aci.health.nsw.gov.au

Development and Validation of the Brief Executive Function Assessment Tool (BEAT)

Alcohol and Drug Cognitive Enhancement (ACE) Program

February 2021

Drug and Alcohol Network

NSW GOVERNMENT
AGENCY FOR CLINICAL INNOVATION
ANTS ADVANCED NEUROPSYCHOLOGICAL TREATMENT SERVICES Knowledge, Compassion, Excellence
WHOS® helping people help themselves

The Agency for Clinical Innovation (ACI) is the lead agency for innovation in clinical care.

We bring consumers, clinicians and healthcare managers together to support the design, assessment and implementation of clinical innovations across the NSW public health system to change the way that care is delivered.

The ACI's clinical networks, institutes and taskforces are chaired by senior clinicians and consumers who have a keen interest and track record in innovative clinical care.

We also work closely with the Ministry of Health and the four other pillars of NSW Health to pilot, scale and spread solutions to healthcare system-wide challenges. We seek to improve the care and outcomes for patients by re-designing and transforming the NSW public health system.

Our innovations are:

- person-centred
- clinically-led
- evidence-based
- value-driven.

www.aci.health.nsw.gov.au

AGENCY FOR CLINICAL INNOVATION

1 Reserve Road St Leonards NSW 2065

Locked Bag 2030, St Leonards NSW 1590
T +61 2 9464 4666
E aci-info@nsw.gov.au | aci.health.nsw.gov.au

SHPN (ACI) 200818 | ISBN 978-1-76081-546-2

Produced by: Drug and Alcohol Network and Advanced Neuropsychological Treatment Services

Further copies of this publication can be obtained from
the Agency for Clinical Innovation website at www.aci.health.nsw.gov.au

Version: 1 **Trim:** ACI/D21/396

Date amended: February 2021

Published Mar 2018. Next Review 2028.

Suggested citation:

Berry J, Nardo T, Sedwell A, Lunn J, Marceau E, Wesseling A, Zucco M, Sugden-Lingard S, Borchard T, Batchelor J, Shores, E. A. Development and validation of the Brief Executive function Assessment Tool. Sydney: Agency for Clinical Innovation; 2021.

Contents

Background and aims

Cognitive impairment is common in substance use disorder (SUD) populations and is a predictor of drop out from SUD treatment. [1,2] However, the assessment of cognitive impairment within this population has mainly been limited to research, for example Marceu et al. (2016), rather than as a routine clinical process.[3] There is an increasing focus on implementing assessment and intervention practices informed by neuropsychology and neuroscience-based approaches in SUD treatments.[4]

Screening for cognitive impairment in alcohol and other drug (AOD) services would help to identify individuals who may need support to obtain the greatest benefit from SUD treatment. It is anticipated that early identification of cognitive impairment within this population and subsequent interventions to improve or compensate for the deficits would result in longer time in treatment and therefore more successful outcomes.[5]

Although the *Montreal Cognitive Assessment* (MoCA) was originally developed to detect mild cognitive impairment in older individuals at risk of dementia, it has been validated for use with younger individuals in SUD treatment.[6,7,8] The MoCA has shown good sensitivity and specificity, and therefore has great promise as a cognitive screening tool within AOD services.[9,10] Furthermore, it has a short administration time of 10 to 20 minutes. However, as the MoCA was not developed specifically to detect cognitive impairment in an SUD population, some items may be redundant and it may not test all cognitive functions that are impaired in SUD.

The goal of the current project was to develop a cognitive screening measure that:

- samples a broad range of executive functions known to be affected in individuals with SUD
- expands on the cognitive abilities, sampled in other screening tools, such as the MoCA (particularly executive function)
- is sensitive to change following neuropsychological intervention
- has robust psychometric properties, including good test-retest reliability, inter-rater reliability, construct validity, criterion validity and classification statistics.

Test-retest reliability relies on items tapping into constructs that are not expected to vary over time, such as latent cognitive constructs. Inter-rater reliability requires clear and easy to implement administration and scoring procedures. Construct validity requires the tool to demonstrate convergence with existing tools that purport to measure similar constructs. Criterion validity requires the tool to distinguish between different populations with and without the condition of interest. Classification statistics such as sensitivity and specificity relate to the tool's ability to detect abnormality when present and to detect abnormality only in the population of interest.

Because executive functioning represents the domain of greatest cognitive impairment in an SUD population, the tool under development focused on executive functions.[11,12] Hence, it was named the *Brief Executive Function Assessment Tool (BEAT)*.

Development of the BEAT

The BEAT was developed over three years and multiple iterations across three pilot versions, denoted v1, v2 and v3. With each new version of the tool, items that were insensitive to cognitive impairment in SUD samples or failed to discriminate between SUD and normal control samples were discarded and new items were added based on ongoing literature review and consultation. Below is a description of the methodology adopted to develop the BEAT, including details about item inclusion, modification or removal.

BEAT v1

The MoCA is more sensitive to executive function impairment than similar screening measures, such as the *Folstein Mini-mental State Examination*.[13] As such, it presented a good starting basis for the development of the BEAT. The *Behavior Rating Inventory of Executive Functioning, Adult version (BRIEF-A)* is a self-report inventory that has been shown to be more sensitive to executive function deficits than performance-based tests.[1114] As such, it also presented a good starting point for investigating self-report items that may help identify cognitive impairment in an SUD population.

The first step in developing BEAT v1 was to analyse group differences across items on the MoCA and the BRIEF-A. Initial analysis comprised a comparison of mean scores for MoCA items across an SUD and normal control sample. The samples comprised n=128 individuals with SUD in residential treatment at a female-only residential SUD treatment facility, We Help Ourselves (WHOS)), and n=37 normal controls. For full sample demographics, see Marceau, et al. (2016).[3]

Table 1. MoCA item means for the SUD and normal control groups as well as mean differences, from Marceau et al. (2016)[3]

Item	Normal control (n=37) mean	Substance use disorder (n=128) mean	Mean difference
Trails	0.87	0.71	0.16
Cube	0.71	0.45	0.26
Clock	2.56	2.14	0.42
Naming	2.88	2.95	-0.07
Memory	3.73	3.52	0.21
Digit span	1.82	1.79	0.03
Letter A	0.97	0.92	0.05
Serial 7	2.89	2.67	0.22
Sentence rep	1.83	1.51	0.32
Fluency	0.87	0.57	0.3
Abstraction	1.83	1.19	0.64
Orientation	5.99	5.89	0.1

The results presented in Table 1 show the greatest mean difference was detected for the trails, cube copy, clock drawing, memory, serial 7s, sentence repetition, fluency and abstraction items on the MoCA. There was less difference between the groups in naming, digit span, letter tapping and orientation items.

Using data collected from a sample of n=33 residents of WHOS (see Marceau et al., 2017 for sample characteristics), BRIEF-A responses were analysed at an item level to identify items that were most frequently endorsed and most sensitive to change following cognitive remediation.[15] See Table 2 for the identified items. These items were used to develop a questionnaire item for the BEAT.

Table 2. Most sensitive BRIEF-A items in a SUD sample

BRIEF-A item	Problem area
4	Concentration difficulties
15	Prioritisation difficulties
22	Inflexibility or problem-solving difficulties
26	Topic maintenance difficulties
28	Emotional reactivity
47	Planning difficulties
49	Initiation difficulties
52	Task finalisation difficulties
54	Task finalisation difficulties
63	Planning difficulties
69	Frequent mood changes
71	Organisation difficulties
72	Emotional reactivity
75	Task finalisation difficulties

The main purpose of BEAT v1 was to establish feasibility (acceptability) on the part of both examiners and examinees.

General

In order to maximise standardisation of administration, and hence increase inter-rater reliability, instructions for some items (such as word list learning, attention, working memory and number tapping) were included on the record form.

Research assistants were trained in the administration of all measures by trained neuropsychologists or neuropsychology students. To ensure proficiency in administration, research assistants administered the test battery to one of the trainers.

During the practice administration, the trainer tested the trainees to ensure the correct prompts were used and mock queries were answered in a way that did not exhibit coaching behaviours, which can impact test performance. The responses recorded on the response forms were also checked by trainers to ensure trainees had recorded them correctly. Feedback was provided throughout the training process to ensure any mistakes in administration were corrected.

Measures administered to research participants by recently trained research assistants were routinely checked for errors by a trained neuropsychologist or individuals who were experienced in the administration of the test battery. Research assistants involved in data entry had additional training in scoring the measures. The participant data files entered by newly trained research assistants were double scored to ensure data integrity.

BEAT v1 pilot sample

BEAT v1 data were collected in a sample of n=41 individuals with SUD attending residential treatment at WHOS. See Table 3 for sample characteristics. Score distributions were explored at the item level. Testing was feasible and acceptable to participants.

Table 3. Characteristics of the BEAT v1 pilot sample

Characteristic	Figures (n=41)	
Gender (%Male)	Female 41	0%
	Mean	Standard deviation
Age	33.4	10.0
Education	10.3	2.2
Test of Premorbid Functioning	89.0	10.0
Primary substance of use	Number	%
Alcohol	10	24%
Methamphetamine	25	61%
Other stimulants	1	2%
Heroin	4	10%
Other opiates or opioids	1	2%

Items for which more than 90% of the sample achieved full credit were considered redundant and excluded from BEAT v1. See Table 4 for descriptive statistics and percentage of the sample achieving full credit for each item.

Table 4. Descriptive statistics and percentage of the sample achieving full credit for each item

Item	Mean	Median	Standard deviation	Inter-quartile range	% Achieving full credit
1. Questionnaire	29.95	28	14.5	21	0%
2. Trail making	0.9	0	2.7	0	81%
3. Clock drawing	1.2	1	0.7	1	39%
4. Figure copy	0.8	1	0.4	0	78%
5. Word list learning	17.6	18	2.3	4	10%
6. Attention span	5.9	6	0.9	2	32%
7. Working memory	3.8	4	1	2	5%
8. Number tapping	1.1	1	1.1	2	34%
9. Abstraction	2.7	3	1.6	3	17%
10. Letter fluency	13.6	13	3.9	6	N/A
11. Category fluency	20.1	19	5.4	7	N/A
12. Orientation	5.9	6	0.3	0	93%
13. Reading	4.1	4	1	1	0%
14. Word list recall (free recall)	5	5	1.6	2	20%
15. Word list recall (memory index score)	18.2	19	2.5	3	20%
16. Visual memory	1.8	2	1.1	2	66%
17. Naming	14.4	15	0.9	1	61%
18. Visual search	0.8	0	1	2	37%

BEAT v1 Item 1: Questionnaire

One of the difficulties associated with detecting cognitive impairment in any population is the difference in results from performance-based and inventory-based tests. Although Hagen et al. found that the BRIEF-A (an inventory-based measure) was more sensitive in assessing executive functioning than performance-based tests in their SUD sample, the same research group has cautioned against the sole use of inventory-based measures, especially when there is comorbid psychological distress.[11,16] The sole use of more objective performance-based tests is less likely to be impacted by self-reporting biases. However, a positive result in the absence of self-reported cognitive difficulties is insufficient to meet the threshold for neuropsychological impairment as defined by the American Academy of Clinical Neuropsychology consensus conference, which

requires both impairment on performance-based tests and 'other findings related to functional capacity'.[17] Similarly, it is impossible for a person to meet criteria for *Diagnostic and Statistical Manual of Mental Disorders Manual of Mental Disorders* major neurocognitive disorder without evidence of an impact on everyday functioning.[18] Self-report inventories are a useful method for establishing functional impact.

Based on these factors, it was desirable to include a self-report component in the BEAT. Thus, item 1 of BEAT v1 comprised 10 questions based on the problem areas corresponding to the BRIEF-A items in Table 2. In order to reduce the subjectivity of a frequency response format (for example, 'How often did you…?'), the questions were worded in the form of 'On how many days in the past week did you…?' The period of one week was chosen in order to reduce error associated with poorer recall of behavioural phenomena over a longer period of time. Some items were combined into a single question. Specifically, items 4 and 26 of the BRIEF-A were collapsed into Question 1, items 52, 54 and 75 were collapsed into Question 5, and items 47 and 63 were collapsed into Question 6. Question 2 was based on BRIEF-A item 15, Question 3 on BRIEF-A item 28, Question 4 on BRIEF-A item 49, Question 7 on BRIEF-A item 69, Question 8 on BRIEF-A item 71, Question 9 on BRIEF-A item 72 and Question 10 on BRIEF-A item 22. See the 10 BEAT Questionnaire items in Table 5.

