



## The reliability of the Standard for Clinicians' Interview in Psychiatry (SCIP): A clinician-administered tool with categorical, dimensional and numeric output



Ahmed Aboraya <sup>a,\*</sup>, Ahmed El-Missiry <sup>d</sup>, Johnna Barlowe <sup>g</sup>, Collin John <sup>a</sup>, Alireza Ebrahimian <sup>f</sup>, Srinivas Muvvala <sup>c</sup>, Ja'me Brandish <sup>h</sup>, Hader Mansour <sup>b</sup>, Wanhong Zheng <sup>a</sup>, Paramjit Chumber <sup>a</sup>, James Berry <sup>a</sup>, Daniel Elswick <sup>a</sup>, Cheryl Hill <sup>a</sup>, Lauren Swager <sup>a</sup>, Warda Abo elez <sup>e</sup>, Hala Ashour <sup>e</sup>, Amal Haikal <sup>e</sup>, Ahmed Eissa <sup>e</sup>, Menan Rabie <sup>d</sup>, Marwa El-Missiry <sup>d</sup>, Mona El Sheikh <sup>d</sup>, Dina Hassan <sup>d</sup>, Sherif Ragab <sup>d</sup>, Mohamed Sabry <sup>d</sup>, Heba Hendawy <sup>d</sup>, Rola Abdel Rahman <sup>d</sup>, Doaa Radwan <sup>d</sup>, Mohamed Sherif <sup>d</sup>, Marwa Abou El Asaad <sup>d</sup>, Sherien Khalil <sup>d</sup>, Reem Hashim <sup>d</sup>, Katherine Border <sup>h</sup>, Roberto Menguito <sup>h</sup>, Cheryl France <sup>a</sup>, Wei Hu <sup>a</sup>, Olivia Shuttleworth <sup>a</sup>, Elizabeth Price <sup>a</sup>

<sup>a</sup> West Virginia University

<sup>b</sup> Western Psychiatric Institute and Clinic University of Pittsburgh School of Medicine

<sup>c</sup> Yale University School of Medicine

<sup>d</sup> Ain Shams University

<sup>e</sup> Mansoura University

<sup>f</sup> University Health Network, Toronto-Western Hospital

<sup>g</sup> Therapeutic Youth Day Treatment, Richmond VA, USA

<sup>h</sup> United Summit Center

### ARTICLE INFO

#### Article history:

Received 17 January 2014

Received in revised form 2 April 2014

Accepted 11 April 2014

Available online 17 May 2014

#### Keywords:

Psychiatric interview

dimensional scores

diagnostic criteria

standardized diagnostic interviews

### ABSTRACT

**Background:** Existing standardized diagnostic interviews are not used by psychiatrists in clinical settings. There is an urgent need for a clinician-administered tool for assessment of adult psychopathology that produces dimensional measures, in addition to categorical diagnoses.

**Methods:** The Standard for Clinicians' Interview in Psychiatry (SCIP) was designed to be used in clinical settings and generates dimensional measures. The reliability of the SCIP was tested at six sites: one hospital and two clinics in USA, two hospitals in Egypt and one clinic in Canada. Participants were adult patients who were admitted for inpatient psychiatric treatment or came for regular office visits in the outpatient clinic. Refusal rate was <1%. Missing data were <1.1%. Patients with dementia, mental retardation or serious medical conditions were excluded. A total of 1,004 subjects were interviewed between 2000 and 2012.

**Results:** Inter-rater reliability (Kappa) was measured for 150 SCIP items: 116 items (77.3%) had good reliability (Kappa > 0.7), 28 items (18.7%) had fair reliability (Kappa ranges from 0.5 to 0.7) and six items (4%) had poor reliability (Kappa < 0.5). Cronbach's alpha for internal consistency was measured for the SCIP dimensions: anxiety, posttraumatic stress, depression, mania, hallucinations, Schneider first-rank symptoms, delusions, disorganized thoughts, disorganized behavior, negative symptoms, alcohol addiction, drug addiction, attention and hyperactivity. All of the SCIP dimensions had substantial Cronbach's alpha values (>0.7) with the exception of disorganized thoughts (Cronbach's alpha = 0.375).

