function during activity and participation in TBI individuals. Results reveal a lack of high-quality measurement property studies. Thus, evidence was insufficient to allow recommendation on choice of instrument and further investigations are needed. It is recommended to use tools with the highest possible evidence for positive ratings. Moreover, instruments with limited evidence should be used with caution in clinical practice and research. This overview contributes to such considerations before choosing an instrument.

Acknowledgements

The authors thank health science librarian Maria Oesterbye, who assisted the development of the detailed search strategy.

Disclosure statement

The authors of this review have no conflicts of interests. This systematic review did not require ethical approval.

Funding

The review was initiated and funded by the Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Denmark and Neurorehabilitation Skive, Hammel Neurorehabilitation Centre and University Research Clinic, Skive, Denmark.

References

[1] Peeters W, van den Brande R, Polinder S, et al. Epidemiology of traumatic brain injury in Europe. Acta Neurochir. 2015;157:1683–1696.

[2] Hyder AA, Wunderlich CA, Puvanachandra P, et al. The impact of traumatic brain injuries: a global perspective. NeuroRehabilitation. 2007;22:341–353.

[3] Sashika H, Takada K, Kikuchi N. Rehabilitation needs and participation restriction in patients with cognitive disorder in the chronic phase of traumatic brain injury. Medicine. 2017;96:e5968.
[4] Cicerone KD, Dahlberg C, Malec JF, et al. Evidence-based cognitive rehabilitation: updated review of the literature from 1998 through 2002. Arch Phys Med Rehabil. 2005;86:1681–1692.
[5] World Health Organization. ICF International classification of functioning, disability and health. Geneva, Switzerland: WHO; 2001.
[6] Cassidy JD, Carroll LJ, Peloso PM, et al. Incidence, risk factors and prevention of mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. J Rehabil Med. 2004;36:28–60.
[7] Park HY, Maitra K, Martinez KM. The effect of occupation-based cognitive rehabilitation for traumatic brain injury: a meta-analysis of randomized controlled trials. Occup Ther Int. 2015;22:104–116.
[8] Himanen L, Portin R, Isoniemi H, et al. Longitudinal cognitive changes in traumatic brain injury: a 30-year follow-up study. Neurology 2006; 66:187–192.
[9] Chung CS, Pollock A, Campbell T, et al. Cognitive rehabilitation for executive dysfunction in adults with stroke or other adult non-progressive acquired brain damage. Cochrane Database Syst Rev. 2013;30: CD008391.
[10] Werff KR. The application of the international classification of functioning, disability and health to functional auditory consequences of mild traumatic brain injury. Semin Hear. 2016;37:216–232.
[11] Tate RL, Godbee K, Sigmundsdottir L. A systematic review of assessment tools for adults used in traumatic brain injury research and their relationship to the ICF. NeuroRehabilitation 2013;32:729–750.
[12] Mohammed Alotaibi N, Reed K, Shaban Nadar M. Assessments used in occupational therapy practice: an exploratory study. Occup Ther Health Care. 2009;23:302–318.
[13] Nasreddine ZS, Phillips NA, Bédirian V, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53:695–699.
[14] Wilson B, Cockburn J, Halligan P. Development of a behavioral test of visuospatial neglect. Arch Phys Med Rehabil. 1987;68:98–101.
[15] Hanson CS, Shechtman O, Jackson Foss J, et al. Occupational therapy: current practice and training issues in the treatment of cognitive dysfunction. NeuroRehabilitation 1997;8:31–41.
[16] Poulin V, Korner-Bitensky N, Dawson DR. Stroke-specific executive function assessment: a literature review of performance-based tools. Aust Occup Ther J. 2013;60:3–19.
[17] Sansonetti D, Hoffmann T. Cognitive assessment across the continuum of care: the importance of occupational performance-based assessment for individuals post-stroke and traumatic brain injury. Aust Occup Ther J. 2013;60:334–342.
[18] Bickenbach J, Cieza A, Rauch A, Stucki G, editors. ICF core sets: manual for clinical practice. Göttingen: Hogrefe; 2012.
[19] Linacre JM, Heinemann JW, Wright BD, et al. The structure and stability of the functional independence measure. Arch Phys Med Rehabil. 1994;75: 127–132.
[20] Chapparo C. Perceive, recall, plan and perform (PRPP): occupation-centred task analysis and intervention system. In: Rodger S, editor. Occupation-centred practice with children. Chichester, UK: Wiley-Blackwell; 2010.
[21] Bayley MT, Tate R, Douglas JM, Turkstra LS, et al. INCOG guidelines for cognitive rehabilitation following traumatic brain injury: methods and overview. J Head Trauma Rehabil. 2014;29:290–306.
[22] Laxe S, Tschiesner U, Zasler N, et al. What domains of the International Classification of Functioning, Disability and Health are covered by the most commonly used measurement instruments in traumatic brain injury research? Clin Neurol Neurosurg. 2012; 114:645–650.
[23] Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols [PRISMA-P] 2015: elaboration and explanation. BMJ. 2015;349:g7647.
[24] Muren MA, Kristensen LQ, Petersen AK, et al. Measurement properties of instruments to assess mental function during activity and participation in individuals surviving traumatic brain injury: a systematic review protocol. Scand J Occup Ther. 2019; 1–5. [E-pub ahead of print]
[25] Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22.
[26] Terwee CB, Jansma EP, Riphagen II, et al. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18: 1115–1123.
[27] AOTA. Role of Occupational therapy in Assessing Functional Cognition [internet]. [cited 2019 June 6]. Available from: https://www.aota.org/Advocacy-Policy/Federal-Reg-Affairs/Medicare/Guidance/role-OT-assessing-functional-cognition.aspx.
[28] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–745.
[29] Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21: 651–657.
[30] Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60:34–42.

