Research paper

Internet-based cognitive assessment tool: Sensitivity and validity of a new online cognition screening tool for patients with bipolar disorder

K.W. Miskowiak [a,b,*], A.E. Jespersen [a,b], K. Obenhausen [a,b], P. Hafiz [c,d], E. Hestbæk [a,b], L. Gulyas [d], L.V. Kessing [b], J.E. Bardram [c,d]

[a] Department of Psychology, University of Copenhagen, Copenhagen, Denmark
[b] Copenhagen Affective Disorder Research Centre (CADIC), Psychiatric Centre Copenhagen, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, DK-2100, Copenhagen, Denmark
[c] Digital Health Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark
[d] Copenhagen Center for Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

ABSTRACT

Background: The International Society for Bipolar Disorders Targeting Cognition Task Force recommends the Screen for Cognitive Impairment in Psychiatry (SCIP) to screen for cognitive impairment in bipolar disorder. However, SCIP must be administered by a healthcare professional, which is often impossible due to time and resource constraints. Web-based, self-administered cognition screening tools may enable assessment and monitoring of patients' cognition at a much larger scale to a reduced cost. For this purpose, we developed the Internet-Based Cognitive Assessment Tool (ICAT) as a modified web-based version of SCIP. This study aimed to investigate the sensitivity and validity of ICAT for cognition assessment in bipolar disorder.
Method: Thirty-five patients with bipolar disorder in full or partial remission and 35 healthy controls completed ICAT on a computer, the standard paper-and-pencil SCIP and a subjective cognition questionnaire and were rated for mood symptoms and functioning at the Copenhagen Affective Disorders Research Centre.
Results: Patients displayed cognitive impairments compared to controls on the ICAT (t (61)=3.67, p<.001, d=0.93). There was a strong correlation between ICAT and SCIP Total Scores (r(61)=.72, p<.000) and moderate to strong correlations on subtest scores (r=.48-.63, ps<.001). Across all participants, lower ICAT scores correlated with more subjective cognitive complaints (r(59)=-.43, p<.001) and poorer psychosocial functioning (r (62)=-.47, p<.001).
Conclusion: ICAT is a sensitive and valid web-based tool for cognition assessment in patients with bipolar disorder. This highlights ICAT as a novel web-based cognition screening tool that is feasible for largescale assessment and monitoring of cognition in the clinical management of bipolar disorder.

1. Background

Cognitive impairment is a core feature of affective disorders, including unipolar depression (UD) and bipolar disorder (BD) (Bourne et al., 2013; Burdick et al., 2014; Rock et al., 2014). These impairments are evident across several cognitive domains, including memory, attention, and executive functions and often persist during periods of remission (Bora et al., 2013; Cullen et al., 2016; Demmo et al., 2017; Rock et al., 2014). Cognitive impairments contribute to poor quality of life and impede patients' functional recovery including work capacity, which poses the largest socio-economic burden of these disorders

(McIntyre et al., 2013; Olesen et al., 2012). Notwithstanding the importance of cognitive functions for patients' daily functioning and societal costs, cognition is often not assessed in the clinic. This is due to lack of consensus on how to assess cognition and restricted time and resources for such assessments (Miskowiak et al., 2018; Miskowiak et al., 2017). While some clinicians assess cognition by simply asking patients whether they *experience* any cognitive challenges, this may not provide accurate insights into patients' cognitive functions because of a general poor correlation between subjective and objective measures of cognition (Demant et al., 2015). This means that it is not necessarily the patients with most subjective cognitive difficulties who display largest

---

* Corresponding author.
  E-mail address: Kamilla@miskowiak.dk (K.W. Miskowiak).

performance decline on objective cognitive tests and vice versa. Indeed, various factors influence this discrepancy, including the severity of mood symptoms, illness duration, premorbid intelligence, and age (Miskowiak et al., 2016; Petersen et al., 2019). The International Society for Bipolar Disorders (ISBD) Targeting Cognition Task Force therefore recommends clinical implementation of the validated and sensitive cognition screening tool, the Screen for Cognitive Impairment in Psychiatry (SCIP) that assesses several cognitive domains despite its brevity (<20 minutes). However, an impediment to its clinical implementation is that it must be administered by a clinician, which requires training and resources in terms of time and space. This highlights a need for the development of *web-based* patient self-administered cognition screening tools that offer brief, valid and reliable remote testing of patients' objective cognitive functions.

Web-based remote testing options exist for a few cognition test batteries, including CNS Vital Signs (CNSVS; Gualtieri and Johnson, 2006) and CogState (Davis et al., 2017), which are validated in neuropsychiatric populations, including dementia, attention deficit hyperactivity disorder (ADHD) and UD. However, these tools may be suboptimal for detection of more subtle cognitive impairments in affective disorders (Davis et al., 2017) and depend on relatively expensive software (Parsons, 2016). In contrast, the THINC-integrated tool (THINC-it; McIntyre et al., 2017) and MyCognition Quotient (MyCQ; Domen et al., 2019) are web-based tools that have been made freely available and thus offer more accessible cognition assessments to patients. While MyCQ was developed to be used transdiagnostically, THINC-it is the first web-based cognition screening tool designed for patients with affective disorders. THINC-it is sensitive to cognitive impairment in patients with UD in an acute illness-phase (McIntyre et al., 2017) and has high reliability in healthy controls (HC) (Harrison et al., 2018). However, only two of four subtests of THINC-it showed acceptable concurrent validity (Harrison et al., 2018). Another limitation of THINC-it - and MyCQ - is that they lack an assessment of verbal learning and memory, which is likely due to difficulties with online speech recognition (Hafiz et al., 2019). Nevertheless, verbal learning and memory assessments should ideally be integrated into future web-based screening tools because impairments in this domain are common even during remitted states of affective disorders (Arts et al., 2008; Bora et al., 2008) and contribute to poor occupational and daily functioning (Bonnín et al., 2010; López-Villarreal et al., 2020; Tse et al., 2014).

We designed and piloted a web-based self-administered objective screening tool for remote testing; the *Internet-based Cognitive Assessment Tool* (ICAT). The ICAT was designed to resemble the SCIP assessment approach (Miskowiak et al., 2018) and includes real-time assessment of verbal learning and recall using Automatic Speech Recognition (ASR) technology. Version 1 of ICAT showed adequate feasibility and validity in a pilot study on *n*=19 healthy volunteers (Hafiz et al., 2019) and based on results and insights derived from this study, an improved version of ICAT was designed and implemented. The aims of the present study were to investigate (i) the sensitivity of ICAT to cognitive impairments in partially or fully remitted patients with BD, and (ii) to establish the concurrent validity of the ICAT as compared to the SCIP. In addition, we assessed the association between ICAT performance, psychosocial function and subjective cognition, and estimated the preliminary tentative cut-off scores for cognitive impairment for ICAT. Finally, the usability of the ICAT was examined. We hypothesized (i) that patients with BD in full or partial remission would display impaired performance as assessed by ICAT compared to healthy controls (HC), and (ii) that cognitive performance as assessed by ICAT would correlate with cognitive performance as assessed with the corresponding paper-and-pencil-based SCIP test.

2. Methods

2.1. Participants and recruitment

Patients with BD and HC were recruited for the study between May 2019 and March 2020. Patients were recruited through an ongoing longitudinal cohort study of patients with newly diagnosed BD, the Bipolar Illness Onset (BIO) study (Kessing et al., 2017) and were referred by specialized psychiatrists at the outpatient Copenhagen Affective Disorder Clinic, Psychiatric Centre Copenhagen. Patients had an ICD-10 diagnosis of BD and were in full or partial remission upon inclusion, as reflected by scores ≤ 14 on the Hamilton Depression Rating Scale 17-items (HRDS-17; Hamilton, 1960) and on the Young Mania Rating Scale (YMRS; Young et al., 1978). Diagnostic assessment was carried out using the diagnostic interview Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al., 1990). The HC were recruited from the Blood Bank at Copenhagen University Hospital. All participants were 18-60 years of age and had Danish as their mother tongue. Exclusion criteria included any diagnosed comorbid neurological disorder, severe somatic illness, or current substance abuse disorder. Patients were further excluded if they had a daily use of benzodiazepines >22.5 mg oxazepam or >7.5 mg diazepam or if they have had electroconvulsive therapy within the past three months. HC were excluded if they had a personal or first-degree family history of psychiatric illness. Written informed consent was obtained prior to study enrolment. The local ethics committee and data protection agency in the Capital Region of Denmark approved the study (protocol numbers: H-7-2014-007 and RHP-2015-023).