**Conclusions:** The SCIP is a reliable tool for assessing psychological symptoms, signs and dimensions of the main psychiatric diagnoses.

© 2014 Elsevier B.V. All rights reserved.

\* Corresponding author at: William R. Sharpe Jr. Hospital, 936 William R. Sharpe Jr. Road, Weston, WV 26452. Tel.: +1 304 269 1210; fax: +1 304 269 2109.

E-mail address: [Aboraya@scip-psychiatry.com](mailto:Aboraya@scip-psychiatry.com) (A. Aboraya).

<sup>1</sup> Dr. Aboraya is the Chief of Psychiatry at William R. Sharpe Jr. Hospital, Clinical professor of psychiatry at West Virginia School of Osteopathic Medicine in Lewisburg, WV and adjunct faculty in the School of Public Health at West Virginia University in Morgantown, WV. Dr. Aboraya received his medical degree (MD) from Cairo University, a master of public health (MPH) and a doctorate of public health (Dr.PH) from Johns Hopkins University.

## 1. Introduction

The science of classification of mental disorders has expanded in the twentieth century due to the efforts of the World Health Organization (WHO) and the American Psychiatric Association (APA). The World Health Organization (WHO) published the International Classification of Diseases (ICD), with the latest version (ICD-10) in 1993 (WHO, 1993). In the United States, the American Psychiatric Association published the Diagnostic and Statistical Manual of Mental Disorders (DSM), with the latest version (DSM-5) in 2013 (American Psychiatric Association, 2013). Consequently, Standardized Diagnostic Interviews (SDIs) were developed with the main goal of diagnosing disorders based upon the existing classification systems. For clinicians, semi-structured interviews were designed to allow for considerable variation in the interviewing style, depth of probing and clinical judgment as to whether a patient's description of a particular behavior meets the relevant diagnostic criterion (Williams et al., 1992). The two widely used semi-structured interviews in the assessment literature are the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (Wing et al., 1990; Spitzer et al., 1992; Williams et al., 1992).

The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) was designed with a top-down approach: questions are grouped by diagnosis and criteria; within each diagnosis, if a required criterion is not met, the interviewer skips the remaining questions assessing the other criteria for that diagnosis (Spitzer et al., 1992). The SCID-I has been used by researchers all over the world (Picardi et al., 2013; Pinna et al., 2013). However, the SCID-I is not practical to use in clinical settings because it is lengthy, cumbersome and interferes with establishing rapport with patients. The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) is another semi-structured interview developed under the auspices of the World Health Organization (WHO). It utilizes the bottom-up approach: after comprehensive baseline symptoms assessment, algorithms are used to make psychiatric diagnoses (Wing et al., 1990). Similarly, the SCAN is not practical to use in clinical settings because it is lengthy and requires extensive training. The Mini-International Neuropsychiatric Interview (M.I.N.I.) was developed by Sheehan as a short structured diagnostic interview (Sheehan et al., 1998). Overall, the M.I.N.I. is simpler and shorter than the SCID-I and the SCAN and has some acceptance in clinical settings (Aboraya, 2009). On the downside, the M.I.N.I. is simply a checklist of symptoms and the clinical approach of probing and exploring the symptoms is largely lost (Nienhuis et al., 2010).

Considering the most popular existing standardized diagnostic interviews (SDIs) and the efforts to modify them for clinicians' use, the literature indicates that psychiatrists do not use the existing structured interviews or rating scales in real clinical settings (Gilbody et al., 2002; Duffy et al., 2008; Morrison, 2008; Aboraya, 2009; Nasrallah, 2009; Rettew et al., 2009; Bastiaens, 2011; Busner et al., 2011). Considering these serious gaps in the literature, namely the lack of an instrument designed for psychiatrists to use in clinical settings and the lack of a clinician-administered tool with dimensional measures, the first

author devised the Standard for Clinicians' Interview in Psychiatry (SCIP).