Table 5. BEAT item 1: questionnaire items

Questions	
1	On how many days in the past week did you have trouble concentrating or staying on the same topic during a conversation?
2	On how many days in the past week did you have trouble prioritising what you had to do?
3	On how many days in the past week did you have a problem because you reacted emotionally to a situation?
4	On how many days in the past week did you have trouble starting something you wanted to do?
5	On how many days in the past week did you have trouble completing something you wanted to achieve?
6	On how many days in the past week did you fail to plan properly for something you wanted to do?
7	On how many days in the past week did your mood dramatically or frequently change?
8	On how many days in the past week did you not have the things you needed with you?
9	On how many days in the past week did you get upset easily at something little?
10	On how many days in the past week did you have difficulties with problem-solving or decision-making?

The questions were read to the examinee, and the examiner marked the response on the form. The number of days was summed across the 10 questions. The maximum total score was 70.

BEAT v1 Item 2: Trail making

This item was modelled on the Trail Making Test – Part B, which requires a person to form a trail, alternating between numbers and letters, in ascending order.[19] The original version contains numbers 1 to 13 and letters A to L. The MoCA includes an abbreviated trail making item, with numbers to 5 and letters to E. Because the difference between the normal control and SUD groups in the pilot data was small for the MoCA trail making item, it was desirable introduce more stimuli to increase the cognitive demand on examinees. Thus, an item with numbers to 9 and letters to I was developed. Similar to the MoCA, the number of errors was the outcome variable.

The instructions were: '*I'd like you to start at number 1 and draw a line from 1 to A, A to 2, 2 to B and so on. Keep alternating between number and letter until you reach the end here (point to I).*'

BEAT v1 Item 3: Clock drawing

In an attempt to reduce the subjectivity and therefore variability in scoring the reproduction of a clock, a pre-drawn circle representing the clock face was provided as per Shulman et al. (1986).[20] Because an examinee's approach to drawing the clock (that is, the order of number placement) has been shown to be an index of the executive function of planning, number ordering was the only outcome of interest for this item.[21] Such simplifications of approaches to scoring clock reproductions has been advocated, and is expected to result in higher rates of inter-rater reliability.[22]

The instructions were: '*This circle represents a clock. Put in all the numbers so that it looks like a clock and then set the time to ten minutes past eleven.*'

The order in which the numbers were written was recorded. The following variables were explored to code for planning on the clock drawing item.

1. First number an anchor (that is, 12, 3, 6 or 9)
2. First two numbers an anchor
3. First four numbers anchors.

Variables 1 and 3 resulted in too few and too many errors respectively. A middle error rate was observed with variable 2, so it was chosen as the way to code responses on this item. Specifically, the coding was:

2 = each of the first two numbers were anchors (12, 3, 6 or 9)

1 = one of the first two numbers was an anchor

0 = neither of the first two numbers were anchors.

BEAT v1 Item 4: Figure copying

The cube copy in the MoCA may prejudice those with limited schooling or exposure to geometry education. As such, a novel design was chosen to assess copying. As was the case with clock drawing, the response data type of interest was the approach taken to copy the figure. Hudson et al. provided simple geometric figures with data regarding the approach for young adults with Williams Syndrome and typically developing children.[23] One of their figures was chosen for the BEAT (see Figure 1), where the commonest approach to copying the design was to start with the outer figure (the diamond) and proceed towards the centre, with the centre-most element (the rectangle) being drawn last. The data were coded as binary, where 1 = started with the outer diamond and 0 = started with another element.

Instructions were: '*Copy this diagram as accurately as you can over here (point to the space to the right of the design).*'

Following the figure copy, examinees were instructed: '*I would like you to remember that last drawing you copied because I will ask you to draw it again later on.*'

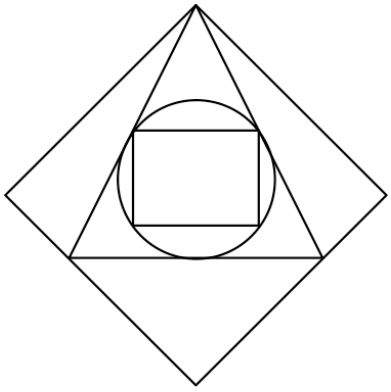

Figure 1. Item 4: Figure copy design

BEAT v1 Item 5: Word list learning

The list of words was increased to seven, compared to five in the MoCA, to better differentiate immediate auditory attention span (an aspect of working memory) from memory encoding. The two words 'lettuce' and 'nurse' were added to the five MoCA words. They were chosen because they are common nouns that belong to different categories, vegetable and profession, respectively, which was deemed important for cueing procedures (see Word List Recall section below). The number of learning trials was increased from two to three. The response variable was the total number of words recalled across the three trials. Maximum total score was 21.

The instructions were: '*Now I'm going to say some words. Listen carefully, because after I've said them, I want you to tell me as many of them as you can remember, in any order you like. Ready?*'

(Read the words at the rate of one per second.)

After trials 1 and 2, say: '*I'm going to say all of the words again and when I stop, I want you to tell me all of the words from the list, including the ones you've already told me.*'

BEAT v1 Item 6: Attention span

The MoCA requires only five digits to be repeated in forward order to attract one point of credit for the first Attention item. However, this represents performance at the fourth percentile for the longest forward digit span according to the Wechsler Adult Intelligence Scale, fourth edition (WAIS-IV) normative sample across all ages.[24] Given the younger age of individuals seeking treatment for SUD, it was considered prudent to assess their forward digit span up to a maximum of seven digits forward, which was the average longest forward digit span in the WAIS-IV normative sample across all ages.[24] For the sake of brevity, only one trial at each digit sequence length was given. If the examinee responded incorrectly to the first item (five digits), then the four-digit item was given next. If the examinee responded correctly to the first item, then the six-digit item was given next. The examiner proceeded in this fashion until an error was made or the seven-digit item was repeated correctly in the case of a good performance or until a correct response was achieved or the three digit item was administered in the case of a poor performance.

Instructions were: '*I'll say some numbers and I want you to repeat them back to me in the same order. For example, if I say 3-5, what would you say?*' (Start at five items then go up and down based on result until an error occurs then stop).

BEAT v1 Item 7: Working memory

The MoCA requires only three digits to be repeated in backward order to attract one point of credit for the first Working Memory item. However, this represents performance at the fourth percentile for the longest backward digit span according to the WAIS-IV normative sample across all ages.[24] Similarly to the Attention Span item, it was considered prudent to assess backward digit span up to a maximum of six digits backwards, which was at the seventy-fourth percentile in the WAIS-IV normative sample across all ages.[24] For the sake of brevity, only one trial at each digit sequence length was given, in the same manner as for Attention Span. If the examinee responded incorrectly to the first item (three digits), then the two-digit item was given next. If the examinee responded correctly to the first item, then the four-digit item was given next. The examiner proceeded in this fashion until an error was made or the six-digit item was repeated correctly, in the case of a good performance, or until a correct response was achieved or the two-digit item was administered, in the case of a poor performance.

Instructions were: '*I'll say some more numbers and this time I want you to say them in reverse order. For example, if I say 7-2, what would you say?*' (Start at three items then go up and down based on result until an error occurs then stop).

BEAT v1 Item 8: Number tapping

A number tapping test, modelled on the Sustained Attention to Response Task was devised.[25] The Sustained Attention to Response Task has been shown to be sensitive to impairments in those with attention deficit hyperactivity disorder and right hemisphere attentional difficulties.[26,27,28] It assesses sustained attention and inhibitory control. The item developed for the BEAT required a person to clap or tap on the tabletop after every number that was read (at the rate of one number per 1-2 seconds), except for the number 1. There were 100 numbers read out in total, including 10 instances of the number 1. Examiners were to note the number of errors (instances when the person tapped incorrectly after 1), for a maximum error score of 10.

Instructions were: '*I'm going to say some numbers and I want you to clap (or tap) after each number except the number one. Ready?*' (Read at the rate of one number per 1-2 seconds).

BEAT v1 Item 9: Abstraction

Two items (salt-pepper and salt-seaweed) were added to the MoCA items, bringing the item total to four. The two final items included the word salt, which required examinees to shift their cognitive set to answer the final item correctly. This change was made to broaden the skills assessed by this item to include cognitive flexibility, in addition to abstraction. If these final two items were correct, then a bonus point was credited, bringing the maximum total score to five.

Instructions were: *'In what way are a banana and orange alike?'* (Note the response and if anything other than 'fruit', say, *'Yes, and they're both fruit')*.

'In what way are a train and bicycle alike?' (etc.)

(Record responses verbatim. Provide no further prompts).

BEAT v1 Item 10: Letter fluency

Letter fluency was retained, as per the MoCA.

The instructions were: '*I'd like you to tell me as many words as you can think of that begin with a particular letter of the alphabet. Only give me words that start with that letter, except for the names of people or places. Also, don't give me the same word with different endings, like love, lover, loving, etc. Tell me as many words as you can think of that start with F.*'

BEAT v1 Item 11: Category fluency

A category fluency item was added to the BEAT. This was based on the finding that category fluency is easier for most examinees because they can rely on the automatic activation of semantically related words to complete the task.[29,30,31] In the letter fluency task, an examinee must rely more on executive skills such as strategy generation to complete the task efficiently. As such, the difference between letter and category fluency may provide a useful index of relative executive function difficulties.

Instructions were: '*Now tell me as many different names of animals as you can think of. Go ahead.*'

BEAT v1 Item 12: Orientation

The MoCA version of orientation was used. Total score was the number of correct items, with a maximum score of six.

Instructions were: '*What is the date/month/year/day of the week?*'

'*What's the name of this place?*'

'*What city are we in?*'

This item was excluded in v2 because 93% participants achieved full credit. This confirmed the finding from the MoCA item analysis, which showed that SUD participants are at or close to ceiling on assessment of orientation.

BEAT v1 Item 13: Reading

Reading correlates very highly with global cognitive constructs such as intelligence.[32,33] Seven items were chosen on the basis of pilot data from established reading tests including the National Adult Reading Test, Wechsler Test of Adult Reading, and Test Of Premorbid Functioning (TOPF).[34,35,24,36] Items were selected to represent a range of difficulty. The total score was the number of correctly read words, with a maximum value of seven.

Examinees were presented with a stimulus card containing the seven words in print and instructed: *'Read these words for me out loud, starting at number 1.'*

BEAT v1 Item 14: Word list recall

Both free recall and a more complex scoring scheme, modelled on the MoCA-MIS (memory index score) were used as response outcomes.[37] The latter involved summing the free recall score multiplied by three, the category cued recall score multiplied by two, and the multiple choice cued recall score. The maximum total score was 21.

Instructions were: '*A little while ago I read you some words several times. I want you to tell me all the words again in any order.*'

For words that are not freely recalled, provide a category cue, that is:

'*One of the words was a part of the body/type of material/type of building/type of flower/colour/ vegetable/profession.*'

For words that are not recalled with a category cue, provide the following multiple choice responses.

> *'Was it:*
> *a nose, face or hand?*
> *denim, cotton or velvet?*
> *a church, school or hospital?*
> *a rose, daisy or tulip?*
> *red, blue or green?*
> *a carrot, lettuce or cucumber?*
> *lawyer, doctor or nurse?'*

Both the free recall and MIS scoring schemes were retained for future inspection of discriminative properties.

BEAT v1 Item 15: Visual memory

Visual processing deficits, including visual memory deficits, are more common than verbal processing deficits in individuals with alcohol use disorders, which is related to reduced integrity of white matter fibre bundles connecting the frontal and occipital lobes.[38,393340]

Visual memory was assessed by asking the examinee to reproduce from memory the figure they copied earlier (Item 4). Scoring was based on the number of elements correctly reproduced and their placement, for a total score of seven (see scoring key below).