2.2. Procedure

Participants attended the Psychiatric Centre Copenhagen on two occasions. On the first visit, participants were assessed for subjective cognitive functions and functional capacity (details below). On the second visit, participants were assessed with the paper-and-pencil-based screening SCIP test and subsequently completed the ICAT test on a 13" laptop. For this second assessment, participants had been instructed to avoid caffeine intake. The ICAT was set up to be fully self-administered and included a short introduction video with instructions for the test. To investigate the accuracy of the ICAT automatic speech recognition (ASR) component, participants' verbal responses were also recorded by the assessor. Finally, participants provided feedback on the system usability of the ICAT. Premorbid verbal intelligence was estimated with the Danish Adult Reading Test (DART; Crawford et al., 1987). All participants were rated on depressive and manic symptoms using the HDRS-17 and YMRS, respectively.

3. Materials

3.1. Internet-based Cognitive Assessment Tool

The Internet-based Cognitive Assessment Tool (ICAT) is a web-based cognitive test battery designed to resemble the cognitive tasks of the Screen for Cognitive Impairment in Psychiatry (SCIP; Purdon, 2005). The ICAT is fully self-administered and requires no involvement by an assessor. Participants are introduced to the application through a homepage and fill out an informed consent form compliant with the European data protection law (general data protection regulation, GDPR). Since the ICAT uses Automatic Speech Recognition (ASR) to assess verbal recall, a technical setup page ensures that the microphone and speaker are correctly configured. A detailed description of the ICAT design process, system descriptions, and the use of ASR in cognitive assessment is reported in Hafiz et al. (2019).

Four of the five cognitive tests in ICAT were designed based on the subtests of SCIP. A detailed description of the ICAT tests and their paper-and-pencil counterpart can be found in the supplementary material,

Table S1. For each test in ICAT, the participant is provided with a detailed guide consisting of pre-recorded video-and audio instructions designed to closely resemble the face-to-face instruction normally issued by a human assessor. An assessment in ICAT takes approximately 35 minutes, with the longer duration than SCIP being due to the introductory video, both auditory and written instructions before each test and longer duration of the adapted LNS than the SCIP verbal fluency test. The ICAT is comprised of five subtests: (i) List Learning (LL; Fig. 1. A), (ii) Consonant repetition (CR; Fig. 1.B), modified from the Weschler Adult Intelligence Scale III: (iii) letter-number sequencing task (adapted
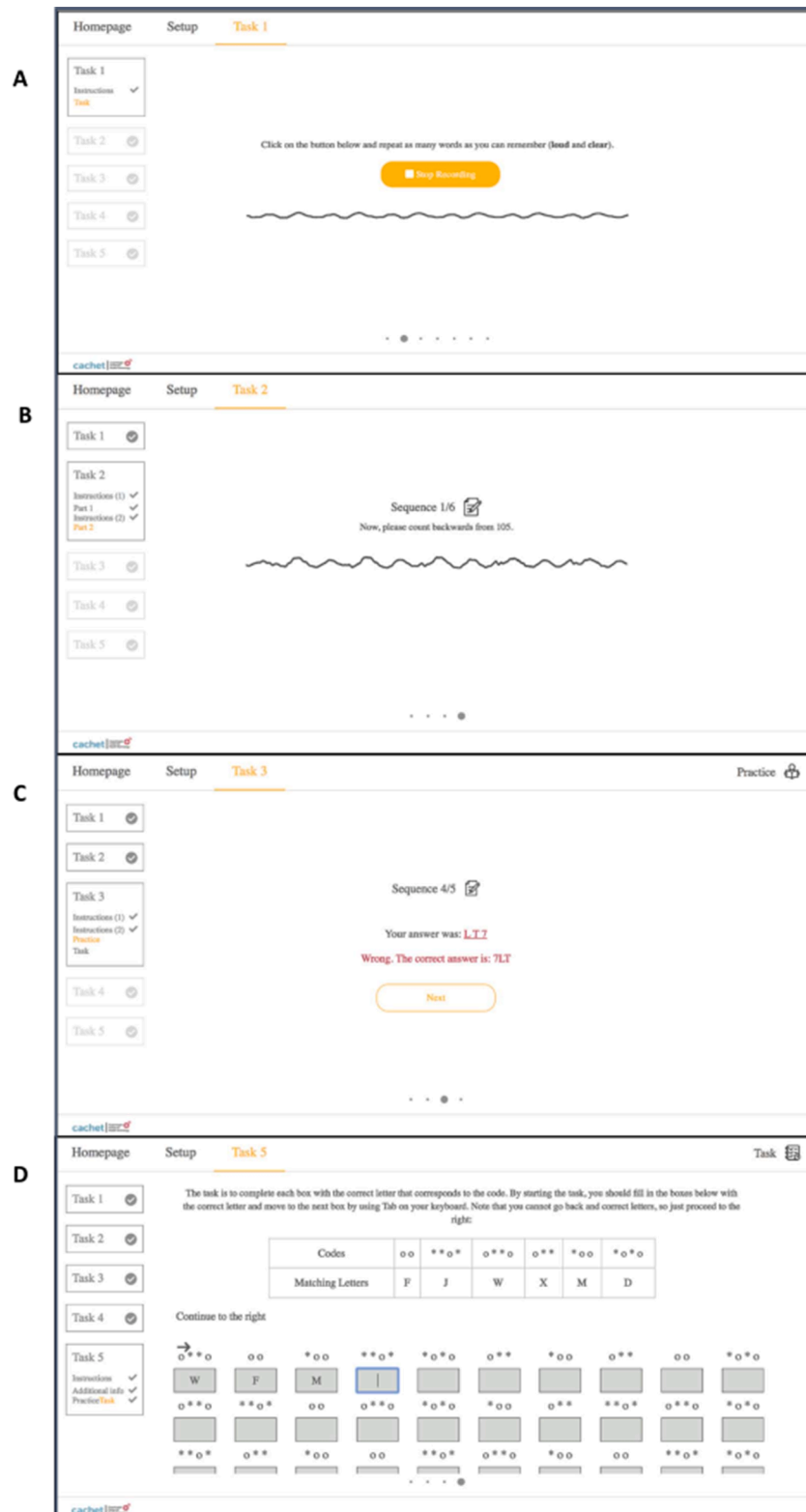

Fig. 1. A. The speech interface of the ICAT List Learning and Delayed List Learning subtests. B The interface of the ICAT Consonant Repetition subtest. C. A practice sequence in the ICAT adapted WAIS Letter/number-sequencing subtest. D. The interface of the ICAT Visuomotor Tracking subtest.

LNS, Fig. 1.C), (iv) Delayed List Learning (DLL; Fig. 1.A), and (v) Visuomotor Tracking (VMT; Fig. 1.D). Together, the ICAT tests probe verbal learning (LL), two aspects of working memory: 'transient online storage and retrieval' and 'executive function working memory' (CR and adapted LNS), delayed verbal memory (DLL) and psychomotor speed (VMT), respectively. The adapted LNS test replaced the SCIP Verbal Fluency subtest, since a digital format of verbal fluency could not be adequately implemented due to insufficient accuracy of ASR for rapidly uttered words with sometimes only slight pronunciation differences (Hafiz et al., 2019).

3.2. Revision of tests for ICAT version 2

The ICAT was optimized from its original version published in Hafiz et al. (2019) through improvement of the DLL and CR tests. In our feasibility study, we had found no association between the ICAT DLL and SCIP Verbal Learning Test (VLT) -Delayed (Hafiz et al., 2019). This was due to: (i) participants uttering words too fast, quiet, and far away from the microphone and (ii) low accuracy of the ASR to certain words. To mitigate such limitations to the automatic speech recognition technology, updated video and audio instructions were included which emphasized how the test participant should speak clearly and separate words with small pauses in-between. To further improve the precision of the speech recognition, we updated the list of words provided in the SCIP-manual (Purdon, 2005) with a new list that yielded a higher accuracy in the speech recognition component. The new words for the revised ICAT word list were chosen from the Rey Auditory Verbal Learning Test (RAVLT) and had an approximately equal frequency in the Danish language compared to the old ICAT list to ensure a similar difficulty level. Another benefit of this approach was that the revised word list had no overlap with words on the SCIP form 3 administered in the study.