## 2. Design, description and output from the SCIP

### 2.1. The SCIP materials and output

The SCIP method of psychiatric assessment consists of **the SCIP interview and the SCIP instruction manual**. The SCIP generates three main outputs: approximately 300 questions and observational items, dimensional scores for symptoms clusters, and the main psychiatric diagnoses in modules. In general, symptoms and categorical diagnoses correspond to the DSM criteria. Symptoms are measured with Likert-type scales with the exception of the alcohol and drug modules, which are binary. Symptoms are assessed for the past month with the exception of the alcohol and drug modules (past year or life-time). Symptoms are coded zero (absent or non-significant symptom), one (symptom present less than half the time) or two (symptom present more than half the time). For the alcohol and drug modules, symptoms are coded zero for absent or non-significant symptom and one for present. The SCIP dimensions depend upon the number and severity of positive items. Details of the phases of the development of the SCIP, design of the SCIP questions and dimensions, and the SCIP instruction manual are available from the author. Psychiatrists and experienced clinicians can read the SCIP materials and use the SCIP without training.

### 2.2. "Bottom First Then Top (BFTT)": the SCIP approach to psychiatric diagnosis

Two approaches to psychiatric diagnosis have been described in literature: the "top-down" approach and the "bottom-up" approach. The advantages and disadvantages of each have been described elsewhere (Wing et al., 1990; Spitzer et al., 1992; Ustun and Tien, 1995; Helzer et al., 2008; McHugh and Slavney, 2012). Seasoned, competent psychiatrists actually use what the first author describes as the **"Bottom First Then Top (BFTT)" approach** to diagnostic assessment. An ideal diagnostic psychiatric interview starts with a bottom-up approach: the psychiatrist establishes rapport with the patient and inquires about chief complaint(s) and history of the present illness. The patient is allowed to take the lead first and express feelings, thoughts, current stressors and other problems. The psychiatrist continues the bottom-up approach by obtaining a detailed life history, screening for symptoms, examining mental status, exploring potential causes of symptoms and utilizing records and informants as needed. As the psychiatrist narrows down the potential diagnoses, the top-down approach takes over. The psychiatrist checks the symptoms and decides whether or not the patient meets the diagnostic criteria for a disorder. **The motto of the BFTT approach is to understand the patient first, then make a diagnosis.**

**Table 1**  
Patient demographic characteristics by site.

	Site 1 Sharpe Hospital- Inpatient	Site 2 Chestnut Ridge Center-Outpatient	Site 3 Ain Shams University Hospital-Inpatient	Site 4 Mansoura University Hospital-Inpatient	Site 5 Fairmont Office- Outpatient	Site 6 Rothbart Center- Outpatient
Sample size	780	30	52	69	42	31
Mean age	33	42	30	35	41	44
Gender, males, %	57	53	69	53	48	61
Education, <12 grade, %	62	29	41	55	61	79
Ethnicity, Whites, %	96	100			100	55
Contribution to Kappa	Yes	Yes	Yes	Yes	No (*)	No (*)
Contribution to alpha	Yes	Yes	Yes	Yes	Yes	Yes

\* Each patient was interviewed one time by one rater. No inter-rater comparisons at site 5 or site 6.

**Table 2**

Raters' qualifications, number of raters, total and average number of patients per rater across six sites (1,004 patients).

Rater qualifications	Total number of raters	Total number of patients	Average number of patients per rater
Author of SCIP	1	579	579
Psychiatrists with >4 years' experience	12	216	18
Senior psychiatry residents (3rd, 4th year)	4	70	17
Doctoral –level clinical psychologist with >4 years' experience	1	31	31
Masters-level clinical psychologist with one year experience	2	42	21
Masters-level clinical social worker with one year experience in mental health	1	18	18
Researchers with one or more year experience in mental health	5	48	10
Total number of raters = 26		Total number of patients = 1004	

### 3. Methods

#### 3.1. Study design and participants

##### 3.1.1. Sites

The SCIP was tested in an international multi-site study (3 hospitals and 3 clinics) in three countries (USA, Canada and Egypt).

Site 1: William R. Sharpe, Jr. Hospital, Weston, West Virginia (a 150-bed, JCAHO accredited, state psychiatric facility).

Site 2: Chestnut Ridge Center Clinic, Morgantown, West Virginia.

Site 3: Ain Shams University Hospital, Cairo, Egypt.

Site 4: Mansoura University Hospital, Mansoura, Egypt.

Site 5: Fairmont Physician Office, Fairmont, West Virginia.