Instructions were: '*A little while ago, you copied a diagram. I want you to draw that diagram now as best as you can from memory*.'

Scoring key

Allocate one point for each of the following conditions.

- Diamond present
- Triangle present
- Circle present
- Rectangle present
- Triangle immediately inside of diamond
- Circle immediately inside of triangle
- Rectangle immediately inside of circle.

BEAT v1 Item 16: Naming

Pilot data indicated that almost all SUD participants received full credit on the MoCA Naming item, and that they outperformed the normal controls on this measure. Recent findings have suggested that Korsakoff patients have difficulties on object perception subtests from the Visual Object and Space Perception, particularly in naming silhouette drawings of common objects.[41,42] This type of naming task overlaps with visuo-spatial functioning, which is impaired relative to verbal functioning in individuals with alcohol use disorder, and therefore may be more sensitive to detecting impairment in an SUD population.[38,39,40]

The silhouette naming pretest from the Bells Test was used as a naming task, as it was required for the Visual Search item in any case.[43] The score was the total number of correctly named silhouette drawings (from a total of 15).

Instructions were: *'What's this a picture of?'* (Work your way through the pictures in the order they are recorded on the form).

BEAT v1 Item 17: Visual search

To further explore visuo-spatial abnormalities in SUD screening, a visual scanning task was included. One-third of the array from the Bells Test was used for this purpose, including 20 targets among 196 distractors.[43]

The instructions were: '*There are 20 bells in this array. I want you to find all 20 and number them as you find them. So, put a "1" next to the first one, "2" next to the second, and so on.*'

After connecting the bells in the order that they were identified, a subjective rating was made by the examiner as below.

2 = Clearly horizontal, starting at top of array

1 = Clearly vertical (starting on left or right of array) or horizontal starting at bottom of array

0 = Any other approach.

BEAT v2

BEAT v2 pilot sample

Data collected in a sample of n=29 individuals with SUD and n=94 normal controls (see Table 6 for sample characteristics) were examined by visual inspection of frequency histograms and consideration of descriptive statistics at the item level (see Table 7). It was presumed that items that were able to discriminate between SUD and normal control groups would likely be more sensitive to the cognitive impairments characteristic of the SUD population. Hence, establishing the known-groups validity at the item level was a primary aim of developing BEAT v2. Items were either discarded or modified to improve the tool's ability to discriminate between these populations. Other items were added based on an ongoing consultation process and literature review of potentially suitable items.

Table 6. Characteristics of BEAT v2 pilot sample

Characteristic	Substance use disorder (n=129)		Normal control (n=94)	
Gender (%Male)	64%		42%	
	Mean	Standard deviation	Mean	Standard deviation
Age	35.5	9.20	31.9	13.2
Education	11.1	1.8	13.3	2.1
Test of Premorbid Functioning	93.2	13.5	98.3	14.6
Primary substance of use	Number	%		
Alcohol	41	32		
Methamphetamine	35	27		
Other stimulants	3	2		
Heroin	30	23		
Other opiates or opioids	7	5		
Sedatives, hypnotics or tranquilisers	4	3		
Cannabis	9	7		

Table 7. Descriptive statistics for the SUD and normal control groups for BEAT v2 items

Item	Substance use disorder (n=129)				Normal control (n=94)			
	Mean	Median	Standard deviation	Inter-quartile range	Mean	Median	Standard deviation	Inter-quartile range
1. Questionnaire	23.1	19	17	21	15.1	13	10.1	14
2. Trail making	0.5	0	1.1	1	0.4	0	1.3	0
3. Clock drawing	1.4	1	0.6	1	1.6	2	0.5	1
4. Figure copy	0.7	1	0.5	1	0.7	1	0.5	1
5. Word list learning	17	17	2.5	4	17.5	18	2.2	3
6. Attention span	5.7	5	0.9	1	5.7	6	0.9	1
7. Working memory	3.9	4	1.1	2	4.3	4	1	1
8. Number tapping	1.1	1	1.5	2	1.5	1	1.6	2
9. Abstraction	2.6	3	1.1	1	2.7	3	1.1	1
10. Letter fluency	15.8	15	5.1	7	16.7	16	4.8	7
11. Category fluency	15	15	3.6	5	16	16	3.6	5
12. Reading	4.2	4	1.1	1	4.7	5	1.3	2
13. Word list recall free recall	4	4	1.8	2	4.9	5	1.5	2
14. Word list recall (MIS)	16.2	17	3.4	5	17.8	18	2.8	4
15. Visual memory	5.9	7	1.5	2	6.3	7	1.2	2
16. Naming	14.1	14	1	1	14.4	15	0.9	1
17. Visual search	1.2	2	1	2	0.9	0	1	2
18. Prospective memory	0.6	1	0.5	1	0.6	1	0.5	1
19. Attention span - Working memory	1.7	2	1.2	2	1.5	1.5	1.2	1
20. Letter fluency - Category fluency	0.8	0	5	6	0.7	1	4.8	6

BEAT v2 Item 1: Questionnaire

To improve readability and comprehension, slight changes were made to the wording of Item 3 (changed to '*On how many days in the past week did you have an exaggerated emotional reaction to a situation?*'), Item 5 (changed to '*On how many days in the past week did you have trouble completing a task?*'), Item 8 (changed to '*On how many days in the past week did you not have the things you needed with you (for example, didn't have wallet at shops)?*' and Item 9 (changed to '*On how many days in the past week did you get easily upset at something insignificant?*').

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, so it was retained.

BEAT v2 Item 2: Trail making

No changes were made to this item.

The distribution of scores and descriptive statistics across groups suggested this item had limited known-groups validity. This suggested a revision of the item was required, such as an exploration of alternative response outcomes (for example, time taken to complete the trail), so the item was retained.

BEAT v2 Item 3: Clock drawing

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, so it was retained.

BEAT v2 Item 4: Figure copying

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had poor known-groups validity. The SUD group performed better than the normal control group, suggesting a change to this item was required.

BEAT v2 Item 5: Word list learning

Given the feasibility of a seven-word list across three successive trials was established, a new list of words was used in BEAT v2 that differed from the words used in the MoCA. The words were chosen to represent exemplars of different semantic categories to facilitate cueing processes for Item 14 Word list recall.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, so it was retained.

BEAT v2 Item 6: Attention span

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had poor known-groups validity. However, it was retained because attention span was considered an important aspect of cognitive assessment for an SUD population.

BEAT v2 Item 7: Working memory

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, so it was retained.

BEAT v2 Item 8: Number tapping

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had poor known-groups validity, with the SUD group performing better than normal controls, so it was discarded.

BEAT v2 Item 9: Abstraction

The MoCA items 'train-bicycle' and 'watch-ruler' were replaced with two new word pairs. Analysis revealed no significant benefit of including the bonus item, so the score was based on the simpler score of total number of correct items, with a maximum total score of four.

The distribution of scores and descriptive statistics across groups suggested this item had limited known-groups validity. However, it was retained because abstraction was considered an important aspect of cognitive assessment for an SUD population.

BEAT v2 Item 10: Letter fluency

Words starting with the letter 'S' and fruits were selected for letter and category fluency tasks, respectively, because they had comparable normative values and it was desirable to be able to determine at face value (that is, on the basis of raw scores) whether there was a 'significant' difference between the measures.[44]

The distribution of scores and descriptive statistics across groups suggested this item had reasonable known-groups validity, so it was retained.

BEAT v2 Item 11: Category fluency

As per Item 10 above, the category was changed from animals to fruits. The distribution of scores and descriptive statistics across groups suggested this item had reasonable known-groups validity, so it was retained.

BEAT v2 Item 12: Orientation

Following the considerable ceiling effects observed in BEAT v1 for this item, Item 12: Orientation was removed.

BEAT v2 Item 13: Reading

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had reasonable known-groups validity, so it was retained.

BEAT v2 Item 14: Word list recall

As outlined above regarding Item 5 Word list learning, a different list of seven words was included in BEAT v2. As such, category cues were updated as follows.

'*One of the words was a musical instrument/something you find in a kitchen/an animal/part of a face/an item of clothing/a vegetable/a profession*.'

For words that are not recalled with a category cue, provide the following multiple choice responses.

> '*Was it:*
> *guitar, piano or violin?*
> *sink, stove or fridge?*
> *frog, monkey or snake?*
> *cheek, lip or nose?*
> *shirt, jeans or singlet?*
> *carrot, lettuce or cucumber?*
> *lawyer, doctor or nurse?'*

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, so it was retained. MIS produced a greater range of scores, and therefore larger score differences between groups. As such, it was chosen as the score of preference for this item.

BEAT v2 Item 15: Visual memory

No changes were made to this item. Although this item discriminated reasonably well between groups, there was a ceiling effect, suggesting a requirement to modify the stimulus to increase the complexity of the task.

BEAT v2 Item 16: Naming

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had reasonable known-groups validity, so it was retained.

BEAT v2 Item 17: Visual search

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had poor known-groups validity, with the SUD group performing better than normal controls, suggesting a need for revision.

BEAT v2 Item 18: Prospective memory

Due to findings in the literature that prospective memory is impaired in SUD, a single event-based prospective memory item was added.[45,46] When the instructions were provided for word list learning, the following was instruction was added: '*Also, if you see a bell, I want you to say to me, "That rings a bell*"'. The cue for the correct response was during the naming task (one of the items is a bell).

Responses were coded as:

1 = stated 'that rings a bell'

0.5 = recalled that something had to be stated but could not recall the precise phrase

0 = did not respond to the cue at all

The distribution of scores and descriptive statistics across groups showed that this item had poor known-groups validity, suggesting a need for revision.

BEAT v2 Item 19: Attention span and Working memory difference

The decomposition paradigm of the process approach to neuropsychological assessment was used to investigate differences between attention span and working memory scores. [47,48] A larger difference between forward and reverse digit span has been observed in some clinical groups and there is evidence of particularly reduced backward digit span scores in university students who binge drink. [49,50,51,52]

This item was calculated by subtracting the Working Memory raw score from the Attention Span raw score.

The distribution of scores and descriptive statistics across groups suggested this item had promising known-groups validity, given larger differences (of four or more) were more common in the SUD group. Hence, the item was retained.

BEAT v2 Item 20: Letter fluency and Category fluency difference

The decomposition paradigm of the process approach to neuropsychological assessment was used to investigate differences between letter and category fluency scores.[47,48] It has long been established that letter fluency is more difficult than category fluency.[53] This is due to the relative reliance on frontal systems in generating words that start with a particular letter and temporal lobe systems in generating items from familiar categories.[54] In support of this model, individuals with temporal lobe lesions have been shown to have greater deficits in category, compared with letter

fluency.[55] Because the letter S and category fruits were chosen due to their relatively similar fluency counts, low, and possibly negative, Letter fluency and Category fluency scores would be associated with relatively poorer generation of novel, compared to familiar, ideas, implicating poorer executive functioning.[44]

This item was calculated by subtracting the Category Fluency raw score from the Letter Fluency raw score.

The distribution of scores and descriptive statistics across groups suggested this item had promising known-groups validity, and so it was retained.

BEAT v3

BEAT v3 pilot sample

Data collected in a sample of n=527 individuals with SUD and n=150 normal controls were examined by visual inspection of frequency histograms and consideration of descriptive statistics at the item level. As with BEAT v2, known-groups validity at the item level was the main criterion for item inclusion or revision. See Table 8 for sample characteristics and Table 9 for descriptive statistics.