The ICAT CR test originally included a drag-and-drop number sorting module - instead of having the participant count backwards as in the SCIP - as a distraction before recalling the letter sequence. However, no association was observed between the ICAT CR test and the SCIP Working Memory Test (WMT) (Hafiz et al., 2019), possibly due to lower cognitive load of the drag-and-drop number sorting distractor task. We therefore redesigned the distraction task to include a speech interface, with participants counting backwards out loud as in the SCIP WMT. To ensure that participants did in fact count backwards, they were told by the system that their responses were recorded (even though they were not). Further, during this "recording" a sound wave was shown to illustrate that sound was being recorded (Fig. 1.B).

3.3. Manual neuropsychological tests

Participants were assessed with the validated cognition screening test, SCIP – Danish version (SCIP-D; Jensen et al., 2015; Purdon, 2005). The SCIP is a short (<20 minutes) paper-and-pencil test that comprises of five subtests: Verbal Learning Task - Immediate (VLT-I), Working Memory Task (WMT), Verbal Fluency Task (VFT), Verbal Learning Task – Delayed (VLT-D), and Psychomotor Speed Task (PST) (Purdon, 2005). They were also assessed with the Wechsler Adult Intelligence Scale III: Letter-Number Sequencing test (Wechsler and Psychological, 1997).

3.4. Assessment of functional capacity and subjective cognitive complaints

Participants' functioning was assessed with the Functioning Assessment Short Test (FAST; Rosa et al., 2007), a brief observer-based rating scale that patients complete with a trained clinician. The FAST is designed to measure main functional impairments typically experienced by patients with BD and is recommend as a functional capacity assessment tool by the ISBD Targeting Cognition Task Force (Miskowiak et al., 2017). Participants also filled out the Cognitive Complaints in Bipolar Disorder Rating Assessment (COBRA; Rosa et al., 2013), a brief

questionnaire assessing subjective cognitive difficulties in daily life situations.

3.5. Assessment of usability

System usability of the revised ICAT was assessed with the Poststudy System Usability Questionnaire (PSSUQ), assessing computer application usability satisfaction along four dimensions: overall usability, system usefulness, information quality, and interface quality (Lewis, 1995). Upon completing the questionnaire, qualitative feedback was collected from patients who were asked to further comment on their experience of using ICAT.

3.6. Statistical analyses

Statistical analyses were carried out using IBM SPSS statistics 25 for windows (Field, 2013). Statistical significance for all analyses was set to an alpha-level of $p<.05$ (two-tailed). The two groups were compared using independent samples t-test for normally distributed data. If the assumption of normality was violated, non-parametric Mann-Whitney U tests were carried out instead. Groups were compared on demographic (age, years of education) and clinical variables (subsyndromal depressive and manic symptoms) as well as on premorbid intellectual ability estimated from participants' error score on the DART using the following formula (IQ estimate=128-(0.83*DART error score) as proposed by Nelson and Willison (1991). A $\chi 2$ test was applied to investigate any potential differences in gender distribution between the two groups.

The sensitivity of the ICAT and SCIP tools for assessment of cognitive impairment was assessed for each subtest, as well as for the total score of each instrument. For independent sample t-tests, Cohen's d was calculated as a measure of effect size. For Mann-Whitney U tests, r was calculated as an appropriate effect size using the formula $r=Z / \sqrt{N}$ (Fritz et al., 2011). Significant between-group differences were followed up with post-hoc ANCOVAs with ICAT or SCIP score (total or subtest scores) as the dependent variable, group (BD vs. HC) as the independent variable, and any demographic or clinical variables on which the two groups differed as covariates.

The concurrent validity of the ICAT compared to SCIP was investigated with correlation analyses using Pearson's r or Spearman's σ for normally and non-normally distributed data, respectively. Associations were investigated between the total scores of ICAT and SCIP, as well as between performance on each of the corresponding subtests. Performance on the ICAT adapted WAIS LNS subtest was also correlated with the original pen-and pencil version. For the ICAT LL and DLL subtests, words recalled by the ICAT ASR component were correlated with manual transcripts recorded by the assessor. Finally, objective neuropsychological performance was correlated with observer-based measures of functional capacity and subjective cognitive impairments, respectively. Significant associations were followed up by partial correlations controlling for any demographic or clinical variables in which the groups differed.

Tentative cut-off scores for cognitive impairments were estimated for ICAT based on the standard deviation (SD) of the HCs. The tentative cut-off scores were established at performance level$\geq$1 SD and$\geq$0.5 SD below the normative mean for the subtests and the total score, respectively, as recommended by the ISBD Task Force Targeting Cognition for clinically relevant thresholds for cognitive impairments in BD (Miskowiak et al., 2018).

Word error rate (WER) was used as the primary performance measure of the ICAT ASR component using the formula WER=(S+D+I)/N, where N is the total number of words, D is the number of deletions, S is the number of substitutions, and I is the number of insertions. WER was calculated manually by comparing ASR transcripts to verbal responses recorded during the ICAT LL and DLL subtests.

4. Results

4.1. Demographics and clinical variables

Demographic and clinical data is presented in Table 1. We included $n=35$ participants with BD (26 females; age, mean ± standard deviation [SD]: 31.7±9.0) and $n=35$ HC (21 females; age, mean ± SD: 28.4±6.0). Comparisons between BD and HC revealed that the two groups were comparable for age, gender, education, and premorbid IQ ($p_s≥.092$). Patients with BD were in full or partial remission, but nevertheless they displayed more subsyndromal symptoms of depression ($U=148.0$, p<.001, $r=0.68$) and mania ($U=444.0$, p=.018, $r=0.28$) compared to HCs.

4.2. Discriminative Ability of the ICAT

The BD group displayed a lower ICAT total performance score than HC with a large effect size ($t(61)=3.67$, p=.001, $d=.93$; Table 2 and Fig. 2.A). Adjusting for subsyndromal mood symptoms, this group difference was reduced to a strong trend ($F(1,61)=3.91$, p=.053, $\eta p^2=.06$). Assessments of the individual ICAT sub-tests revealed that he BD group displayed moderately impaired performance on LL ($t(65)=2.34$, $p=.001$, $d=.66$), which rendered non-significant after adjustment for subsyndromal mood symptoms ($F(1, 65)=.85$, p=.361, $\eta p^2=.01$). The BD group also displayed moderately impaired performance on LNS ($t(67)=2.45$, p=.02, $d=.62$), which prevailed after adjusting for subsyndromal mood symptoms ($F(1, 67)=4.78$, p=.03, $\eta p^2=.07$). Patients also showed moderately impaired VMT performance ($t(66)=2.49$, p=.02, $d=.58$), which was reduced to a trend-level after adjusting for subsyndromal mood symptoms ($F(1, 66)=2.92$, p=.09, $\eta p^2=.044$). In contrast, BD patients showed no performance reductions on CR and DLL ($p_s≥.17$; Table 2, Fig. 2.B). Notably, 23% of patients achieved a maximum score on the CR, which was indicative of ceiling effects on this test (see Table 3). Post hoc statistical power calculations showed that the power (i.e., the probability of correctly rejecting the null hypothesis) was high for the ICAT total score ($1-\beta=.95$), moderate for the ICAT LL ($1-\beta=.75$), LSN sequencing ($1-\beta=0.71$) and VMT ($1-\beta=.65$), but low for the CR ($1-\beta=.45$) and DLL subtests ($1-\beta=.31$).

Table 1
Demographic and clinical characteristics of the BD and HC group.

	Group		
	HC (n = 35)	BD (n = 35)	p-value
Demographic and clinical variables			
Age, mean (SD)	28.5 (6.0)	31.7 (9.0)	.092
Years of education, mean (SD)	16.2 (2.2)	16.3 (5.9)	.233
Est. premorbid intellectual ability, mean (SD)[a]	112.8 (6.2)	112.8 (5.6)	.869
Gender, no women (%)	21 (60.2)	26 (74.3)	.203
HDRS-17 baseline, mean (SD)	0.51 (1.1)	4.3 (3.5)	<.001**
YMRS baseline, mean (SD)	0.54 (1.3)	1.6 (2.3)	.018*
Diagnosis, no (%)[b]			
BD type I	-	13 (39)	-
BD type II	-	20 (61)	-
Duration since diagnosis in years, mean (SD)[c]		3.1(3.7)	
Illness duration in years, mean (SD)[c]		9.0 (7.93)	

note. BD: Bipolar Disorder. HC: Healthy Control. SD: Standard Deviation. HDRS-17: 17 item Hamilton Depression Rating Scale. YMRS: Young Mania Rating Scale. * = p < .05 (two-tailed), ** = p < .01 (two-tailed).
[a] Missing data for $n = 2$ BD patients and $n = 2$ HCs.
[b] Missing data for $n = 2$ BD patients.
[c] Duration since diagnosis and illness duration was defined as the time from the first manic, hypomanic, or mixed episode to the time of diagnosis or assessment, respectively.