Site 6: Rothbart Center, Toronto, Canada.

The SCIP was translated into the Arabic language and was tested at sites 3 and 4 in Egypt. Table 1 shows the demographic characteristics of participants from each site.

##### 3.1.2. Subjects

###### 1. Patient recruitment:

The SCIP project was conducted in natural clinical settings and patients were recruited as they came for regular outpatient visits or admitted to the hospital. The investigators were allowed to review the patients' records as they usually do. Less than 1% of the patients approached refused to participate. No advertising for the project was done and no financial rewards were given to the patients. A total of 1,004 subjects were recruited for the SCIP study between 2000 and 2012.

2. Inclusion criteria: patients who were 18 years and older at the time of the interview and competent to sign the consent were enrolled in the study.

3. Exclusion criteria: patients with dementia, mental retardation or serious medical conditions were excluded from the study.

##### 3.1.3. Raters

Table 2 shows the profile of raters who participated at all six sites (26 raters interviewed 1,004 patients). Table 3 shows the profile of the raters who participated in the inter-rater reliability testing (22 raters interviewed 303 patients and generated 606 interviews). Inter-rater reliability was tested at site 1, site 2, site 3 and site 4. No inter-

rater reliability testing was performed at site 5 or site 6 because each of these two sites had one rater who interviewed the patient one time.

##### 3.1.4. Ethics

The SCIP study has been carried out in accordance with The Code of Ethics of the World Medical Association. The SCIP study was approved and renewed annually by the West Virginia University Institution Review Board (IRB). Patients who had capacity signed the consent. The SCIP interview was also approved as part of the patients' health records at Sharpe Hospital and is part of the Electronic Health Records (EHR) of West Virginia University Hospital. The first author used the SCIP routinely with most of his patients (inpatients and outpatients). The first author also obtained another IRB approval for exempt research, which allowed retrospective review of the existing data that were part of the health records. In Canada and Egypt, the local investigators obtained the approval for the SCIP project as required by each institution's guidelines for research on human subjects.

#### 3.2. Reliability

##### 3.2.1. Inter-rater reliability

To measure inter-rater reliability, the patient was interviewed in the presence of two or more investigators at the same time in the same room. One interviewer conducted the interview and asked the questions (lead interviewer) and the other interviewers observed and rated the SCIP items as the interview proceeded (observers). When the interview was finished, observers were allowed to ask clarifying questions if needed. The lead interviewer and the observers rated responses independently and were never allowed to discuss their ratings. To measure reliability, we used the index of agreement Kappa (K) which measures the proportion of agreement corrected for chance agreement (Cohen, 1960; Fleiss, 1981; Kraemer et al., 2002). In general, Kappa values greater than 0.7 indicate good agreement, Kappa values ranging from 0.5 to 0.7 indicate fair agreement, and Kappa values less than 0.5 indicate poor agreement (Williams et al., 1992). Kappa is considered stable if the number of positive cases for a given symptom is 10 or more, regardless of agreement or disagreement of the raters on the individual items being assessed (Williams et al., 1992).

**Table 3**

For 606 interviews from 303 patients providing inter-rater reliability data: Raters' qualifications, number of raters, total and average number of patients per rater and the total number of interviews.

Rater qualifications	Total number of raters	Total number of patients	Average number of patients per rater	Total number of interviews
Author of SCIP	1	64	64	128
Psychiatrists with >4 years' experience	12	153	13	306
Senior psychiatry residents (3rd, 4th year)	3	38	13	76
Masters-level psychologist with one year experience	2	7	3	14
Masters-level clinical social worker with one year experience in mental health	1	4	4	8
Researchers with one or more year experience in mental health	3	37	12	74
Total # of raters = 22		Total number of patients = 303		Total number of interviews = 606

**Table 4**

Inter-rater reliability agreement (Kappa) and standard error (SE) for the SCIP items (symptoms and signs) in patients at site 1 (Sharpe Hospital), site 2 (Chestnut Ridge Center), site 3 (Ain Shams University Hospital) and site 4 (Mansoura University Hospital).  
Stable Kappa Only (\*).