Table 8. Characteristics of the BEAT v3 pilot sample

Characteristic		Substance use disorder (n=527)		Normal control (n=150)	
Gender (%Male)		62%		41%	
		Mean	Standard deviation	Mean	Standard deviation
Age		36.5	10.8	28.5	13.1
Education		10.7	2.1	13.5	2.0
Test of Premorbid Functioning		92.3	12.2	107.5	12.7
Primary substance of use		Number	%		
	Alcohol	201	38		
	Methamphetamine	231	44		
	Other stimulants	17	3		
	Heroin	21	4		
	Other opiates or opioids	7	1		
	Sedatives, hypnotics or tranquilisers	4	1		
	Cannabis	43	8		

Table 9. Descriptive statistics for the SUD and normal control groups for BEAT v3 items

Item	Substance use disorder (n=527)				Normal control (n=150)			
	Mean	Median	Standard deviation	Inter-quartile range	Mean	Median	Standard deviation	Inter-quartile range
1. Questionnaire	25.2	22	15.4	23	19.3	17	11.8	17
2. Trail making	36.2	30	28.1	20	27.9	24	14.7	14.3
3. Clock drawing	1.3	1	0.6	1	1.6	2	0.5	1
4. Figure copy (trapezium start)	1.4	1	1.6	0	1.3	1	1.3	0
4. Figure copy (accuracy)	8.8	9	0.6	0	9	9	0.2	0
5. Word list learning	16.6	17	2.7	4	17.9	19	2.6	3
6. Attention span	5.6	6	1	1	5.9	6	1	2
7. Working memory	3.8	4	1.1	1	4.5	4	1	1
9. Abstraction	2.4	3	1.2	1	2.8	3	1	2
10. Letter fluency	15.3	15	4.9	7	17.7	18	4.8	7
11. Category fluency	14.9	15	4.3	6	16.3	16	4.5	6
13. Reading	4.2	4	1.2	1	5.3	5	3.5	2
14. Word list recall (MIS)	16.1	17	3.6	5	17.1	18	3	5
15. Visual memory	6.4	7	1.9	3	7.2	7	1.9	3
16. Naming	14.7	15	0.7	0	15	15	0.2	0
17. Visual search	106.4	89	59.8	65.8	103	90.5	53.9	56.8
18. Prospective memory	1	1	0.7	1.1	1.5	1.5	0.6	1
19. Attention span - Working memory	1.9	2	1.2	2	1.4	1	1.2	1
20. Letter fluency - Category fluency	0.3	0	5	7	1.3	1	4.8	7

BEAT v3 General changes

Instructions were added to the protocol to increase inter-rater and test-retest reliability. Checkboxes were added to assist the examiner to track progress during administration of the test.

BEAT v3 Item 1: Questionnaire

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 2: Trail making

A timed component was added to introduce a measure that captured processing speed. This conformed more to the original Trail Making Test administration procedure of recording the time taken to complete the task (Army Individual Test Battery).[19] The instructions were changed such that the examiner pointed out an error when it was made and prompted the examinee to return to the previous successfully selected item. This prompting was included in the time taken to complete the task, and hence captured errors in the overall time. This was considered a useful modification to the BEAT, as there were previously no timed items to capture processing speed. Given the errors were captured in the time taken to complete the trail, error data were no longer used.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 3: Clock drawing

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 4: Figure copying

In BEAT v2, the SUD group performed better than the normal control group on this item. As such, a more complex figure was developed for BEAT v3.

The new figure had nine elements and the order of completion was recorded by the examiner on the record form.

The initial element copied was also recorded, with the prediction that most non-cognitively impaired individuals would start by copying the outer figure, the large trapezium, before proceeding to the other elements. The Figure Copy (Trapezium Start) score was one if the person started by drawing the trapezium before the other elements.

Copying accuracy was also examined. This was the sum of all elements copied, with a maximum score of nine.

Although both measures (Trapezium Start and Accuracy) failed to classify large numbers of the SUD group, Accuracy was retained, because it did so at a slightly higher rate, and it was considered important to retain at least one measure of copying. Figure copying was also necessary for assessment of Item 15: Visual Memory.

BEAT v3 Item 5: Word list learning

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 6: Attention span

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 7: Working memory

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 8: Number tapping

N/A as this item was previously removed.

BEAT v3 Item 9: Abstraction

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 10: Letter fluency

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 11: Category fluency

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 12: Orientation

N/A, as this item was discarded in a previous version.

BEAT v3 Item 13: Reading

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 14: Word list recall

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 15: Visual memory

A more complex geometric figure replaced the previous figure because both the normal control and SUD groups performed at, or close to, the ceiling with the previous figure. The number of elements in the new figure was nine, compared to four in the previous figure.

One point was credited for each of the nine elements that was recognisable and present, even if placement was incorrect.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 16: Naming

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 17: Visual search

Given the poor discriminative validity of this item in BEAT v2, the outcome of interest was changed to the time taken to identify the targets.

The distribution of scores and descriptive statistics across groups suggested this item had poor known-groups validity.

BEAT v3 Item 18: Prospective memory

In order to improve the known-groups validity of this item, an additional prospective memory instruction was added. After introducing the test, the following was stated by the examiner, '*At the end of this test, I will say "that's the end". When I say that, I want you to remind me to ask you about your schooling*.'

As per the coding of the existing prospective memory instruction, responses to the new instruction were coded as:

1 = reminded the examiner to ask about their schooling

0.5 = recalled that something had to be stated but could not recall what it was

0 = did not respond to the cue at all

The results of both instructions were summed, with a maximum score of two.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 19: Attention span and Working memory difference

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 20: Letter fluency and Category fluency difference

No changes were made to this item. The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 21: Motor series

An item based on the Luria Motor Series test was added.[56] This was based on findings that individuals with SUD have difficulty establishing and maintaining motor movement patterns.[57,58] The score comprised the number of correct repetitions of the sequence completed out of six.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 22: Incidental learning

Because memory impairment is relatively mild in individuals with SUD, it was considered that an incidental memory paradigm may better elicit impairment. The item comprised asking the examinee to recall the 15 silhouette drawings they were asked to identify in the naming item. They were given credit if they recalled the item, even if they named it incorrectly during the naming item.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

BEAT v3 Item 23: Delay discounting

Delay discounting refers to a person's tendency to discount the value of a future reward in favour of more immediate rewards.[59] Delay discounting is seen in higher rates among people with substance use disorders.[59,60,61,62,63] Although delay discounting is not typically included in the assessment of cognition per se, the concept is related to cognitive constructs including IQ and the executive function of impulse control.[64]

A series of 27 delay discounting forced choice items was developed, modelled on examples in Johnson & Bickel.[65] These were based on 22 'now' amounts (ranging from $1 to $90) and three 'later' amounts ($10, $50, $100) over three time periods (one week, one month, six months).

Examinees were asked to indicate whether they would prefer a certain amount of money now or a different amount of money at some point in the future for 27 different hypothetical situations where they were offered money.

The instruction was: '*In a hypothetical situation where someone is offering you money, would you prefer to receive…*'

The sum of choices at the delayed period of time was the score for this item, with a maximum score of 27.

The distribution of scores and descriptive statistics across groups suggested this item had good known-groups validity, and so it was retained.

Education

Years of education according to the Australian Qualifications Framework was included at the end of the form following the final prospective memory prompt '*That's the end*'.[66] When formal education is not otherwise recorded, this provides the examiner with clear questions to accurately calculate the total years of education completed.

Validation of the BEAT

Validity

A series of item-level and test-level analyses were conducted on BEAT v3 to finalise the item list for the BEAT release version and establish validity.

Item-level analyses

One of the purposes of developing the BEAT was to delineate differences in cognition between individuals with SUD and normal controls. If the tool is able to distinguish between these groups, then it could be said to have known-groups validity, a form of construct validity. Another purpose of BEAT development was to identify individuals with cognitive impairment among those seeking treatment for SUD. If the tool is able to distinguish between those with and without cognitive impairment within an SUD population, then it could be said to have criterion validity.

Both known-groups validity and criterion validity were examined at the item level in order to exclude items that did not demonstrate at least one of these forms of validity.

Known-groups validity – item-level analyses

Two separate series of analyses were conducted to examine known-groups validity:

- univariate analyses of variance to examine absolute differences in cognition between an SUD and normal control sample
- univariate analyses of covariance to examine differences in cognition between an SUD and normal control sample with relevant factors accounted for.

Each of these analyses was considered important for different reasons. Establishing absolute differences in cognition is important at an ecological level in that a person below a certain threshold of cognition may be less able to process information relevant to SUD treatment and therefore not derive as much benefit from it. Their substance use history and other factors such as age, education and premorbid intelligence may all contribute to the person being below threshold. Regardless of the factors that may bring them below threshold, what is ecologically meaningful is that they are less able to benefit from treatment and may therefore require additional supports to get the most out of treatment. A separate but important question is whether or not there are differences in scores between an SUD and normal control group when factors apart from SUD-related cognitive impairment are accounted for, such as age, education and premorbid intelligence. Items that reveal such differences are particularly sensitive to SUD-related cognitive decline.

To examine known-groups validity at the item level, univariate analyses of variance with the item score as the dependent variable and group (SUD, normal control) as the independent variable were conducted. The first column of Table 10 shows all items that significantly differed between the groups in absolute terms. All items except for Item 17: Visual search differed significantly between the groups in the expected direction.

Between the SUD and normal control samples, there was a significant difference between age, $F(1,675) = 57.47$, $p < 0.001$, education, $F(1,669) = 210.49$, $p < 0.001$, TOPF score, $F(1,651) = 174.83$, $p < 0.001$ and gender, $\chi 2(1,675) = 21.09$, $p < 0.001$. As such, analyses of covariance with the item score as the dependent variable and group as the independent variable were conducted with age, education, TOPF and gender entered as covariates. The second column of Table 10 shows the items that significantly differed between the groups when these factors were controlled for. Item 1: Questionnaire, Item 3: Clock drawing, Item 7: Working memory, Item 11: Category fluency, Item 16: Naming, Item 19. Attention span - Working memory, Item 21: Motor series and Item 23: Delay discounting significantly distinguished between groups when age, education,

premorbid intelligence and gender were accounted for. This suggests that these eight items are particularly sensitive to specific SUD-related cognitive impairment.

Table 10. Results of item-level univariate analyses of variance and covariance

Item	Known-groups validity		Criterion validity			Included in final version
	Difference between Normal control and SUD samples	Difference between Normal control and SUD samples (corrected for age, education, gender and TOPF)	Difference between Impaired and Intact levels of executive functioning criterion	Difference between Intact and Some Impairment levels of executive functioning criterion	Difference between Some Impairment and Impaired levels of executive functioning criterion	
1. Questionnaire	✓	✓	✓	✓	✓	✓
2. Trail making	✓		✓			✓
3. Clock drawing	✓	✓		✓		✓
4. Figure copy	✓		✓			✓
5. Word list learning	✓		✓		✓	✓
6. Attention span	✓		✓			✓
7. Working memory	✓	✓	✓	✓	✓	✓
9. Abstraction	✓		✓		✓	✓
10. Letter fluency	✓		✓	✓	✓	✓
11. Category fluency	✓	✓	✓	✓	✓	✓
13. Reading	✓		✓	✓	✓	✓
14. Word list recall	✓		✓		✓	✓
15. Visual memory	✓		✓			✓
16. Naming	✓	✓				✓
17. Visual search	✓					
18. Prospective memory	✓		✓	✓	✓	✓
19. Attention span - Working memory	✓	✓	✓			✓
20. Letter fluency - Category fluency	✓					✓
21. Motor series	✓	✓	✓		✓	✓
22. Incidental learning	✓		✓		✓	✓
23. Delay discounting	✓	✓				✓

✓ Significant at $p<0.05$ level, * Marginally significant ($p=0.052$)

Criterion validity – item level analyses

The first step in establishing criterion validity is to clearly define the criterion. In line with the recommendation of the American Academy of Clinical Neuropsychology consensus conference, requiring impairment on both performance-based tests and 'other findings related to functional capacity', a variable combining the results of performance- and inventory-based measures was constructed.[17] Adopting such an approach potentially improves both ecological and predictive validity and is consistent with the vector approach of neuropsychological assessment.[67,68]

The first step in constructing this variable was to convert three performance-based measures and one inventory-based measure to standard scores. The performance-based measures were:

- the Interference trial of the Stroop Test, Golden version, a measure of response inhibition
- Alpha Score, a measure of working memory
- number of unique designs on the Five Point Test, a measure of nonverbal idea generation.[69,70,71]

Stroop Interference was corrected by age according to Morrow.[72] Alpha Score was corrected for age as per Craik et al.[70] The Five Point Test Unique Designs was corrected for age and education as per Goebel et al.[71] The inventory-based measure was the Global Executive Composite (GEC) of the BRIEF-A, which was age corrected as per the manual.[14]

Consistent with established practice for neuropsychological tests, a cut score of 1.5 standard deviations below the mean was used to indicate impairment on each of these measures; 1.5 standard deviations above the mean was applied to the BRIEF-A GEC, as higher scores indicated greater impairment on this measure. [73,74]

The criterion variable, named Executive Functioning Criterion (EFC) had the following three levels.