Table 2
Scores on the Internet-Based Cognitive Assessment Tool (ICAT), corresponding subtests on the Screen for Cognitive Impairment in Psychiatry – Danish (SCIP-D), functional capacity and subjective cognitive functioning between the HC and BD groups.
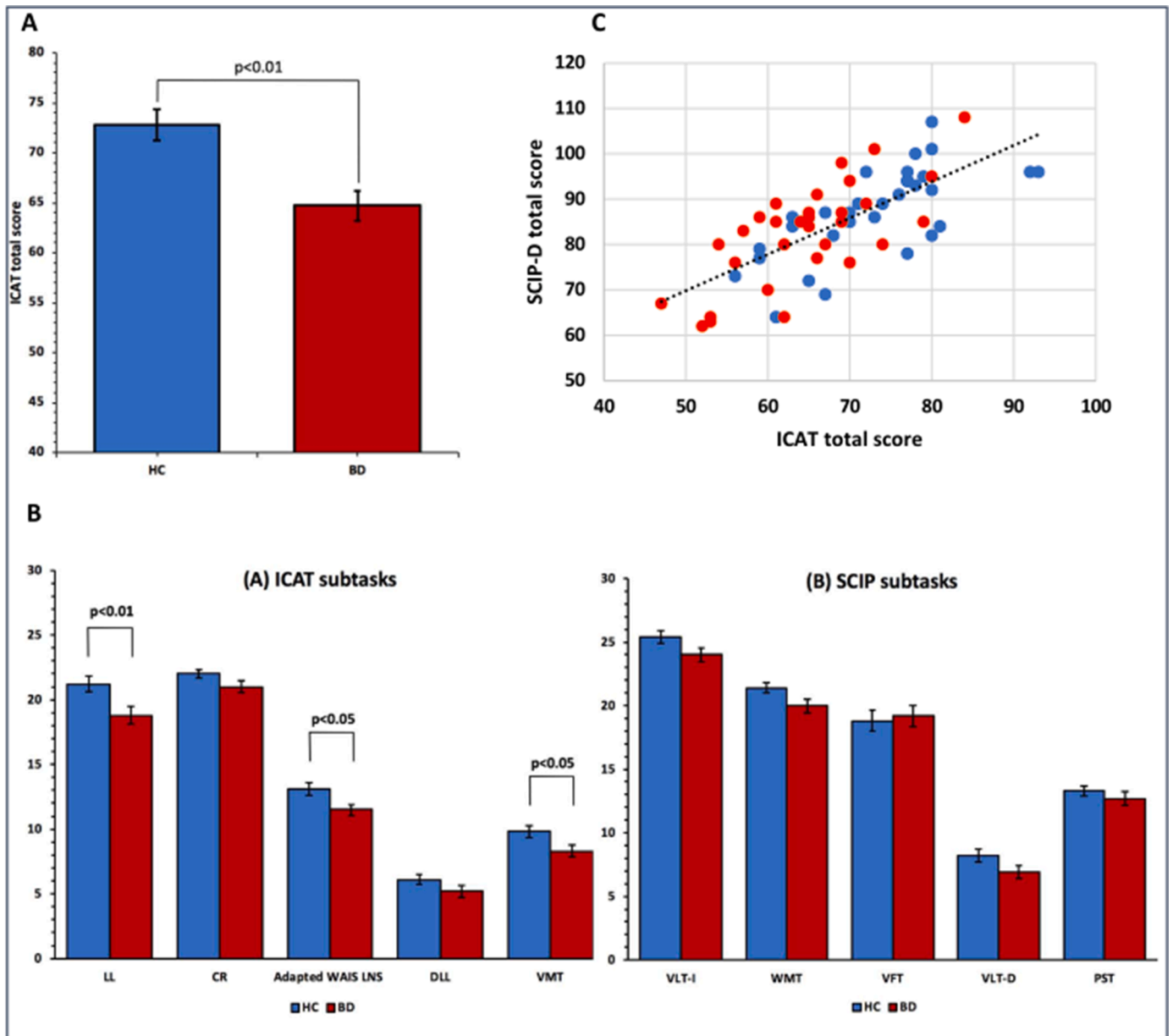
	Group		
	HC (n = 35)	BD (n = 35)	p-value
Internet-based Cognitive Assessment Tool (ICAT)			
List learning (LL), mean (SD)[a]	21.2 (3.5)	18.8 (3.8)	.009**
Consonant repetition (CR), mean (SD)[b]	22.0 (1.8)	21.0 (2.6)	.174
Adapted WAIS IV LNS, mean (SD)[c]	13.1 (2.7)	11.5 (2.5)	.017*
Delayed list learning (DLL), mean (SD)[d]	6.1 (2.4)	5.2 (2.5)	.373
Visuomotor Tracking (VMT), mean (SD)[e]	9.8 (2.5)	8.3 (2.7)	.015*
ICAT total score, mean (SD)[f]	72.8 (8.9)	64.7 (8.6)	.001**
Screen for Cognitive Impairment in Psychiatry (SCIP)			
Verbal Learning Task - immediate (VLT-I), mean (SD)	25.4 (2.8)	24.0 (3.03)	.052
Working Memory Task (WMT), mean (SD)	21.4 (2.2)	20.0 (3.2)	.076
Verbal fluency Task (VFT), mean (SD)	18.8 (4.9)	19.2 (4.9)	.698
Verbal Learning Task - delayed (VLT-D), mean (SD)	8.2 (2.9)	6.9 (2.9)	.144
Psychomotor Speed Task (PST), mean (SD)	13.3 (2.5)	12.7 (3.3)	.189
SCIP total score, mean (SD)	87.2 (9.7)	82.5 (11.2)	.080
Manual neuropsychological measures and transcripts			
WAIS IV LNS, mean (SD)	13.2 (3.2)	11.6 (66)	.033*
ICAT LL manual transcripts, mean (SD)	22.21 (3.4)	20.6 (4.08)	.087
ICAT DLL manual transcripts, mean (SD)	6.5 (2.4)	5.6 (2.5)	.231
Functional capacity and subjective cognitive complaints			
FAST[g]	1.7 (3.2)	14.1 (9.1)	<.001**
COBRA[h]	7.5 (3.8)	20.6 (7.4)	<.001**

Note. HC: Healthy Control. BD: Bipolar Disorder. SD: Standard Deviation. HDRS-17: 17 item Hamilton Depression Rating Scale. YMRS: Young Mania Rating Scale. WAIS IV LNS: Wechsler Adult Intelligence Scale III: Letter-Number Sequencing. FAST: Functional Assessment Short Test. COBRA: Cognitive Complaints in Bipolar Disorder Rating Assessment. * = p < .05 (two-tailed), ** = p < .01 (two-tailed).
[a] Missing data for $n = 2$ BD patients and $n = 1$ HCs.
[b] Missing data for $n = 1$ HC.
[c] Missing data for $n = 1$ HC.
[d] Missing data for $n = 2$ BD patients and $n = 2$ HC.
[e] Missing data for $n = 1$ BD patient and $n = 1$ HC.
[f] Missing data for $n = 4$ BD patient and $n = 3$ HC.
[g] Missing data for $n = 1$ BD patient and $n = 3$ HC.
[h] Missing data for $n = 1$ BD patient.

4.3. Between-group differences on SCIP performance, functioning and cognitive complaints

Patients with BD displayed a non-significant trend towards impairments on the SCIP tests, as reflected by a lower SCIP total score compared to HC ($t(61)=1.78$, p=.08, $d=.45$). Similar non-significant trends towards impaired performance in the BD group were also found for the SCIP VLT-I ($t(65)=1.98$, $p=.052$, $d=.48$) and WMT tests ($U=448.5$, p=.08, $r=.21$). No group differences were observed for the remaining SCIP subtests ($p_s≥.145$; Table 2 and Fig. 2.B). The group differences remained non-significant when adjusting for subsyndromal

Fig. 2. A. Performance on the ICAT total score in the HC and BD groups. Error bars represent standard errors of the mean (SEM). B. Performance on the five ICAT subtests (A) and SCIP subtests (B) in the HC and BD groups. Error bars represent SEM. C. Scatterplot of the correlation between the ICAT and SCIP total scores across the entire sample (r = .72). Patients in the BD group are represented by red dots. Participants in the HC group are represented by blue dots.

mood symptoms ($p_s \geq .28$).