[31] Furlan AD, Malmivaara A, Chou R, et al. Updated method guideline for systematic reviews in the Cochrane Back and Neck Group. Spine 2015;40: 1660–1673.

[32] Dawson DR, Anderson ND, Burgess P, et al. Further development of the multiple errands test: standardized scoring, reliability, and ecological validity for the Baycrest version. Arch Phys Med Rehabil. 2009; 90:S41–S51.

[33] McKeon A, Terhorst L, Skidmore E, et al. A novel tool for naturalistic assessment of behavioural dysregulation after traumatic brain injury: a pilot study. Brain Inj. 2017;31:1781–1790.

[34] Malec JF, Hammond FM, Giacino JT, et al. Structured interview to improve the reliability and psychometric integrity of the disability rating scale. Arch Phys Med Rehabil. 2012;93:1603–1608.

[35] Rappaport M, Hall KM, Hopkins K, et al. Disability rating scale for severe head trauma: coma to community. Arch Phys Med Rehabil. 1982;63:118–123.

[36] Boyd TM, Sautter SW. Route-finding: a measure of everyday executive functioning in the head-injured adult. Appl Cognit Psychol. 1993;7:171–181.

[37] Labi MLC, Brentjens M, Shaffer K, et al. Functional cognition index©: a new instrument to assess cognitive disability after traumatic brain injury. Neurorehabil Neural Repair. 1998;12:45–51.

[38] LeBlanc J, de Guise E, Lamoureux J, et al. Criterion validity of a functional cognitive task in patients with severe traumatic brain injury. Brain Inj. 2012; 26:1143–1154.

[39] Pretz CR, Kean J, Heinemann AW, et al. A multidimensional rasch analysis of the functional independence measure based on the national institute on disability, independent living, and rehabilitation research traumatic brain injury model systems national database. J Neurotrauma. 2016;33: 1358–1362.

[40] Bamdad MJ, Ryan LM, Warden DL. Functional assessment of executive abilities following traumatic brain injury. Brain Inj. 2003;17:1011–1020.

[41] Kean J, Malec JF, Altman IM, et al. Rasch measurement analysis of the Mayo-Portland adaptability inventory (MPAI-4) in a community-based rehabilitation sample. J Neurotrauma. 2011;28:745–753.

[42] Murrey GJ, Hale FM, Williams JD. Assessment of anosognosia in persons with frontal lobe damage: clinical utility of the Mayo-Portland adaptability inventory [MPAI]. Brain Inj. 2005;19:599–603.

[43] Wallace JJ, Caroselli JS, Scheibel RS, et al. Predictive validity of the Neurobehavioural Cognitive Status Examination (NCSE) in a post-acute rehabilitation setting. Brain Inj. 2000;14(1):63–69.

[44] Zgaljardic DJ, Yancy S, Temple RO, et al. Ecological validity of the screening module and the daily living tests of the neuropsychological assessment battery using the Mayo-Portland adaptability inventory-4 in postacute brain injury rehabilitation. Rehabil Psychol. 2011;56:359–365.

[45] O'Neil-Pirozzi TM, Goldstein R, Strangman GE, et al. Test-re-test reliability of the virtual planning test in individuals with traumatic brain injury. Brain Inj. 2010;24:509–516.

[46] Zhang L, Abreu BC, Seale GS, et al. A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: reliability and validity. Arch Phys Med Rehabil. 2003;84:1118–1124.