- Intact (Intact on the Alpha Score, Stroop Interference, Five Point Test Unique Designs and BRIEF-A GEC).
- Some Impairment (Impaired on the Alpha Score, Stroop Interference, Five Point Test Unique Designs or BRIEF-A GEC).
- Impaired (Impaired on the BRIEF-A GEC and at least one of Alpha Score, Stroop Interference and Five Point Test Unique Designs).

Due to missing data, EFC could only be calculated for n=467 SUD and n=140 normal control participants. The characteristics of the three EFC level subsamples for the SUD group are shown in Table 11. Because the standard scores used to calculate EFC were already demographically corrected, univariate analyses of variance were conducted to examine criterion validity at the item level, with the item score as the dependent variable and EFC (Intact, Some Impairment, Impaired) as the independent variable. The third column of Table 10 shows all items that significantly differed between the Intact and Impaired levels of EFC. The fourth and fifth columns show items that differed between the Intact and Some Impairment, and Some Impairment and Impaired, levels of EFC, respectively.

Item 17: Visual Search was the only item that lacked both known-groups and criterion validity, so it was removed from the BEAT release version.

Table 11. Characteristics of the EFC subsamples within the SUD sample

Characteristic	Intact (n=147)		Some Impairment (n=221)		Impaired (n=99)	
Gender (%Male)	56%		60%		69%	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Age	38.3	11.8	35.9	10.2	34.7	11.1
Education	11.1	2.3	[a]10.9	1.8	9.9	1.9
Test of Premorbid Functioning	[b]96.7	12.1	[c]92.3	11.8	[d]86.3	10.6
Primary substance of use	Number	%	Number	%	Number	%
Alcohol	73	49.7	87	39.4	26	26.3
Methamphetamine	52	35.4	92	41.6	54	54.5
Other stimulants	3	2.0	12	5.4	1	1.0
Heroin	6	4.1	9	4.1	5	5.1
Other opiates or opioids	4	2.7	2	0.9	1	1.0
Sedatives, hypnotics or tranquilisers	2	1.4	1	0.5	1	1.0
Cannabis	7	4.8	18	8.1	11	11.1

a based on n=219, b based on n=145, c based on n=218, d based on n=97

Final (release) version of the BEAT

Thus, the final (release) version of the BEAT comprised 20 items. See Table 12, which shows the included and excluded items for each version of the BEAT and Table 13, which shows the descriptive statistics for the re-numbered items in the final (release) version of the BEAT for the n=501 SUD participants and n=145 normal control participants who had BEAT data for every item.

Table 12. Items included and excluded for each version of the BEAT

	Item	Version 1	Version 2	Version 3	Final version
1	Questionnaire	✓	✓	✓	✓
2	Trail making	✓	✓	✓	✓
3	Clock drawing	✓	✓	✓	✓
4	Figure copy	✓	✓	✓	✓
5	Word list learning	✓	✓	✓	✓
6	Attention span	✓	✓	✓	✓
7	Working memory	✓	✓	✓	✓
8	Number tapping	✓	✓		
9	Abstraction	✓	✓	✓	✓
10	Letter fluency	✓	✓	✓	✓
11	Category fluency	✓	✓	✓	✓
12	Orientation	✓			
13	Reading	✓	✓	✓	✓
14	Word list recall	✓	✓	✓	✓
15	Visual memory	✓	✓	✓	✓
16	Naming	✓	✓	✓	✓
17	Visual search	✓	✓	✓	
18	Prospective memory		✓	✓	✓
19	Attention span - Working memory		✓	✓	✓
20	Letter fluency - Category fluency		✓	✓	✓
21	Motor series			✓	✓
22	Incidental learning			✓	✓
23	Delay discounting			✓	✓

Table 13. Descriptive statistics for the SUD and normal control groups for BEAT final (release) version items

Item	Substance use disorder (n=501)				Normal control (n=145)			
	Mean	Median	Standard deviation	Inter-quartile range	Mean	Median	Standard deviation	Inter-quartile range
1. Questionnaire	25.3	23	15.3	23	19.5	18	11.9	17
2. Trail making	36.0	30	27.6	20	27.8	24	14.9	14.5
3. Clock drawing	1.3	1	0.6	1	1.6	2	0.5	1
4. Figure copy	8.9	9	0.6	0	9.0	9	0.2	0
5. Word list learning	16.6	17	2.6	4	17.8	19	2.6	3.5
6. Attention span	5.7	6	.9	1	5.9	6	1	2
7. Working memory	3.8	4	1.1	1	4.5	4	1	1
8. Abstraction	2.4	3	1.2	1	2.9	3	1	2
9. Letter fluency	15.3	15	4.9	7	17.8	18	4.8	6.5
10. Category fluency	15.0	15	4.2	6	16.5	16	4.5	6.5
11. Reading	4.3	4	1.1	1	5.3	5	3.5	2
12. Word list recall	16.2	17	3.4	5	17.2	18	3	4.5
13. Visual memory	6.4	7	1.9	3	7.2	7	1.8	3
14. Naming	14.7	15	0.7	0	15.0	15	0.2	0
15. Prospective memory	1.0	1	0.7	1.5	1.5	1.5	0.6	1
16. Attention span - Working memory	1.9	2	1.2	2	1.4	1	1.2	1
17. Letter fluency - Category fluency	0.3	0	5	7	1.3	1	4.9	6.5
18. Motor series	4.5	6	2	2	5.4	6	1.4	0
19. Incidental learning	6.8	7	2.8	4	8	8	2.3	2
20. Delay discounting	11.5	11	6.5	8.5	17.0	16.5	7.3	12

Test-level analyses

In order to investigate known-groups and criterion validity of the BEAT final (release) version, the sample was randomly split into two approximately equally sized samples. One sample served as the training sample and the other as the validation sample. See Table 14 for subsample characteristics.

Table 14. Training and validation subsamples characteristics

Characteristic	Training sample				Validation sample			
	SUD (n=259)		Normal control (n=72)		SUD (n=268)		Normal control (n=78)	
Gender (%Male)	61%		39%		63%		42%	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Age	36.8	11.0	28.1	12.7	36.2	10.7	28.9	13.5
Education	10.7	2.0	13.3	2.0	10.7	2.2	13.6	2.1
Test of Premorbid Functioning	91.7	11.6	108.6	12.4	92.8	12.7	106.4	13.0
Primary substance of use	No.	%			No.	%		
Alcohol	96	37			105	39		
Methamphetamine	111	43			120	45		
Other stimulants	7	3			10	4		
Heroin	12	5			9	3		
Other opiates or opioids	3	1			4	2		
Sedatives, hypnotics or tranquilizers	3	1			1	0.4		
Cannabis	25	10			18	7		

Score transformation

Based on data from the training sample of normal controls, the 20 items in the final version of the BEAT were transformed to standard scores on the basis of quartiles. First to fourth quartile raw scores were converted to scores of 0, 1, 2 and 3, respectively. Reverse scoring was applied to items 1. Questionnaire, 2. Trail Making and 19. AS-WM. The cut scores are presented in Table 15. The BEAT score was derived by summing the transformed item scores.

Table 15. Raw to transformed score conversions

Item	Raw score	Transformed score
1. Questionnaire	0-9	3
	10-18	2
	19-29	1
	30-70	0
2. Trail making	0-16	3
	17-23	2
	24-33	1
	>33	0
3. Clock drawing	0	0
	1	1
	2	3
4. Figure copy	0-8	0
	9	3
5. Word list learning	0-16	0
	17-18	1
	19-20	2
	21	3
6. Attention span	0-4	0
	5	1
	6	2
	7	3
7. Working memory	0-3	0
	4	1
	5	2
	6	3
8. Abstraction	0-1	0
	2	1
	3	2
	4	3
9. Letter fluency	0-14	0
	15-17	1
	18-20	2
	>20	3
10. Category fluency	0-13	0
	14-15	1
	16-18	2
	>18	3
11. Reading	0-3	0
	4	1
	5	2
	6-7	3
12. Word list recall	0-15	0
	16-17	1
	18-19	2
	20-21	3
13. Visual memory	0-6	0
	7	1
	8	2
	9	3
14. Naming	0-14	0
	15	3
15. Prospective memory	0-0.5	0
	1-1.5	1
	2	3
16. Attention span - Working memory	<3	3
	>2	0
17. Letter fluency - Category fluency	<-2	0
	-2-0	1
	1-3	2
	>3	3
18. Motor series	0-5	0
	6	3
19. Incidental learning	0-5	0
	6-7	1
	8	2
	9-15	3
20. Delay discounting	0-10	0
	11-15	1
	16-21	2
	22-27	3

Known-groups validity – test level

The scoring transformations based on the training sample were applied to the validation sample.

In order to examine known-groups validity at the test level, univariate analysis of covariance with the BEAT score as the dependent variable and group (SUD, normal control) as the independent variable with age, education, TOPF and gender entered as covariates was conducted on the validation sample. There was a significant effect of group, $F(1,321) = 7.64$, $p = 0.006$, establishing known-groups validity.

Table 16 shows the descriptive statistics for the SUD and normal control validation samples.

Table 16. BEAT descriptive statistics for the SUD and normal control validation samples (known groups validity data) and the EFC severity levels for the SUD validation sample (criterion validity data)

Statistic	Known-groups validity		Criterion validity		
			Executive functioning criterion (SUD sample)		
	SUD	Normal control	Intact	Some impairment	Impaired
Number	258	74	81	113	46
Mean	28.2	37.8	33.2	27.3	21.0
Standard deviation	8.4	7.0	7.3	7.5	6.8
Minimum	8	21	17	10	8
Maximum	48	50	48	46	38

Criterion validity – Test level

In order to examine criterion validity at the test level, univariate analysis of covariance with the BEAT score as the dependent variable and EFC level (intact, some impairment, impaired) as the independent variable with age, education, TOPF and gender entered as covariates was conducted on the validation sample. There was a significant effect of EFC level, $F(2,231) = 22.08$, $p < 0.001$, establishing criterion validity. Pairwise comparisons using Bonferroni corrections revealed significant differences between all levels of EFC. Table 15 shows the BEAT descriptive statistics at each level of the EFC for the SUD validation sample.

Another aspect of validity is the extent to which a test classifies individuals with a condition of interest versus those that do not have the condition of interest. Because the some impairment (n=131; 113 SUD and 18 normal control participants) group of EFC did not fall into either of these categories, it was eliminated from classification statistics analysis, and only the Intact (n=132; 81 SUD and 51 normal control participants) and impaired (n=49; 46 SUD and 3 normal control participants) groups were included.

Receiver operating characteristic (ROC) curve analysis between the EFC intact and impaired group for the whole validation sample revealed an optimal cut score of ≤30. When this cut score was applied, it resulted in 86% sensitivity and 77% specificity, and explained 88% area under the

curve. Negative predictive power was 94% and overall classification accuracy was 79% as depicted in Table 16. See the ROC curve in Figure 2.