Patients with BD experienced more functional difficulties in everyday life compared to HC with a moderate effect size ($U=56.5$, $p<.001$, $r=.78$; Table 2), which prevailed after adjusting for subsyndromal mood symptoms ($F(1,65)=20.08$, $p<.001$, $\eta p^2=.245$). They also reported more subjective cognitive difficulties than HC ($t(67)=13.43$, $p<.001$, $d=1.65$), which again prevailed after adjustment for subsyndromal mood symptoms ($F(1, 67)=42.51$, $p<.001$, $\eta p^2=.40$).

4.4. Validity of the ICAT

Across the entire sample, we found a strong positive correlation between the total scores of the ICAT and SCIP ($r(61)=.72$, $p<.001$; Fig. 2.C). This association remained significant when adjusting for subsyndromal mood symptoms ($r(61)=.69$, $p<.001$). Analyses also revealed moderate to strong positive correlations between performance on individual ICAT subtests and their corresponding SCIP subtests for

verbal learning and memory (LL vs. VLT-I; ($r(65)=.54$, $p<.001$), working memory (CR vs. WMT; $r(67)=.48$, $p<.001$), delayed working memory (DLL vs. VLT-D; $r(64)=.55$, $p<.001$), and psychomotor speed (VMT vs. PST; $r(66)=.55$, $p<.001$). These associations all prevailed after adjusting for subsyndromal mood symptoms ($p_s< .001$). The ICAT LNS test was weakly correlated with the SCIP Verbal Fluency Task ($r(67)=.27$, p=.03) but displayed a moderate to strong association with the paper-and-pencil version of the WAIS LNS task from which it was adapted ($r(67)=.63$, $p<.001$). The latter correlation also prevailed after adjusting for subsyndromal symptoms (p<.001). Participants generally recalled fewer words from the ICAT than the SCIP word lists in the learning phase (21.4 ± 3.8 vs. 24.7 ± 3.0; paired t-test: $t=7.9$, df=65, $p<.001$) and delayed recall phase (6.0 ± 2.5 vs. 7.6 ± 2.9, $t=4.6$, df=64, p<.001).

Across the entire sample, verbal responses recorded by the ICAT ASR component were strongly correlated with verbal responses recorded manually by the assessor for both the LL ($r(65)=.91$, $p<.001$) and DLL (r

Table 3

Number and percentages of patients and HC achieving a maximum score on the SCIP and ICAT subtests with fixed score range.

	Maximum score achieved (100% accuracy)	
	BD group $n = 35$ No. (%)	HC group $n = 35$ No. (%)
SCIP		
Verbal Learning Task - Immediate	1 (3.0)	3 (8.8)
Working Memory Task	5 (14.3)	5 (14.7)
Verbal Learning Task – Delayed	6 (18.2)	7 (21.2)
ICAT		
List Learning (manual transcripts)[a]	0 (0.0)	3 (8.8)
Consonant Repetition[b]	8 (22.9)	11 (32.4)
Delayed List Learning (manual transcripts)[c]	3 (8.8)	4 (12.1)

ASR: Automatic Speech Recognition; HC: healthy controls.

SCIP data excluded on corresponding subtests for these participants.

[a] Missing data for 2 BD and 1 HC.

[b] Missing data for 1 HC.

[c] Missing data for 2 BD and 2 HC.

(64)=.96, p<.001) subtests (Table 4). Participants' recall and recognition accuracy of the ASR component is presented in Fig. 3. Overall, 1,640 Danish words were received by the ASR component during the ICAT LL and DLL tests with an average word error rate (WER) of 8.8%. For the individual Danish words included on the list, eight out of 10 words had an accuracy >90%, whereas two words ('*dør*' and '*kylling*') had slightly lower but still acceptable ASR accuracy of 85% and 88%, respectively (Fig. 3). Across the LL and DLL, the percent correctly identified words by the ASR was 92%.

4.5. Association between ICAT, functioning and cognitive complaints

Across the entire sample, there was a moderate correlation between lower ICAT total scores and more functional impairments (*r*(59)=-.43, p<.001). This correlation was also significant within the patient group (*r*(30)=-.32, p=.04). In contrast, SCIP total scores were not significantly associated with functional impairments across the entire sample or within the patient group (*p*s≥.10).

Across the entire sample, lower ICAT total scores correlated moderately with more subjective cognitive complaints (*r*(62)=-.47, *p* < .001). Further, lower SCIP total scores correlated weakly with more cognitive complaints (*r*(62)=-.28, p<.03). However, within the patient group, ICAT or SCIP scores showed no correlation with subjective complaints (*p*s≥.20). Finally, more subjective cognitive complaints correlated with greater functional impairments across the entire sample (*r*(64)=.72 p<.001) and in the patient group alone (*r*(31)=.48 p=.01).

4.6. Tentative cut-off scores for cognitive impairment

Based on the performance of the HCs, we propose the cut-off of <68

points for the ICAT total score using a threshold for cognitive impairment on this global cognition measure of 0.5 SD under the HC mean, in line with the ISBD Targeting Cognition Task Force recommendations (Miskowiak et al., 2018). The following cut offs for the individual ICAT sub-tasks: <18 for the LL subtest, <20 for the CR subtest, <10 for the adapted WAIS LNS subtest, <4 for the DLL subtest, and <7 for the VMT subtest, in line with the recommended threshold for cognitive impairment of ≥1 SD under the HC mean (Miskowiak et al., 2018).

4.7. Usability of the ICAT

A total of 67 (96%) participants completed the PSSUQ usability questionnaire. They reported high usability of the ICAT system as measured by their perceived overall usability (*M*=4.1, *SD*=0.5 of max 5 on the PSSUQ). In the qualitative feedback, seven participants reported that it was more challenging to memorize words being read aloud by the ICAT application during the LL subtest compared to the face-to-face format of the SCIP VLT-I subtest. Additionally, four participants reported that the ICAT VMT subtest was more difficult to complete on a keyboard than the paper-and-pencil format of the SCIP PST subtest.

5. Discussion

We investigated the sensitivity and validity of a novel web-based and self-administered cognitive screening tool (the revised ICAT) in fully or partially remitted patients with BD and HC. Consistent with our first hypothesis, patients with BD displayed impaired cognitive performance compared to HCs, as measured by the ICAT total score comprising all five subtests as well as on three out of five subtests tapping into verbal learning and memory, working memory, and psychomotor speed, respectively. This suggests that ICAT is sensitive to cognitive impairments in BD. In line with our second hypothesis, performance on ICAT was strongly correlated with performance on the validated objective screening tool SCIP, indicating that ICAT provides a valid assessment of cognitive functioning. Finally, participants reported high levels of satisfaction with the ICAT system, thus indicating good feasibility and user friendliness.