[47] Doninger NA, Bode RK, Heinemann AW, et al. Rating scale analysis of the neurobehavioral cognitive status examination. J Head Trauma Rehabil. 2000;15:683–695.

[48] Doig E, Fleming J, Ownsworth T, et al. An occupation-based, metacognitive approach to assessing error performance and online awareness. Aust Occup Ther J. 2017;64:137–148.

[49] Nott MT, Chapparo C, Heard R. Reliability of the perceive, recall, plan and perform system of task analysis: a criterion-referenced assessment. Aust Occup Ther J. 2009;56:307–314.

[50] Renison B, Ponsford J, Testa R, et al. The ecological and construct validity of a newly developed measure of executive function: the virtual library task. J Int Neuropsychol Soc. 2012;18:440–450.

[51] Canty AL, Fleming J, Patterson F, et al. Evaluation of a virtual reality prospective memory task for use with individuals with severe traumatic brain injury. Neuropsychol Rehabil. 2014;24:238–265.

[52] Hawley CA, Taylor R, Hellawell DJ, et al. Use of the functional assessment measure [FIM + FAM] in head injury rehabilitation: a psychometric analysis. J Neurol Neurosurg Psychiatry. 1999;67:749–754.

[53] Cardoso CO, Zimmermann N, Parana CB, et al. Brazilian adaptation of the hotel task: a tool for the ecological assessment of executive functions. Dement Neuropsychol. 2015;9:156–164.

[54] Donaghy S, Wass PJ. Interrater reliability of the functional assessment measure in a brain injury rehabilitation program. Arch Phys Med Rehabil. 1998;79:1231–1236.

[55] Lewis MW, Babbage DR, Leathem JM. Systematic behavioural observation of executive performance after brain injury. Brain Inj. 2017;31:639–648.

[56] Martinez-Pernia D, Nunez-Huasaf J, Del Blanco A, et al. Using game authoring platforms to develop screen-based simulated functional assessments in persons with executive dysfunction following traumatic brain injury. J Biomed Inform. 2017;74:71–84.

[57] van Baalen B, Odding E, van Woensel MP, et al. Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. Clin Rehabil. 2006;20:686–700.

[58] Tesio L, Cantagallo A. The functional assessment measure (FAM) in closed traumatic brain injury outpatients: a Rasch-based psychometric study. J Outcome Meas. 1998;2:79–96.

[59] Gouvier WD, Blanton PD, LaPorte KK, et al. Reliability and validity of the disability rating scale and the levels of cognitive functioning scale in monitoring recovery from severe head injury. Arch Phys Med Rehabil. 1987;68:94–97.

[60] Roozenbeek B, Lingsma HF, Maas AI. New considerations in the design of clinical trials for traumatic brain injury. Clin Invest. 2012;2:153–162.

[61] Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27:1171–1179.

[62] Granger CV, Hamilton BB, Keith RA, et al. Advances in functional assessment for medical rehabilitation. Top Geriatr Rehabil. 1986;1:59–74.

[63] Hall KM, Bushnik T, Lakisic-Kazazic B, et al. Assessing traumatic brain injury outcome measures for long-term follow-up of community-based individuals. Arch Phys Med Rehabil. 200;82:367–374.

[64] Wales K, Clemson L, Lannin N, et al. Functional assessments used by occupational therapists with older adults at risk of activity and participation limitations: a systematic review. PLoS One. 2016;11: e0147980.

[65] Glenny C, Stolee P. Comparing the functional independence measure and the interRAI/MDS for use in the functional assessment of older adults: a review of the literature. BMC Geriatr. 2009;9:52-2318-9-52.

[66] Hobart JC, Lamping DL, Freeman JA, et al. Evidence-based measurement: which disability scale for neurologic rehabilitation? Neurology 2001;57: 639–644.

[67] Malec JF. Comparability of Mayo-Portland adaptability inventory ratings by staff, significant others and people with acquired brain injury. Brain Inj. 2004;18:563–575.

[68] Duncan PW. Outcome measures in stroke rehabilitation. Handb Clin Neurol. 2013;110:105–111.

[69] Wesson J, Clemson L, Brodaty H, et al. Estimating functional cognition in older adults using observational assessments of task performance in complex everyday activities: a systematic review and evaluation of measurement properties. Neurosci Biobehav Rev. 2016;68:335–360.

[70] Bowman J. Challenges to measuring outcomes in occupational therapy: a qualitative focus group study. Br J Occup Ther. 2006;69:464–472.

[71] Stapleton T, McBrearty C. Use of standardised assessments and outcome measures among a sample of Irish occupational therapists working with adults with physical disabilities. Br J Occup Ther. 2009;72: 55–64.