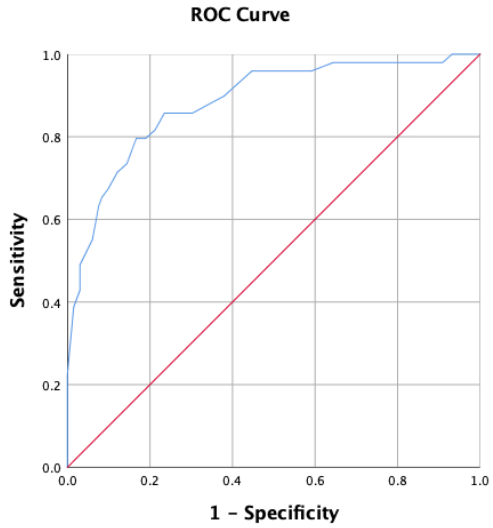

Figure 2. BEAT ROC curve for the impaired and intact groups from the whole validation sample

When the cut score of ≤ 30 was applied to the SUD validation sample, it resulted in 91% sensitivity and 69% specificity, and explained 88% area under the curve. Negative predictive power was 94% and overall classification accuracy was 77% as depicted in Table 17. See the ROC curve in Figure 3.

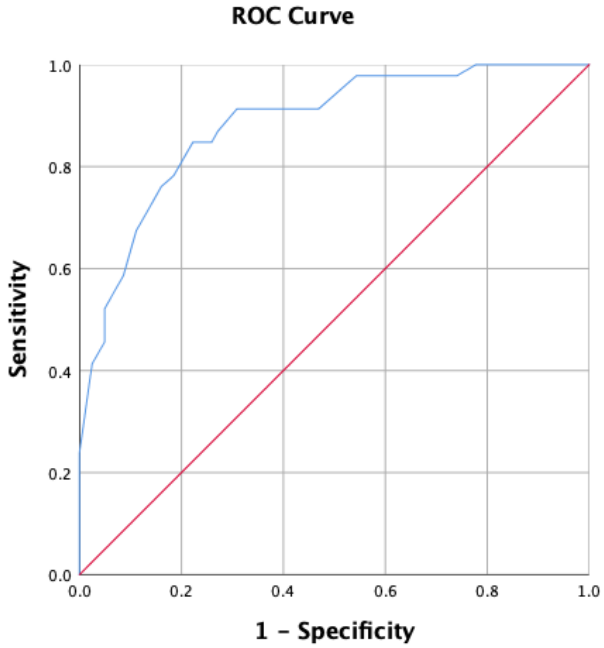

Figure 3. ROC curve for the BEAT for the impaired and intact groups from the SUD validation sample

Table 17. Classification statistics for the whole and SUD validation samples

Statistic	Whole validation sample (N=181)		SUD validation sample (N=130)	
	Value %	95% CI	Value %	95% CI
Sensitivity	85.71	72.76 to 94.06	91.30	79.21 to 97.58
Specificity	76.52	68.35 to 83.45	69.14	57.89 to 78.93
Positive likelihood ratio	3.65	2.63 to 5.07	2.96	2.11 to 4.15
Negative likelihood ratio	0.19	0.09 to 0.37	0.13	0.05 to 0.32
Disease prevalence	27.07	20.75 to 34.16	36.22	27.88 to 45.22
Positive predictive value	57.53	49.38 to 65.30	62.69	54.51 to 70.20
Negative predictive value	93.52	87.83 to 96.65	93.33	84.44 to 97.31
Accuracy	79.01	72.34 to 84.69	77.17	68.88 to 84.14

Based on the above analyses, criterion-related validity was established.

Incremental validity

Applying the cut score of ≤ 30 on the BEAT to the entire sample classified 17% of the normal control sample and 60% of the SUD sample as impaired (See Figure 4). While this cut score likely results in a high rate of false positive results (that is, high negative predictive power), this is desirable of a screening test. However, 60% is a relatively high prevalence rate of cognitive impairment in an SUD population based on previous research.[3,9] As such, it is proposed that the BEAT be used in conjunction with the ACE Screening Tool, a screen of risk factors for cognitive impairment.[75] When the criterion for impairment was impaired status on both the ACE Screening Tool and the BEAT, only 1.4% of the normal control and 46.1% of the SUD sample were classified as impaired.

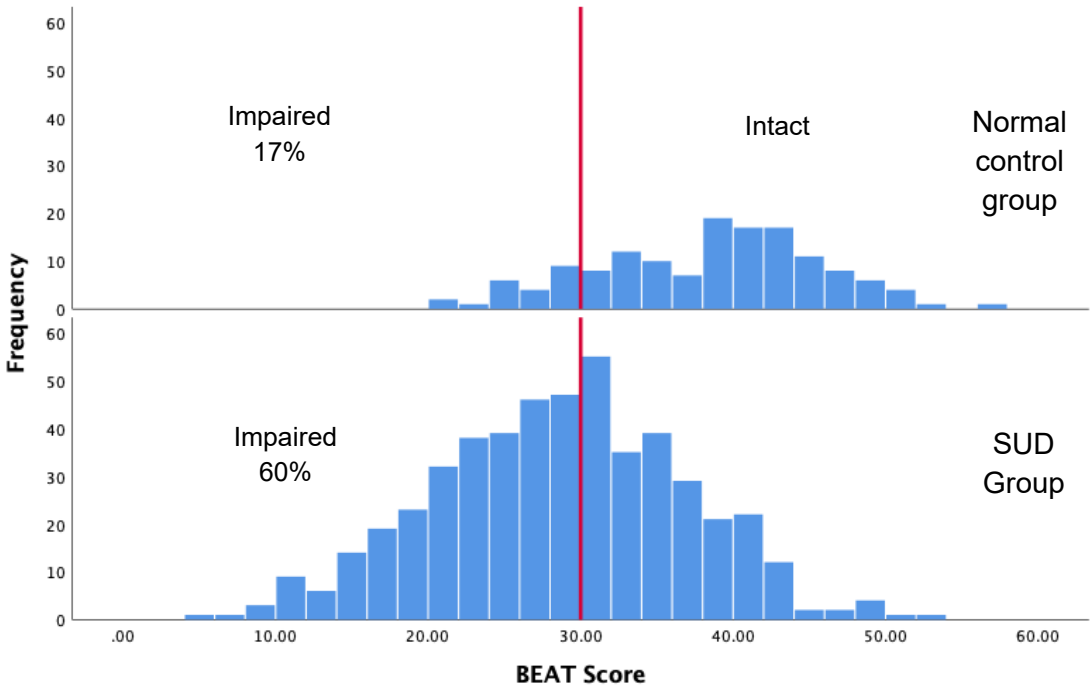

Figure 4. BEAT frequency histograms for the SUD and normal control samples showing percentage classified as impaired

A two-stage screening process

It is recommended that the ACE Screening Tool be used as the first step in screening for cognitive impairment within an SUD population. If a person screens positive on that tool, then it is recommended that they be administered the BEAT. Figures 5 and 6 show the numbers and percentages of the SUD and normal control groups classified as intact or impaired for stages 1 and 2 of the screening process respectively.

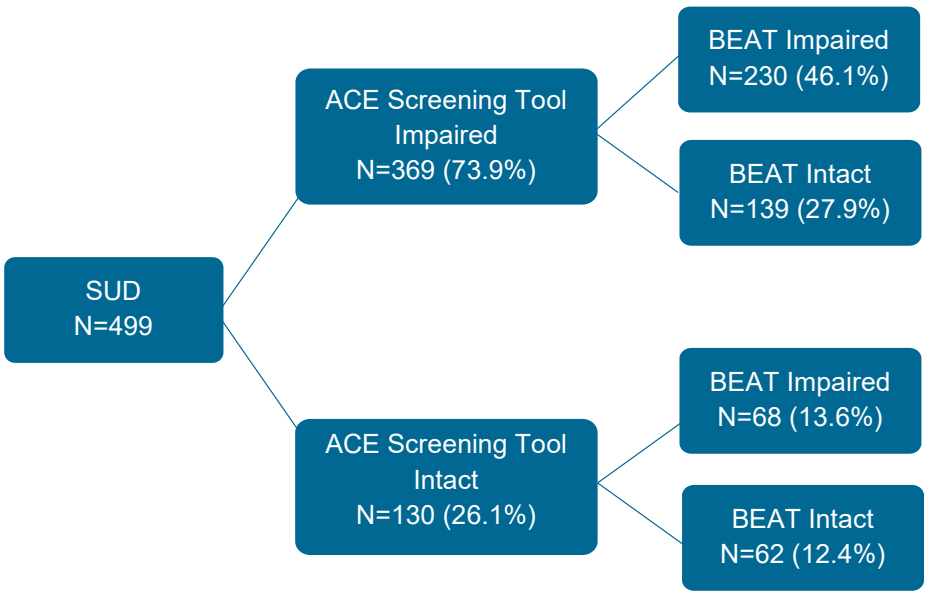

Figure 5. Numbers and percentages of the SUD group classified as intact or impaired for stages 1 and 2 of the screening process

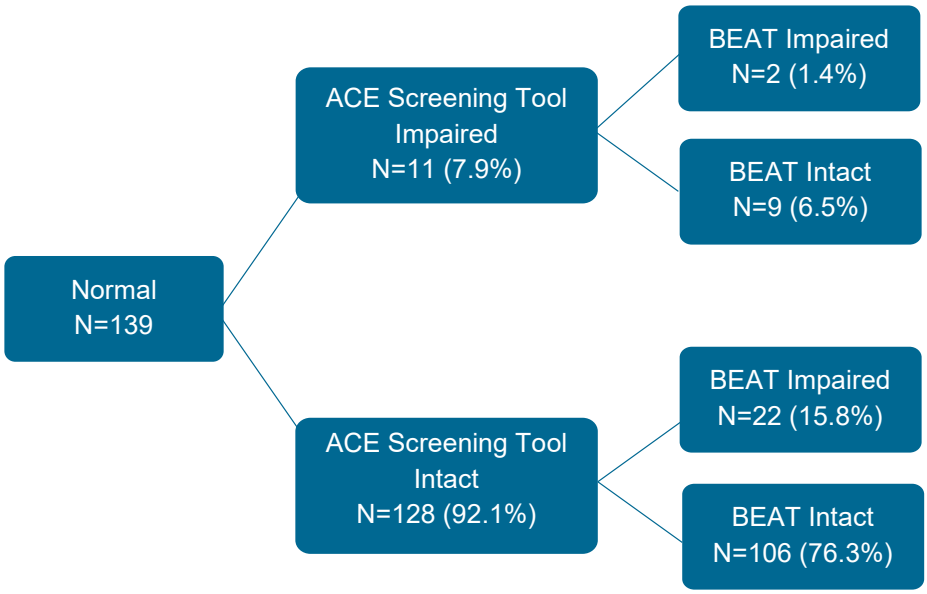

Figure 6. Numbers and percentages of the normal control group classified as intact or impaired for stages 1 and 2 of the screening process

The incremental validity of using this two-staged screening process was examined by calculating the classification statistics, which are presented in Table 18. The two-stage screening process increased sensitivity to 99% and negative predictive power to 98%, which is highly desirable for a screening process. This means that the process is highly likely to detect cognitive impairment when it is present. The lower sensitivity and positive predictive power values of 29% and 36%, respectively, indicate that this increased sensitivity is attained at the cost of a high rate of detecting cognitive impairment when it is not present, as defined by the EF Criterion variable.

Table 18. Classification statistics for combined ACE Screening Tool and BEAT screening for the SUD sample

Statistic	Value %	95% confidence interval
Sensitivity	98.70	92.98 to 99.97
Specificity	29.23	22.95 to 36.15
Positive likelihood ratio	1.39	1.27 to 1.53
Negative likelihood ratio	0.04	0.01 to 0.32
Disease prevalence	28.31	23.03 to 34.06
Positive predictive value	35.51	33.40 to 37.69
Negative predictive value	98.28	88.93 to 99.75
Accuracy	48.90	42.81 to 55.01

Construct validity

BEAT scores were correlated against other performance-based (Stroop Interference, Five Point Test Unique Designs, Alpha Score), inventory-based (BRIEF-A GEC) and screening (MoCA) measures of executive functioning, as shown in Table 19. Each of these correlations was statistically significant at the p ≤0.05 level, which establishes construct validity.