The strong correlation between the ICAT and SCIP total scores, and between ICAT subtests and their paper-and-pencil counterparts, indicates high concurrent validity of ICAT. This reflects an improvement from ICAT version 1, which showed only moderate correlations with the SCIP in a sample of healthy participants (Hafiz et al., 2019). The ICAT also seems to have higher concurrent validity than the other web-based cognition screening tool designed specifically for affective disorders - the THINC-it - for which total scores showed a moderate correlation (*r*=0.4) with standardized neuropsychological tests and only two of four subtests correlated significantly with neuropsychological tests (Harrison et al., 2018). One explanation is that THINC-it employs a gamified test format that may – in contrast with neuropsychological tests – also tap into reward processing that is closely related to depressive symptoms (Nusslock and Alloy, 2017). The ICAT employs a less gamified design
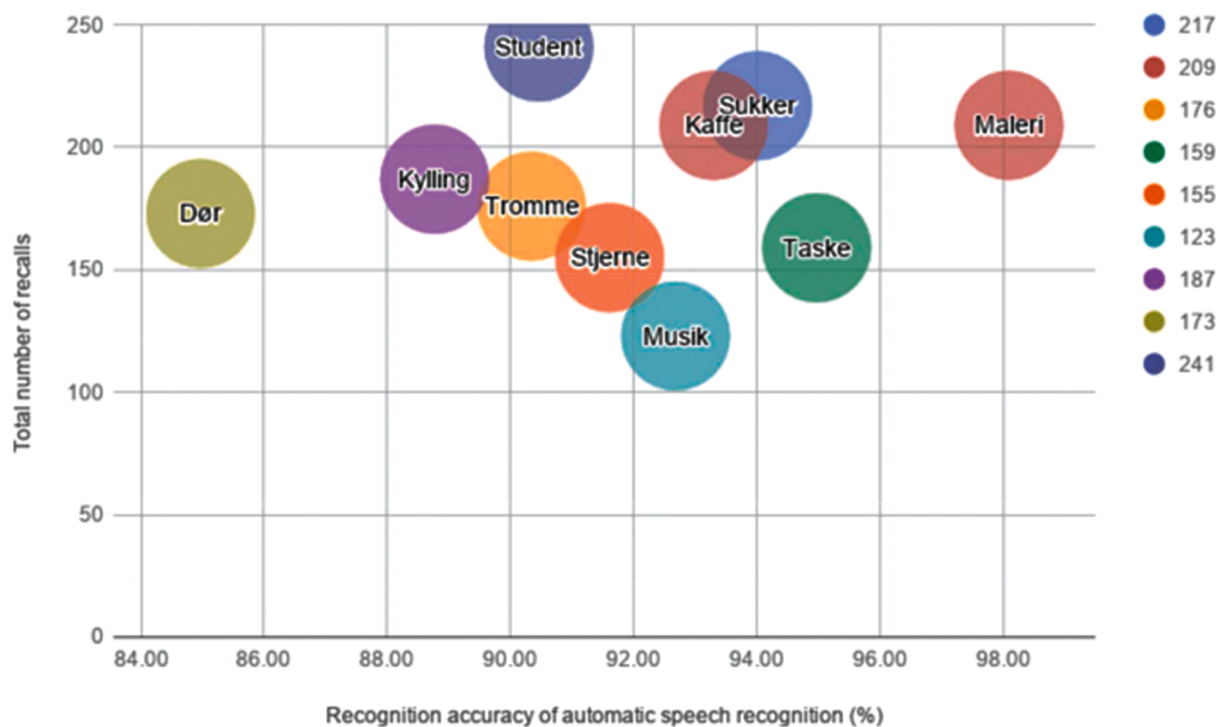
Table 4

ICAT correlations with SCIP-D, manually recorded measures, functional capacity, and subjective cognitive impairment across the entire sample (*n* = 70).

	SCIP subtests						Manually recorded measures			Functional capacity and subjective cognition	
ICAT subtests	VLT-I	WMT	VFT	VLT-D	PST	Total score	LL (manual)	DLL (manual)	WAIS LNS	FAST	COBRA
1. LL (ASR)	.54**	.44**	.36**	.47**	.18	.59**	.91**	.96**	-	-.26*	-.39**
2. CR	.04	.48**	.26	.04	.45**	.38**	-	-	-	-.21	-.31**
3. WAIS LNS	.22	.49**	.27*	.15	.45**	.46**	-	-	.63**	-.25*	-.27*
4. DLL (ASR)	.38**	.31*	.33**	.55**	.13	.51**	-	-	-	-.24	-.17
5. VMT	.37**	.42**	.06	.15	.55**	.44**	-	-	-	-.39**	-.28*
6. Total score	.47**	.62**	.41**	.41**	.50**	.72**	-	-	-	-.43**	-.47**

LL: List Learning. CR: Consonant repetition. WAIS IV LNS: Weschler Adult Intelligence Scale III: Letter-Number sequencing. DLL: Delayed list learning. VMT: Visuomotor Tracking. ASR: Automatic Speech Recognition. VLT-I: Verbal Learning Task - Immediate. WMT: Working Memory Task. VFT: Verbal Fluency Task. VLT-D: Verbal Learning Task - Delayed. PST: Psychomotor Speed Task. FAST: Functional Assessment Short Test. Cognitive Complaints in Bipolar Disorder Rating Assessment.

* = p < .05 (two-tailed), ** = p < .01 (two-tailed).

Fig. 3. Total number of recalls versus the recognition accuracy of the Automatic Speech Recognition component for the Danish words in task 1 (List Learning) and task 4 (Delayed List Learning).

and has greater resemblance with traditional neuropsychological tests, and its results may thus be less influenced by mood symptoms. Another advantage of ICAT compared to other web-based cognitive tests (Domen et al., 2019; McIntyre et al., 2017) is its automatic assessment of verbal recall based on speech recognition. This is an important feature because verbal memory impairments are often pronounced in affective disorders and associated with functional disability (Arts et al., 2008; Bonnín et al., 2010; Tse et al., 2014). Indeed, the ICAT List Learning and Memory tasks – which applies the ASR technology – displayed high concurrent validity and strong correlations with manual transcripts. Notably, participants recalled fewer words from the ICAT than SCIP word lists. This could be due to their greater difficulty with remembering words read aloud from a computer than by the research assistant, from whom they could read mouth movements to assist their learning. It is also possible that their poorer retention of the ICAT words was influenced by their awareness of having to pronounce the words clearly, which could have diminished their attention to memorizing the words.

The finding of a moderate association between psychosocial functioning, measured by the FAST, and the ICAT processing speed test is in accordance with previous reports of associations between this cognitive domain and functioning in BD (Mur et al., 2009). In contrast, the observed significant -albeit only moderate - association between ICAT performance and subjective cognitive difficulties according to the COBRA diverges from previous findings of a poor relationship between objectively and subjectively measured cognitive difficulties (Demant et al., 2015; Van der Elst et al., 2008). A possible explanation is that the self-administered format of ICAT more closely resembles the cognitive challenges that patients must tackle by themselves in their daily life, since completing the ICAT tests involves minimal guidance and expectations from a neuropsychologist. Indeed, guidance and expectations during face-to-face assessment might facilitate motivational processes, which can have a positive confounding influence on test effort during assessment (Greher and Wodushek, 2017).

The finding that patients showed only a trend towards impaired performance on the SCIP diverges from previous studies that have demonstrated good sensitivity of the SCIP to cognitive impairments in

BD (Jensen et al., 2015; Rojo et al., 2010). Notably, 63% of patients in the present study were newly diagnosed with BD (< 3 years) upon study inclusion. Moreover, in the current study, the total illness duration and mean age were lower than in other validation studies of the SCIP in remitted BD (mean illness duration 9 years and mean age 32 years in the present study *vs.* 12 years and 41 years in previous studies) (Rojo et al., 2010). Importantly, longer illness duration has been associated with more cognitive impairments in BD (Cardoso et al., 2015). It is therefore conceivable, that the limited sensitivity of the SCIP and of some of the ICAT subtests was due to the inclusion of recently diagnosed and hence less impaired patients. Indeed, when applying a SCIP total score cut-off point for cognitive impairments in BD of <70, suggested by Jensen et al. (2015) we identified 16% of our patients (*vs.* 6% of HCs) as cognitively impaired compared to 30-59% in other previous validation studies of the SCIP (Jensen et al., 2015; Rojo et al., 2010). Given this, it was remarkable that ICAT was sensitive to cognitive impairment in patients on the total score and on three out of five subtests. Following adjustment for subsyndromal mood symptoms, patients' verbal learning and psychomotor impairments as measured by ICAT rendered non-significant, whereas the significant group difference prevailed on the ICAT working memory test. This finding corroborates with meta-analytic evidence for ameliorated verbal memory and psychomotor speed during remission, whereas executive function deficits are more independent of clinical mood states (Kurtz and Gerraty, 2009).

There are numerous implications of a valid web-based cognitive screening tool. First, assessment of cognition is not systematically implemented in the clinical management of BD, which is partly due to lack of consensus on which cognition screening tools to use, as well as limited resources of health care providers (Miskowiak et al., 2018). Even brief, easily administered tools such as the SCIP require training, time and the facilities and resources for conduct face-to-face assessment. A valid web-based tool such as ICAT can therefore have important implications for assessing objective cognitive functioning in daily settings by enabling remote, self-administered assessment. This enables cost-efficient cognitive screening, providing cognition scores which can be uploaded directly to electronic medical records accessible to clinical

staff in hospitals or practitioners in primary care. Second, ICAT can help monitor patients' cognitive status over time to help detect a potential cognitive decline or improvements following changes in treatment and lifestyle. Third, in research trials investigating pro-cognitive treatments, the possibility to efficiently pre-screen eligible participants for objective cognitive impairments across large geographical distances would represent an immense methodological advancement in recruitment (Miskowiak et al., 2017). Finally, ICAT could enable unprecedented large-scale, longitudinal studies of cognitive functioning in patients with affective disorders in nationwide register-based studies including thousands of participants.