	SCIP items (symptoms and signs)	Total number of positive cases for a given item	Kappa (**)	SE
1	Generalized anxiety	61	0.76	0.05
2	Panic attack	54	0.81	0.05
3	Agoraphobia	26	0.52	0.05
4	Social phobia	22	0.51	0.05
5	Screening for obsessions	38	0.70	0.04
6	Screening for compulsions	31	0.58	0.05
7	Experience traumatic events	69	0.75	0.05
8	Re-experience traumatic events	34	0.89	0.05
9	Depressed mood	158	0.86	0.04
10	Anhedonia	125	0.87	0.04
11	Suicidal ideation, intention, pan	79	0.61	0.04
12	Elated mood	76	0.72	0.05
13	Irritable mood	65	0.75	0.05
14	Mixed mood (same day mood changes)	44	0.50	0.05
15	Paranoid delusions	97	0.83	0.04
16	Other delusions	39	0.77	0.04
17	Auditory hallucination	92	0.76	0.04
18	Visual hallucination	51	0.68	0.05
19	Violence	74	0.64	0.04
20	Disorganized behavior	32	0.54	0.04
21	Disorganized thoughts	39	0.65	0.04
22	Alcohol problems	53	0.89	0.06
23	Opioid problem	17	0.78	0.06
24	Cannabis problem	18	0.87	0.06
25	Cocaine problem	13	0.86	0.06
26	Worry about weight and image	12	0.73	0.05
27	Poor attention	11	0.73	0.05
28	Hyperactivity	14	0.58	0.05
29	Panic attacks	30	0.92	0.06
30	Worry about having another panic attack	25	0.81	0.04
31	Action to prevent panic attacks	26	0.87	0.04
32	Generalized anxiety	25	0.84	0.04
33	Restlessness with anxiety	26	0.74	0.04
34	Tension with anxiety	22	0.77	0.04
35	Exhaustion with anxiety	22	0.79	0.05
36	Poor concentration with anxiety	27	0.76	0.05
37	Irritability with anxiety	28	0.83	0.04
38	Insomnia with anxiety	25	0.82	0.05
39	Obsession	26	0.85	0.04
40	Compulsion	18	0.77	0.04
41	Experience traumatic events	10	0.83	0.05
42	Distressing recollection of events	30	0.88	0.05
43	Bad dreams or nightmares	26	0.94	0.05
44	Flashback	23	0.87	0.05
45	Psychological distress due to events	26	0.91	0.05
46	Physical reactions due to events	24	0.93	0.05
47	Avoidance	27	0.94	0.05
48	Amnesia	15	0.70	0.06
49	Diminished interest	17	0.83	0.05
50	Detachment	22	0.87	0.05
51	Diminished affect	24	0.88	0.05
52	Insomnia	16	0.78	0.05
53	Anger	19	0.80	0.05
54	Poor concentration	14	0.78	0.05
55	Hypervigilance	17	0.87	0.05
56	Startle response	20	0.86	0.05
57	Daze	16	0.82	0.05
58	Depressed mood	128	0.91	0.04
59	Anhedonia	121	0.87	0.04
60	Crying when depressed	11	0.76	0.04
61	Hopelessness	11	0.82	0.04
62	Poor concentration	116	0.80	0.04
63	Psychomotor retardation	97	0.72	0.04
64	Appetite changes when depressed	93	0.79	0.04
65	Weight loss	62	0.71	0.04
66	Weight gain	15	0.76	0.05
67	Initial insomnia	103	0.79	0.04
68	Middle insomnia	79	0.65	0.04
69	Late insomnia	46	0.62	0.04
70	Hypersomnia	26	0.68	0.05
71	Decreased libido	74	0.80	0.04
72	Worthlessness	97	0.78	0.04

(continued on next page)

Table 4 (continued)