Table 13. Correlations between the BEAT and other executive function measures

Test	Pearson r (n)		
	Whole sample	SUD	Normal control
MoCA			0.518** (145)
Stroop Interference trial	0.530** (643)	431** (498)	0.570** (145)
Five Point Test Unique Designs	0.503** (634)	0.465** (490)	0.326** (144)
Alpha score	0.545** (581)	0.518** (498)	0.290* (83)
BRIEF-A GEC	-0.363** (607)	-0.241**(467)	-0.207* (140)

**p≤0.001, *p≤0.05

Reliability

Internal consistency

Cronbach's alpha was 0.734, which represents adequate internal consistency.[76] This is particularly the case given the highly fractionated nature of executive functioning.[77,78]

Test-retest reliability

Test and retest data were available for 137 SUD and 40 normal control individuals. The median test-retest interval was 55 days for the SUD group, 77 days for the normal control group and 56 days for the entire sample. See Table 20 for the sample characteristics.

Test-retest reliability was established by separately examining intra-class correlation coefficients for the normal control, SUD and combined (normal control and SUD) samples, calculated based on an absolute agreement, 2-way mixed effects model. As can be seen in Table 21, the BEAT demonstrated good reliability for the SUD (ICC = 0.812, 95% CI [0.580, 0.899] r = 0.75) and combined samples (r = 0.845, 95% CI [0.707, 0.908]) and acceptable reliability for the normal control sample (r = 0.793, 95% CI [0.610, 0.890]) according to the interpretation guidelines in Strauss et al.[74] In the combined sample, the BEAT was more reliable than the Five Point Test, Alpha Span and the BRIEF-A GEC; only the Stroop test was more reliable than the BEAT. The BEAT was more reliable than the MoCA for the normal control sample.

Table 20. Retest sample characteristics

Characteristic	Substance use disorder (n=340)		Normal control (n=47)	
Gender (%Male)	56%		45%	
	Mean	Standard deviation	Mean	Standard deviation
Age	36.4	10.0	28.2	12.6
Education	10.7	2.0	13.5	1.9
Test of Premorbid Functioning	93.1	12.1	109.6	11.7
Primary substance of use	Number	%		
Alcohol	125	37		
Methamphetamine	147	43		
Other stimulants	7	2		
Heroin	26	8		
Other opiates or opioids	7	2		
Sedatives, hypnotics or tranquilisers	7	2		
Cannabis	21	62		

Table 21. Test-retest reliability coefficients for the BEAT and comparison measures across the normal control, SUD and combined samples

Test	Normal control sample		SUD sample		Combined sample	
	ICC	Number	ICC	Number	ICC	Number
BEAT	0.793**	40	0.812**	137	0.845**	175
MoCA	0.578*	45				
Stroop Interference trial	0.811**	47	0.884**	237	0.888**	284
Five Point Test Unique Designs	0.684**	38	0.793**	235	0.793**	273
Alpha score	0.693*	20	0.763**	234	0.786**	254
BRIEF-A	0.780	40	0.755**	218	0.775**	258

** p≤0.001, * p<0.01

Inter-rater reliability

Inter-rater reliability was calculated by examining score differences across three examiners. Each of the three examiners scored an equal number of baseline BEAT protocols. One examiner then cross-scored a total of n=60 BEAT protocols, divided equally across the other two examiners. The intraclass correlation coefficient was .994, 95% CI [.985, .997], calculated based on an absolute agreement, 2-way mixed effects model, representing excellent inter-rater reliability.[79]

Outpatient sample validation

The BEAT has been validated for use in an outpatient population within the Illawarra Shoalhaven Local Health District with a sample of n=75 clients with SUD. At the time of writing, the results of that study are in the process of preparation for submission to a peer-reviewed journal.

Conclusion

The 20-item BEAT was developed for the purposes of detecting cognitive impairment, and particularly executive function impairment, in an SUD population. It has been shown to demonstrate known-groups validity, criterion validity, good classification statistics, internal consistency, test-retest reliability, inter-rater reliability and construct validity. Furthermore, used in combination with the ACE Screening Tool, it demonstrates incremental validity, with the two-step screening process showing excellent classification statistics (99% sensitivity to cognitive impairment).

Acknowledgements

The authors would like to acknowledge all of the participants in this research, across several NSW residential and non-residential drug and alcohol services; the staff at those facilities, who took valuable time away from their regular roles and responsibilities to facilitate this research, and the tireless efforts of the paid and volunteer research assistants who collected the data used in this research. We would also like to acknowledge the ACE Steering Committee members for their support in setting up the ACE study, from which much of the data for this project has come.

Authors

- Jamie Berry, Advanced Neuropsychological Treatment Services and Macquarie University
- Talia Nardo, Advanced Neuropsychological Treatment Services and Macquarie University
- Antoinette Sedwell, Agency for Clinical Innovation
- Jo Lunn, We Help Ourselves
- Ely Marceau, University of Wollongong
- Ashleigh Wesseling, Macquarie University
- Maia Zucco, Macquarie University
- Sindella Sugden-Lingard, Advanced Neuropsychological Treatment Services
- Thomas Borchard, Advanced Neuropsychological Treatment Services
- Jennifer Batchelor, Macquarie University
- Edwin Arthur Shores, Advanced Neuropsychological Treatment Services and Macquarie University

Alcohol and other drug services

- Adele House (Coffs Harbour)
- Calvary Riverina Centre (Wagga Wagga)
- Illawarra Shoalhaven Local Health District Alcohol and Other Drug Services
    - Bungora Clinic (Wollongong)
    - LAMP Opioid Treatment Unit (Nowra)
    - Orana Centre (Wollongong)
    - Shoalhaven Drug and Alcohol Service (Nowra)
- Jarrah House (Little Bay)
- Kedesh (Wollongong)
- One80TC (Yarramundi)
- Salvation Army – Dooralong (Dooralong)
- Salvation Army – William Booth (Sydney)
- St Vincent de Paul - Freeman House (Armidale)
- The Glen Centre (Chittaway Point)
- WHOS Hunter (Cessnock)
- WHOS New Beginnings (Rozelle)

Glossary

AOD	alcohol and other drug
BRIEF-A	Behavior Rating Inventory of Executive Functioning, Adult version
CI	confidence interval
EFC	Executive functioning criterion
GEC	Global Executive Composite
ICC	intraclass correlation coefficient
MIS	Memory Index Score
MoCA	Montreal Cognitive Assessment
ROC	Receiver operating characteristic
SUD	substance use disorder
TOPF	Test of Premorbid Functioning
WAIS-IV	Wechsler Adult Intelligence Scale, fourth edition
WHOS	We Help Ourselves

References

1 Fernández-Serrano MJ, Pérez-García M, Schmidt Río-Valle J, et al. *Neuropsychological consequences of alcohol and drug abuse on different components of executive functions*. J Psychopharmacol. 2010;24(9), 1317-1332. doi.org/10.1177/0269881109349841

2 Brorson HH, Ajo Arnevik E, Rand-Hendriksen K, Duckert F. *Drop-out from addiction treatment: A systematic review of risk factors*. Clin Psychol Rev. 2013:33(8), 1010-1024. doi.org/10.1016/j.cpr.2013.07.007

3 Marceau EM, Lunn J, Berry J, Kelly PJ, Solowij N. *The Montreal Cognitive Assessment (MoCA) is sensitive to head injury and cognitive impairment in a residential alcohol and other drug therapeutic community*. J Subst Abuse Treat. 2016;66, 30-36. doi.org/10.1016/j.jsat.2016.03.002

4 Kirsch P, Vassileva J, Verdejo-Garcia A, et al. *A roadmap for integrating neuroscience into addiction treatment: a consensus of the neuroscience interest group of the international society of addiction medicine*. Front. Psychiatry. 2019;10(December), 1-23. https://doi.org/10.3389/fpsyt.2019.00877

5 Berry J, Jacomb I, Lunn J, et al. *A stepped wedge cluster randomised trial of a cognitive remediation intervention in alcohol and other drug (AOD) residential treatment services.* BMC Psychiatry. 2019;19(70), 1-11.

6 Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I et al. *The Montreal Cognitive Assessment, MoCA: A Brief Screening*. J Am Geriatr Soc. 2005;695-699. doi.org/10.1111/j.1532-5415.2005.53221.x

7 Deirdre M, Doerflinger C. *Mental Status Assessment in Older Adults : Montreal Cognitive Assessment : MOCA Version 7.1 (Original Version).* The Hartford Institute for Geriatric Nursing. 2012. 104(3), 1-2.

8 Copersino ML, Fals-Stewart W, Fitzmaurice G, et al. *Rapid cognitive screening of patients with substance use disorders*. Exp Clin Psychopharmacol. 2009;17(5), 337-344. doi.org/10.1037/a0017260

9 Ridley N, Batchelor J, Draper B, et al. *Cognitive screening in substance users: Diagnostic accuracies of the Mini-Mental State Examination, Addenbrooke's Cognitive Examination–Revised, and Montreal Cognitive Assessment.* J Clin Exp Neuropsychol. 2017;00(00), 1-16. doi.org/10.1080/13803395.2017.1316970

10 Alarcon R, Nalpas B, Pelletier S, Perney P. *MoCA as a Screening Tool of Neuropsychological Deficits in Alcohol-Dependent Patients. Alcoholism*. Clin Exp Dent Res. 2015;39(6), 1042–1048. doi.org/10.1111/acer.12734

11 Hagen E, Erga AH, Hagen KP, Nesvåg SM, McKay JR, Lundervold AJ, Walderhaug E. *Assessment of executive function in patients with substance use disorder: A comparison of inventory- and performance-based assessment.* J Subst Abuse Treat. 2016;66, 1-8. doi.org/10.1016/j.jsat.2016.02.010

12 Valls-Serrano C, Verdejo-García A, Caracuel A. (2016). *Planning deficits in polysubstance dependent users: Differential associations with severity of drug use and intelligence*. Drug Alcohol Depend. 2016;162, 72–78. doi.org/10.1016/j.drugalcdep.2016.02.027

13 Folstein MF, Folstein SE, McHugh PR. *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician*. J Psychiat Res. 1975;12(3), 189-198. doi.org/10.1016/0022-3956(75)90026-6

14 Roth R, Isquith P, Gioia G. *Behavior Rating Inventory of Executive Functioning - Adult version (BRIEF-A).* Lutz: PAR. 2005.

15 Marceau EM, Berry J, Lunn J, Kelly PJ, Solowij N. *Cognitive remediation improves executive functions, self-regulation and quality of life in residents of a substance use disorder therapeutic community*. Drug Alcohol Depend. 2017;178(June), 150-158. doi.org/10.1016/j.drugalcdep.2017.04.023

16 Hagen E, Sømhovd M, Hesse M, et al. *Measuring cognitive impairment in young adults with polysubstance use disorder with MoCA or BRIEF-A – The significance of psychiatric symptoms.* J Subst Abuse Treat. 2019;97(August 2018), 21-27. doi.org/10.1016/j.jsat.2018.11.010

17 Guilmette TJ, Sweet JJ, Hebben N, et al. *American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance*. Clin Neuropsychol. 2020;0(0), 1-17. doi.org/10.1080/13854046.2020.1722244

18 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (5th ed). Arlington, VA. 2013.

19 Army Individual Test Battery. *Manual of directions and scoring*. Washington, DC. War Department, Adjutant General's Office. 1944.

20 Shulman KI, Shedletsky R, Silver IL. *The challenge of time: Clock-drawing and cognitive function in the elderly*. Int J Geriatr Psychiatry. 1986;1(2), 135-140. doi.org/10.1002/gps.930010209

21 Shulman KI, Pushkar Gold D, Cohen CA, Zucchero CA. *Clock-drawing and dementia in the community: A longitudinal study*. Int J Geriatr Psychiatry. 1993;8(6), 487-496. doi.org/doi:10.1002/gps.930080606

22 Mainland BJ, Amodeo S, Shulman KI. *Multiple clock drawing scoring systems: Simpler is better.* Int J Geriatr Psychiatry. 2014;29(2), 127–136. doi.org/10.1002/gps.3992

23 Hudson KD, Farran EK. *Drawing the line: Drawing and construction strategies for simple and complex figures in Williams syndrome and typical development*. Br J Dev Psychol. 2011;29(4), 687-706. doi.org/10.1016/j.ridd.2013.04.004

24 *WAIS-IV Administration and Scoring Manual.* San Antonio, Texas: NCS Pearson Inc. 2008.