The present study has several limitations. First, the sample size, with 35 patients and 35 healthy controls, was relatively small for this ICAT validation study. Post hoc power calculations showed that the power was high for estimating the sensitivity to cognitive impairment in BD with the ICAT total scores, but only moderate for three (LL, LSN and VMT) and low for two ICAT subtests (CR and DLL). Further, the suggested cut-off scores can only be considered tentative since optimal cut-off scores for cognitive impairments must be determined in a larger sample by means of receiver operating characteristic (ROC) analysis. Second, it was not possible to counterbalance the order of assessment, and all participants therefore completed the SCIP first, which may have introduced a learning effect, such that performance on the ICAT tests was possibly better than it would have been had it not been proceeded by the SCIP tests. However, speaking against this, participants displayed more deficits on the ICAT compared to the SCIP. Third, although the brevity of the ICAT tests represents a clear advantage, the tool does not measure all aspects of cognitive functioning, such as problem solving or shifting aspects of EFs. Moreover, ICAT does not address deficits within affective cognition although such deficits have been linked to functional impairments in BD (Miskowiak et al., 2019). Nevertheless, ICAT is meant for screening purposes only and cannot replace a complete neuropsychological examination. Fourth, the ICAT ASR component displayed slight inconsistencies for the List Learning scores which may be unavoidable due to the early stage of ASR technology (König et al., 2018; Pakhomov et al., 2015). Finally, ICAT testing was conducted in a clinical setting rather than in patients' homes and may thus not be representative of scores obtained in home settings with different physical (and social) contexts. Future studies are thus required to established whether the ICAT has adequate validity and reliability in a remote administration setting. Finally, a limitation of the ICAT is that a lack of computer skills would impact on participants' ICAT scores. However, participants in the present study did not experience difficulties using the computer. This was perhaps because of their relatively young age (years, mean±SD: 30±8) and because of the specific instructions in the introductory ICAT video clip and in the auditory and written instructions before each ICAT test regarding which key(s) to press (Hafiz et al. 2019). Nevertheless, potential technical difficulty is an important aspect which we will assess with a brief online questionnaire in our planned next study in which participant will complete the ICAT in their home settings.

In conclusion, ICAT showed adequate concurrent validity, suggesting that this tool can be used for remote, self-administered assessments of objective cognitive functioning in fully or partially remitted patients with BD. The ICAT was even more sensitive to cognitive impairments than the SCIP and may be more sensitive in capturing perceived real-life functional and cognitive difficulties. The use of real-time ASR for assessment of verbal memory, as well as the close resemblance to a standard neuropsychological screening tool, represents important advantages of the ICAT system as compared to existing web-based cognitive screening tools. Based on the insights derived from this study, a slight optimization of the ICAT is now possible as well as the development of a parallel version for repeated testing. Future larger studies are now warranted to investigate the test-retest reliability of the ICAT and psychometric properties of a parallel ICAT test, identify optimal cut-off scores for cognitive impairments using receiver operator characteristic (ROC) analysis, and address the important question of how the tool fares

outside of a clinical setting (i.e., with remote testing) as well as in other patient groups that may benefit from a screening for cognitive impairments.

Authorship contribution statement

Kamilla Miskowiak: Conceptualization, design of study and ICAT, resources, supervision, writing – original draft, review and editing. Andreas Jespersen: recruiting, testing, statistical analysis, writing – original draft. Kia Obenhausen and Emilie Jacobsen: recruiting, writing – review and editing. Pegah Hafiz and Lorant Gulyas: design and programming of ICAT, piloting, writing - review and editing. Lars Kessing: conceptualization, supervision, writing – review and editing. Jakob Bardram: conceptualization, design of ICAT, resources, supervision, writing – review and editing. All authors have read and approved the final version of the manuscript.

References

Arts, B., Jabben, N., Krabbendam, L., van Os, J., 2008. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. Psychol. Med. 38, 771–785.

Bonnín, C., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A., Franco, C., Murru, A., Sanchez-Moreno, J., Vieta, E., 2010. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, 156–160.

Bora, E., Harrison, B.J., Yucel, M., Pantelis, C., 2013. Cognitive impairment in euthymic major depressive disorder: a meta-analysis. Psychol Med 43, 2017–2026.

Bora, E., Yucel, M., Pantelis, C., 2008. Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. J. Affect. Disord. 113, 1–20.

Bourne, C., Aydemir, O., Balanza-Martinez, V., Bora, E., Brissos, S., Cavanagh, J.T., Clark, L., Cubukcuoglu, Z., Dias, V.V., Dittmann, S., Ferrier, I.N., Fleck, D.E., Frangou, S., Gallagher, P., Jones, L., Kieseppa, T., Martinez-Aran, A., Melle, I., Moore, P.B., Mur, M., Pfennig, A., Raust, A., Senturk, V., Simonsen, C., Smith, D.J., Bio, D.S., Soeiro-de-Souza, M.G., Stoddart, S.D., Sundet, K., Szoke, A., Thompson, J. M., Torrent, C., Zalla, T., Craddock, N., Andreassen, O.A., Leboyer, M., Vieta, E., Bauer, M., Worhunsky, P.D., Tzagarakis, C., Rogers, R.D., Geddes, J.R., Goodwin, G. M., 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr. Scand. 128, 149–162.

Burdick, K.E., Russo, M., Frangou, S., Mahon, K., Braga, R.J., Shanahan, M., Malhotra, A. K., 2014. Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. Psychol. Med. 44, 3083–3096.

Cardoso, T., Bauer, I.E., Meyer, T.D., Kapczinski, F., Soares, J.C., 2015. Neuroprogression and cognitive functioning in bipolar disorder: a systematic review. Curr. Psychiatry Rep. 17, 75.

Crawford, J.R., Besson, J.A., Parker, D.M., Sutherland, K.M., Keen, P.L., 1987. Estimation of premorbid intellectual status in depression. Br. J. Clin. Psychol. 26, 313–314.

Cullen, B., Ward, J., Graham, N.A., Deary, I.J., Pell, J.P., Smith, D.J., Evans, J.J., 2016. Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder: a systematic review. J. Affect. Disord. 205, 165–181.

Davis, M.T., DellaGioia, N., Matuskey, D., Harel, B., Maruff, P., Pietrzak, R.H., Esterlis, I., 2017. Preliminary evidence concerning the pattern and magnitude of cognitive dysfunction in major depressive disorder using cogstate measures. J. Affect. Disord. 218, 82–85.

Demant Vinberg, M., Kessing, L.V., Miskowiak, K.W., 2015. Assessment of subjective and objective cognitive function in bipolar disorder: Correlations, predictors and the relation to psychosocial function. Psychiatry Res. 229, 565–571.

Demmo, C., Lagerberg, T.V., Aminoff, S.R., Hellvin, T., Kvitland, L.R., Simonsen, C., Haatveit, B., Andreassen, O.A., Melle, I., Ueland, T., 2017. Course of neurocognitive function in first treatment bipolar I disorder: one-year follow-up study. Psychiatry Res. 249, 286–292.

Domen, A.C., van de Weijer, S.C.F., Jaspers, M.W., Denys, D., Nieman, D.H., 2019. The validation of a new online cognitive assessment tool: The MyCognition Quotient. Int. J. Methods Psychiatr. Res. 28, e1775.

Field, A., 2013. Discovering Statistics using IBM SPSS Statistics. Sage Publications Ltd.

Fritz, C., Morris, P., Richler, J., 2011. Effect size estimates: current use, calculations, and interpretation. J. Exp. Psychol. Gen. 141, 2–18.

Greher, M.R., Wodushek, T.R., 2017. Performance validity testing in neuropsychology: scientific basis and clinical application-a brief review. J. Psychiatr. Pract. 23, 134–140.

Gualtieri, C.T., Johnson, L.G., 2006. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. Arch. Clin. Neuropsychol. 21, 623–643.

Hafiz, P., Miskowiak, K.W., Kessing, L.V., Jespersen, A.E., Obenhausen, K., Gulyas, L., Żukowska, K., Bardram, J.E., 2019. The Internet-Based Cognitive Assessment Tool: system design and feasibility study. JMIR Formative Res. 3, e13898.

Hamilton, M., 1960. A rating scale for depression. J Neurol Neurosurg.. Psychiatry 23, 56–62.