	SCIP items (symptoms and signs)	Total number of positive cases for a given item	Kappa (**)	SE
73	Guilt	86	0.80	0.04
74	Suicide	68	0.64	0.04
75	Elated mood	71	0.75	0.04
76	Irritable mood	70	0.76	0.04
77	Mixed mood (same day mood changes)	41	0.58	0.05
78	Racing thoughts	71	0.85	0.04
79	Pressured speech	53	0.72	0.04
80	Clanging	12	0.49	0.04
81	Distraction	63	0.79	0.04
82	Overactivity	68	0.83	0.04
83	Grandiosity	40	0.81	0.04
84	Over spending	49	0.74	0.04
85	Decreased sleep	56	0.78	0.04
86	Hypersexuality	24	0.69	0.04
87	Hallucination quality	54	0.90	0.04
88	Hallucination frequency	54	0.93	0.05
89	Internal hallucination	50	0.84	0.04
90	Voices commenting	40	0.77	0.04
91	Second and third hallucination	45	0.78	0.04
92	Visual hallucination	27	0.81	0.04
93	Tactile hallucination	10	0.95	0.05
94	Observed hallucination	12	0.55	0.04
95	Reading thoughts	17	0.83	0.04
96	Thought insertion	16	0.76	0.04
97	Thought broadcast	16	0.71	0.04
98	Paranoid delusion	50	0.86	0.04
99	Persecutory delusion	56	0.85	0.04
100	Conspiracy delusion	49	0.84	0.04
101	Delusion of reference	31	0.81	0.05
102	Religious delusion	17	0.80	0.04
103	Grandiose delusion	16	0.77	0.05
104	Other delusion	12	0.40	0.05
105	Bizarreness of delusion	14	0.43	0.05
106	Derailment	37	0.65	0.06
107	Flight of ideas	15	0.62	0.06
108	Tangentiality	28	0.57	0.06
109	Incoherent speech	18	0.41	0.06
110	Illogical speech	13	0.25	0.05
111	Agitation	33	0.48	0.04
112	Violence	25	0.64	0.04
113	Odd behavior	19	0.67	0.06
114	Inappropriate affect	14	0.77	0.06
115	Blunted affect	42	0.68	0.05
116	Avolition	35	0.74	0.04
117	Alogia	29	0.62	0.05
118	Poor self care	27	0.79	0.06
119	Alcohol tolerance	39	0.99	0.06
120	Alcohol withdrawal	33	0.93	0.06
121	Drinking alcohol to avoid withdrawal	29	0.96	0.06
122	Unable to control alcohol	51	0.96	0.06
123	Unable to reduce or stop alcohol	47	0.85	0.06
124	Time spent to drink alcohol	37	0.94	0.06
125	Less time working	36	0.92	0.06
126	Work problem	29	0.83	0.06
127	Fighting when intoxicated	31	0.90	0.06
128	Family problem	51	0.82	0.06
129	Legal problem	29	0.92	0.06
130	Medical problem	11	0.70	0.06
131	Depression problem	24	0.90	0.06
132	Continue alcohol with problem	57	0.87	0.06
133	Alcohol in hazardous situations	42	0.77	0.06
134	Alcohol binge	37	0.88	0.06
135	Alcohol blackout	53	0.98	0.06
136	Drug tolerance	49	0.95	0.06
137	Drug withdrawal	46	0.97	0.06
138	Using drug to avoid withdrawal	40	0.94	0.06
139	Unable to control drug use	55	0.97	0.06
140	Unable to reduce or stop drug use	54	0.97	0.06
141	Time spent to use drug	56	0.88	0.06
142	Less time working	50	0.95	0.06
143	Work problem	34	0.83	0.06
144	Fighting when using drug	22	0.80	0.06
145	Family problem	58	0.80	0.06
146	Legal problem	22	0.80	0.06
147	Depression problem	19	0.76	0.06
148	Delusion problem	16	0.59	0.06
149	Drug use with problem	64	0.91	0.06

Table 4 (continued)

SCIP items (symptoms and signs)	Total number of positive cases for a given item	Kappa (**)	SE
150 Drug use in hazardous situations	57	0.90	0.06

\* Kappa is stable if the number of positive cases for a given item is 10 or more.

\*\* Kappa values were calculated based upon inter-rater interviews of 303 patients at site 1, site 2, site 3 and site 4.

### 3.2.2. Internal consistency reliability (Cronbach's alpha)

Cronbach's alpha, which measures the reliability of a dimensional measure composed of several items, was calculated using data from all administrations of the SCIP (Cronbach, 1951; Carmines and Zeller, 1979; Schene et al., 2000). Cronbach's alpha takes into consideration the mean inter-item correlation and its value increases as the average inter-item correlation and the number of items increase (Carmines and Zeller, 1979). The reliability of a dimension is considered substantial if Cronbach's alpha is  $>0.7$ , moderate if alpha ranges from 0.5 to 0.7, and poor if alpha is  $<0.5$  (Schene et al., 2000).