25 Robertson IA, Manly T. Andrade J, Baddeley BT, Yiend, J. *Oops!: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects.* Neuropsychologia. 1997;35(6), 747-758. doi.org/10.1016/S0028-3932(97)00015-8

26 O'Connell RG, Bellgrove M, Dockree PM, et al. *Self-Alert Training: Volitional modulation of autonomic arousal improves sustained attention.* Neuropsychologia. 2008;46, 1379-1390. doi.org/10.1016/j.neuropsychologia.2007.12.018

27 O'Connell RG, Bellgrove M, Dockree PM, Robertson IH. *Cognitive remediation in ADHD: effects of periodic non-contingent alerts on sustained attention to response*. Neuropsychological Rehabilitation. 2006;16(6), 653-665.

28 Robertson IH, Murre JM. *Rehabilitation of brain damage: brain plasticity and principles of guided recovery*. Psychological Bulletin. 1999;125(5), 544-575. doi.org/10.1037/0033-2909.125.5.544

29 Katzev M, Tüscher O, Hennig J, et al. *Revisiting the functional specialization of left inferior frontal gyrus in phonological and semantic fluency: the crucial role of task demands and individual ability*. J Neurosci. 2013;33(18):7837-7845. doi:10.1523/JNEUROSCI.3147-12.2013.

30 Luo L, Luk G, Bialystok E. *Effect of language proficiency and executive control on verbal fluency performance in bilinguals*. Cognition. 2010;114(1):29-41. doi:10.1016/j.cognition.2009.08.014.

31 Shao Z, Janse E, Visser K, Meyer AS. *What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults.* Front Psychol. 2014;5:772. Published 2014 Jul 22. doi:10.3389/fpsyg.2014.00772

32 Pearson Assessment. *Advanced Clinical Solutions for WAIS-IV and WMS-IV: Clinical and Interpretive Manual.* San Antonio, Texas: NCS Pearson Inc. 2009.

33 Wechsler D. *Wechsler Memory Scale – Fourth Edition (WMS–IV) technical and interpretive manual*. San Antonio, TX: Pearson. 2009.

34 Nelson HE. *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual*. Windsor: NFER-Nelson. 1982.

35 Nelson HE, Willison J. *National Adult Reading Test (NART).* Windsor: NFER-Nelson. 1991.

36 Wechsler D. *Wechsler Test of Adult Reading: WTAR.* San Antonio, TX: Psychological Corporation. 2001

37 Julayanont P, Brousseau M, Chertkow H, et al. *Montreal Cognitive Assessment Memory Index Score (MoCA-MIS) as a predictor of conversion from mild cognitive impairment to Alzheimer's disease*. J Am Geriatr Soc. 2014;62(4):679-684. doi:10.1111/jgs.12742

38 Kamarajan C, Ardekani BA, Pandey AK, et al. *Random Forest Classification of Alcohol Use Disorder Using fMRI Functional Connectivity, Neuropsychological Functioning, and Impulsivity Measures*. Brain Sci. 2020;10(2):115. Published 2020 Feb 20. doi:10.3390/brainsci10020115

39 Pandey AK, Ardekani BA, Kamarajan C, et al. *Lower Prefrontal and Hippocampal Volume and Diffusion Tensor Imaging Differences Reflect Structural and Functional Abnormalities in Abstinent Individuals with Alcohol Use Disorder.* Alcohol Clin Exp Res. 2018;42(10):1883-1896.

40 Bagga D, Sharma A, Kumari A, et al. *Decreased white matter integrity in fronto-occipital fasciculus bundles: relation to visual information processing in alcohol-dependent subjects*. Alcohol. 2014;48(1):43-53. doi:10.1016/j.alcohol.2013.10.009

41 Warrington EK, James M. *The visual object and space perception battery*. Suffolk: Bury St. Edmunds. 1991.

42 Kasse E, Oudman E, Olivier M, Wijnia JW, Postma A. *Subtle object location perception deficits in Korsakoff's syndrome*. J Clin Exp Neuropsychol. 2019;41(9):881-887. doi:10.1080/13803395.2019.1640864

43 Gauthier L, Dehaut F, Joanette Y. *The Bells Test: A quantitative and qualitative test for visual neglect*. Int J Clin Neuropsyc. 1989;11(2), 49-54.

44 Baldo JV, Shimamura AP. *Letter and category fluency in patients with frontal lobe lesions*. Neuropsychology. 1998;12(2):259-267. doi:10.1037//0894-4105.12.2.259

45 Platt B, O'Driscoll C, Curran VH, et al. *The effects of licit and illicit recreational drugs on prospective memory: a meta-analytic review.* Psychopharmacology (Berl). 2019;236(4):1131-1143. doi:10.1007/s00213-019-05245-9

46 Rendell PG, Mazur M, Henry JD. *Prospective memory impairment in former users of methamphetamine*. Psychopharmacology (Berl). 2009;203(3):609-616. doi:10.1007/s00213-008-1408-0

47 Poreh AM. *The quantified process approach: an emerging methodology to neuropsychological assessment*. Clin Neuropsychol. 2000;14(2):212-222. doi:10.1076/1385-4046(200005)14:2;1-Z;FT212

48 Poreh AM. *The quantified process approach to neuropsychological assessment. Studies on neuropsychology, neurology, and cognition (1st ed).* New York: Taylor & Francis. 2006. Available from: www.loc.gov/catdir/toc/ecip064/2005035530.html

49 Griffin PT, Heffernan A. *Digit span, forward and backward: separate and unequal components of the WAIS digit span.* Percept Motor Skill. 1983;335-338.

50 Iverson GL, Tulsky DS. *Detecting malingering on the WAIS-III. Unusual digit span performance patterns in the normal population and in clinical groups*. Arch Clin Neuropsychol. 2003;18(1):1-9.

51 Laures-Gore J, Marshall RS, Verner E. *Performance of individuals with left-hemisphere stroke and aphasia and individuals with right brain damage on forward and backward digit span tasks*. Aphasiology. 2011;25(1):43-56. doi:10.1080/02687031003714426

52 Parada M, Corral M, Mota N, et al. *Executive functioning and alcohol binge drinking in university students.* Addict Behav. 2012;37(2):167-172. doi:10.1016/j.addbeh.2011.09.015

53 Gladsjo JA, Schuman CC, Evans JD, et al. *Norms for letter and category fluency: demographic corrections for age, education, and ethnicity*. Assessment. 1999;6(2):147-178. doi:10.1177/107319119900600204

54 Vaughan RM, Coen RF, Kenny R, Lawlor BA. *Preservation of the semantic verbal fluency advantage in a large population-based sample: normative data from the TILDA study*. J Int Neuropsychol Soc. 2016;22(5):570-576. doi:10.1017/S1355617716000291

55 Henry JD, Crawford JR. *A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology.* 2004;18(2):284-295. doi:10.1037/0894-4105.18.2.284

56 Luria AR. *Higher cortical functions in man*. New York: Basic Books. 1966.

57 Cunha PJ, Nicastri S, de Andrade AG, Bolla KI. *The frontal assessment battery (FAB) reveals neurocognitive dysfunction in substance-dependent individuals in distinct executive*

domains: Abstract reasoning, motor programming, and cognitive flexibility. Addict Behav. 2010;35(10):875-881. doi:10.1016/j.addbeh.2010.05.005

58 Fontes MA, Bolla KI, Cunha PJ, et al. *Frontal Assessment Battery (FAB) is a simple tool for detecting executive deficits in chronic cannabis users.* J Clin Exp Neuropsychol. 2011;33(5):523-531. doi:10.1080/13803395.2010.535505

59 Bickel WK, Marsch LA. *Toward a behavioral economic understanding of drug dependence: delay discounting processes.* Addiction. 2001;96(1):73-86. doi:10.1046/j.1360-0443.2001.961736.x

60 Bickel WK, Yi R, Landes RD, Hill PF, Baxter C. *Remember the future: working memory training decreases delay discounting among stimulant addicts*. Biol Psychiatry. 2011;69(3):260-265. doi:10.1016/j.biopsych.2010.08.017

61 Businelle MS, McVay MA, Kendzor D, Copeland A. *A comparison of delay discounting among smokers, substance abusers, and non-dependent controls*. Drug Alcohol Depend. 2010;112(3):247-250. doi:10.1016/j.drugalcdep.2010.06.010

62 Madden GJ, Bickel WK, Jacobs EA. *Discounting of delayed rewards in opioid-dependent outpatients: exponential or hyperbolic discounting functions?*. Exp Clin Psychopharmacol. 1999;7(3):284-293. doi:10.1037//1064-1297.7.3.284

63 Vuchinich RE, Simpson CA. *Hyperbolic temporal discounting in social drinkers and problem drinkers*. Exp Clin Psychopharmacol. 1998;6(3):292-305. doi:10.1037//1064-1297.6.3.292

64 de Wit H, Flory JD, Acheson A, McCloskey M, Manuck SB. *IQ and nonplanning impulsivity are independently associated with delay discounting in middle-aged adults*. Pers Individ Dif. 2007;42(1), 111–121. https://doi.org/10.1016/j.paid.2006.06.026

65 Johnson MW, Bickel WK. *Within-subject comparison of real and hypothetical money rewards in delay discounting*. J Exp Anal Behav. 2002;77(2):129-146. doi:10.1901/jeab.2002.77-129

66 Australian Government Department of Education, Skills and Employment. *Australian Qualifications Framework*. Canberra: Australian Government Department of Education, Skills and Employment; 2013. Available from: https://www.aqf.edu.au/

67 Manchester D, Priestley N, Jackson H. *The assessment of executive functions: coming out of the office*. Brain Inj. 2004;18(11):1067-1081. doi:10.1080/02699050410001672387

68 Chaytor N, Schmitter-Edgecombe M. *The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills*. Neuropsychol Rev. 2003;13(4):181-197. doi:10.1023/b:nerv.0000009483.91468.fb

69 Golden C. *Stroop color and word test: manual for clinical and experimental uses*. Chicago: Stoelting. 1978.

70 Craik FIM, Bialystok E, Gillingham S, et al. *Alpha span: A measure of working memory*. Can J Exp Psychol. 2018;72(3):141-152. doi:10.1037/cep0000143

71 Goebel S, Fischer R, Ferstl R, et al. *Normative data and psychometric properties for qualitative and quantitative scoring criteria of the Five-point Test*. Clin Neuropsychol. 2009;23(4):675-690. doi:10.1080/13854040802389185

72 Morrow SA. *Normative data for the Stroop color word test for a North American population*. Can J Neurol Sci. 2013;40(6):842-847. doi:10.1017/s0317167100015997

73 Lezak MD. *Neuropsychological assessment*. Oxford; New York: Oxford University Press. 2012.

74 Strauss E, Sherman EMS, Spreen O. *A compendium of neuropsychological tests: administration, norms, and commentary*. Oxford; New York: Oxford University Press. 2006.

75 Berry J, Lunn J, Sedwell A, et al. *Technical, Administration and Scoring Manual for the ACE Screening Tool*. Sydney. Agency of Clinical Innovation. 2020.

76 Ponterotto JG, Ruckdeschel DE. *An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures*. Percept Mot Skills. 2007;105:997-1014.

77 Jurado MB, Rosselli M. *The elusive nature of executive functions: a review of our current understanding*. Neuropsychol Rev. 2007;17(3):213-233. doi:10.1007/s11065-007-9040-z

78 Stuss DT. Functions of the frontal lobes: relation to executive functions. J Int Neuropsychol Soc, 2011;17:759-765. doi:10.1017/S1355617711000695

79 Koo TK, Li MY. *A guideline of selecting and reporting intraclass correlation coefficients for reliability research*. J Chiropr Med. 2016;15(2): 155-163. doi:10.1016/j.jcm.2016.02.012