Harrison, J.E., Barry, H., Baune, B.T., Best, M.W., Bowie, C.R., Cha, D.S., Culpepper, L., Fossati, P., Greer, T.L., Harmer, C., Klag, E., Lam, R.W., Lee, Y., Mansur, R.B., Wittchen, H.U., McIntyre, R.S., 2018. Stability, reliability, and validity of the THINC-it screening tool for cognitive impairment in depression: a psychometric exploration in healthy volunteers. Int. J. Methods Psychiatr. Res. 27, e1736.

Jensen, J.H., Støttrup, M.M., Nayberg, E., Knorr, U., Ullum, H., Purdon, S.E., Kessing, L. V., Miskowiak, K.W., 2015. Optimising screening for cognitive dysfunction in bipolar disorder: validation and evaluation of objective and subjective tools. J. Affect. Disord. 187, 10–19.

Kessing, L.V., Munkholm, K., Faurholt-Jepsen, M., Miskowiak, K.W., Nielsen, L.B., Frikke-Schmidt, R., Ekstrom, C., Winther, O., Pedersen, B.K., Poulsen, H.E., McIntyre, R.S., Kapczinski, F., Gattaz, W.F., Bardram, J., Frost, M., Mayora, O., Knudsen, G.M., Phillips, M., Vinberg, M., 2017. The Bipolar Illness Onset study: research protocol for the BIO cohort study. BMJ Open 7, e015462.

König, A., Linz, N., Tröger, J., Wolters, M., Alexandersson, J., Robert, P., 2018. Fully automatic speech-based analysis of the semantic verbal fluency task. Dement. Geriatr. Cogn. Disord. 45, 198–209.

Kurtz, M.M., Gerraty, R.T., 2009. A meta-analytic investigation of neurocognitive deficits in bipolar illness: profile and effects of clinical state. Neuropsychology 23, 551–562.

Lewis, J.R., 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int. J. Hum.–Comput. Interact. 7, 57–78.

López-Villarreal, a., Sánchez-Morla, E.M., Jiménez-López, E., Martínez-Vizcaíno, V., Aparicio, A.I., Mateo-Sotos, J., Rodriguez-Jimenez, R., Vieta, E., Santos, J.L., 2020. Predictive factors of functional outcome in patients with bipolar I disorder: a five-year follow-up. J. Affect. Disord. 272, 249–258.

McIntyre, R.S., Best, M.W., Bowie, C.R., Carmona, N.E., Cha, D.S., Lee, Y., Subramaniapillai, M., Mansur, R.B., Barry, H., Baune, B.T., Culpepper, L., Fossati, P., Greer, T.L., Harmer, C., Klag, E., Lam, R.W., Wittchen, H.U., Harrison, J., 2017. The THINC-Integrated Tool (THINC-it) screening assessment for cognitive dysfunction: validation in patients with major depressive disorder. J Clin Psychiatry 78, 873–881.

McIntyre, R.S., Cha, D.S., Soczynska, J.K., Woldeyohannes, H.O., Gallaugher, L.A., Kudlow, P., Alsuwaidan, M., Baskaran, A., 2013. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. Axiety 30, 515–527.

Miskowiak, K.W., Burdick, K.E., Martinez-Aran, A., Bonnin, C.M., Bowie, C.R., Carvalho, A.F., Gallagher, P., Lafer, B., Lopez-Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R.J., Purdon, S., Torres, I.J., Yatham, L.N., Young, A.H., Kessing, L.V., Vieta, E., 2018. Assessing and addressing cognitive impairment in bipolar disorder: the International Society for Bipolar Disorders Targeting Cognition Task Force recommendations for clinicians. Bipolar Disord. 20, 184–194.

Miskowiak, K.W., Burdick, K.E., Martinez-Aran, A., Bonnin, C.M., Bowie, C.R., Carvalho, A.F., Gallagher, P., Lafer, B., Lopez-Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R.J., Torres, I.J., Yatham, L.N., Young, A.H., Kessing, L.V., Vieta, E., 2017. Methodological recommendations for cognition trials in bipolar disorder by the International Society for Bipolar Disorders Targeting Cognition Task Force. Bipolar Disord. 19, 614–626.

Miskowiak, K.W., Petersen, J., Ott, C.V., Knorr, U., Kessing, L., Gallagher, P., Robinson, L., 2016. Predictors of the discrepancy between objective and subjective cognition in bipolar disorder: a novel methodology. Acta Psychiatr. Scand. 134.

Miskowiak, K.W., Seeberg, I., Kjaerstad, H.L., Burdick, K.E., Martinez-Aran, A., del Mar Bonnin, C., Bowie, C.R., Carvalho, A.F., Gallagher, P., Hasler, G., Lafer, B., López-Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R.J., Purdon, S., Torres, I.J., Yatham, L.N., Young, A.H., Kessing, L.V., Van Rheenen, T.E., Vieta, E., 2019. Affective cognition in bipolar disorder: a systematic review by the ISBD targeting cognition task force. Bipolar Disord. 21, 686–719.

Mur, M., Portella, M.J., Martinez-Aran, A., Pifarre, J., Vieta, E., 2009. Influence of clinical and neuropsychological variables on the psychosocial and occupational outcome of remitted bipolar patients. Psychopathology 42, 148–156.

Nelson, H.E., Willison, J., 1991. National adult reading test (NART). Nfer-Nelson Windsor.

Nusslock, R., Alloy, L.B., 2017. Reward processing and mood-related symptoms: An RDoC and translational neuroscience perspective. J. Affect. Disord. 216, 3–16.

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.U., Jonsson, B., group, C.s., European Brain, C., 2012. The economic cost of brain disorders in Europe. Eur. J. Neurol. 19, 155–162.

Pakhomov, S.V.S., Marino, S.E., Banks, S., Bernick, C., 2015. Using automatic speech recognition to assess spoken responses to cognitive tests of semantic verbal fluency. Speech Commun. 75, 14–26.

Parsons, T.D., 2016. Clinical neuropsychology and technology, what's new and how we can use it. Springer International Publishing, Cham, pp. 47–63.

Petersen, J.Z., Porter, R.J., Miskowiak, K.W., 2019. Clinical characteristics associated with the discrepancy between subjective and objective cognitive impairment in depression. J. Affect. Disord. 246, 763–774.

Purdon, S., 2005. The Screen for Cognitive Impairment in Psychiatry: administration and Psychometric Properties. PNL Inc, Edmonton, Alberta, Canada.

Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol. Med. 44, 2029–2040.

Rojo, E., Pino, O., Guilera, G., Gomez-Benito, J., Purdon, S.E., Crespo-Facorro, B., Cuesta, M.J., Franco, M., Martinez-Aran, A., Segarra, N., Tabares-Seisdedos, R., Vieta, E., Bernardo, M., Mesa, F., Rejas, J., 2010. Neurocognitive diagnosis and cut-off scores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). Schizophr. Res. 116, 243–251.

Rosa, A.R., Mercade, C., Sanchez-Moreno, J., Sole, B., Mar Bonnin, C.D., Torrent, C., Grande, I., Sugranyes, G., Popovic, D., Salamero, M., Kapczinski, F., Vieta, E., Martinez-Aran, A., 2013. Validity and reliability of a rating scale on subjective cognitive deficits in bipolar disorder (COBRA). J. Affect. Disor.d 150, 29–36.

Rosa, A.R., Sanchez-Moreno, J., Martinez-Aran, A., Salamero, M., Torrent, C., Reinares, M., Comes, M., Colom, F., Van Riel, W., Ayuso-Mateos, J.L., Kapczinski, F., Vieta, E., 2007. Validity and reliability of the Functioning Assessment Short Test (FAST) in bipolar disorder. Clin. Pract. Epidemiol. Ment. Health 3, 5.

Tse, S., Chan, S., Ng, K.L., Yatham, L.N., 2014. Meta-analysis of predictors of favorable employment outcomes among individuals with bipolar disorder. Bipolar Disord. 16, 217–229.

Van der Elst, W., Van Boxtel, M.P.J., Van Breukelen, G.J.P., Jolles, J., 2008. A large-scale cross-sectional and longitudinal study into the ecological validity of neuropsychological test measures in neurologically intact people. Arch. Clin. Neuropsychol. 23, 787–800.

Wechsler, D., Psychological, C., 1997. WAIS-III : administration and scoring manual : Wechsler Adult Intelligence Scale. Psychological Corporation, [San Antonio, Tex].

Wing, J.K., Tf, B., Brugha, T.T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch. Gen. Psychiatry 47, 589–593.

Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133, 429–435.