### 3.3. Statistical analyses

STATA was used to calculate the values of Kappa, Cronbach's alpha and other statistical measures (STATA, 2009). Missing data were coded "9" and were not used in the analyses. If the interviewers were unsure about particular ratings, those data were coded "8" and also were not used in analysis. The percentage of "missing" and "unsure" data varied from 0.07% to 1.1%.

## 4. Results

### 4.1. Inter-rater reliability

Table 4 shows inter-rater reliability (Kappa) of the SCIP symptoms and signs based upon 303 patients interviewed in USA and Egypt. Kappa was calculated for 150 psychological symptoms and signs, including 28 SCIP screening questions. Among the 150 SCIP items, six items (4%) had poor reliability: clanging ( $K = 0.49$ ), other delusions ( $K = 0.4$ ), bizarreness of delusions ( $K = 0.43$ ), incoherent speech ( $K = 0.41$ ), illogical speech ( $K = 0.25$ ) and agitation ( $K = 0.48$ ). Of the 150 SCIP items, 28 items (18.7%) had fair reliability ( $K$  ranges from 0.5 to 0.7) and 116 items (77.3%) had good reliability ( $K > 0.7$ ).

### 4.2. Internal consistency (Cronbach's alpha) of the SCIP dimensions

The SCIP interviews of 1,004 patients from all sites were pooled to measure the internal consistency (Cronbach's alpha) of as many dimensions as possible. The SCIP dimensions were created based on three principles. First: the initial composition of each dimension was derived from all possible criteria of the DSM diagnoses. For example, the DSM-IV required one or more symptoms for the diagnosis of alcohol abuse and three or more symptoms for the diagnosis of alcohol dependence. All of the eleven criteria for alcohol abuse and dependence, in the form of questions, were included in initial model. Second, some important questions were added, even though they were not in the DSM criteria. For example, two important questions were added to depression dimension: crying when depressed (important for depression severity) and feeling hopeless (important for suicide prediction). Both questions were retained in the final depression dimension because they had high item-rest correlation (Table 5). Third: some dimensions were predefined in the literature such as Schneider's first rank symptoms. The delusions and hallucinations dimensions are the types of delusions and hallucinations.

The items included in each dimension were measured with Likert-type scales with the exception of items from the alcohol and drug modules, which were binary. The rule of thumb in internal consistency analysis is to drop items with item-total correlation  $<0.4$  (Gliem and

Gliem, 2003). Nunnally and Bernstein (1994) recommended using item-rest correlations (the correlation between an item and the scale that is formed by all other items (Nunnally and Bernstein, 1994)). We dropped items with item-rest correlations below 0.4. Table 5 shows the internal consistency of the following SCIP dimensions: anxiety, post-traumatic stress, depression, mania, hallucinations, Schneider first-rank symptoms, delusions, disorganized thoughts, disorganized behavior, negative symptoms, alcohol addiction, drug addiction, attention and hyperactivity. All of the SCIP dimensions had substantial Cronbach's alphas ( $>0.7$ ), with the exception of disorganized thoughts (Cronbach's alpha = 0.375). Table 5 also shows the item-rest correlation for each item included in the dimension, mean inter-item correlation of all items in each dimension and the values of Cronbach's alpha with 95% confidence interval for 14 SCIP dimensions. These initial results show that the SCIP dimensions, based upon the SCIP items, are reliable.

## 5. Discussion

### 5.1. Main findings

The Standard for Clinicians' Interview in Psychiatry (SCIP) was developed and designed to reliably measure psychological symptoms, signs and dimensions and to transform traditional psychiatric assessments into research data. Over the past few decades, there have been increased efforts to measure the reliability of psychiatric symptoms as the use of rating scales and structured interviews has expanded. To our knowledge, Rodgers and Mann (1986) studied the largest number of psychiatric symptoms for reliability in 1986. The index of association for its presence,  $IA(p)$ , was calculated for 48 anxiety and depressive questions derived from the Present State Examination (Rodgers and Mann, 1986). The SCIP study calculated Kappa for 178 psychological symptoms and signs, including 29 screening questions covering the main areas of adult psychopathology.

Creating reliable psychological dimensions requires that symptoms and signs be measured in a valid and reliable way. The absence of valid and reliable symptoms was the main limiting factor in creating dimensional measures in the past (Andreasen et al., 1992). The SCIP study removed this major obstacle. Based upon reliable SCIP items, the SCIP dimensions were created and have shown to be reliable. In the SCIP study, internal consistency reliability was measured for 14 SCIP dimensions and all of the SCIP dimensions had substantial Cronbach's alphas ( $>0.7$ ), with the exception of disorganized thoughts (Cronbach's alpha = 0.375). This unique achievement of the SCIP study, creating reliable psychological symptoms, signs and dimensions, allows for the formation of a psychological database. Twenty years ago, Nancy C. Andreasen, a renowned researcher, envisioned the importance and the need for a comprehensive archival database of psychopathology that would match sophisticated neurobiological techniques, enabling us to study the interrelationships among psychopathology, treatment response and neurobiological measures (Andreasen et al., 1992). The vision of a comprehensive archival database of psychopathology is a reality today through the SCIP database.

As psychiatrists and experienced clinicians use the SCIP methodology, psychiatric information obtained by interviewing patients can be transformed into useful research data. The SCIP study database contains 1,419 SCIP interviews and 761 variables and is available for statisticians and researchers to analyze and publish on the phenomenology of mental disorders among adults.

**Table 5**  
Internal consistency of the SCIP dimensions. Cronbach's alpha and one-sided 95% confidence interval (CI), based upon SCIP interviews of 1,004 patients at site 1, site 2, site 3, site 4, site 5 and site 6.

SCIP dimension	SCIP items included in dimension	Item-rest correlation	Mean inter-item correlation	Cronbach's alpha for the dimension (one-sided 95% CI)
1 Generalized anxiety dimension	1. Generalized anxiety	0.477	0.706	0.944 ( $\geq 0.939$ )
	2. Restlessness with anxiety	0.905		
	3. Tension with anxiety	0.867		
	4. Exhaustion with anxiety	0.865		
	5. Poor concentration with anxiety	0.896		
	6. Irritability with anxiety	0.845		
	7. Insomnia with anxiety	0.856		
2 Posttraumatic stress dimension	1. Experience traumatic events	0.652	0.576	0.961 ( $\geq 0.953$ )
	2. Distressing recollection of events	0.841		
	3. Bad dreams or nightmares	0.807		
	4. Flashback	0.807		
	5. Psychological distress due to events	0.830		
	6. Physical reactions due to events	0.845		
	7. Avoidance	0.805		
	8. Amnesia	0.576		
	9. Diminished interest	0.726		
	10. Detachment	0.822		
	11. Diminished affect	0.808		
	12. Insomnia	0.804		
	13. Anger	0.760		
	14. Poor concentration	0.693		
	15. Hypervigilance	0.698		
	16. Startle response	0.741		
	17. Daze	0.649		
	18. Derealization	0.513		
3 A Depression dimension (with suicide question)	1. Depressed mood	0.843	0.477	0.927 ( $\geq 0.921$ )
	2. Anhedonia	0.844		
	3. Crying when depressed	0.737		
	4. Hopelessness	0.804		
	5. Poor concentration	0.762		
	6. Psychomotor retardation	0.684		
	7. Appetite changes when depressed	0.541		
	8. Weight loss	0.499		
	9. Initial insomnia	0.663		
	10. Middle insomnia	0.576		
	11. Decreased libido	0.609		
	12. Worthlessness	0.781		
	13. Guilt	0.694		
	14. Suicide	0.286		
3 B Depression dimension (without suicide question)	1. Depressed mood	0.844	0.521	0.934 ( $\geq 0.929$ )
	2. Anhedonia	0.846		
	3. Crying when depressed	0.738		
	4. Hopelessness	0.808		
	5. Poor concentration	0.767		
	6. Psychomotor retardation	0.689		
	7. Appetite changes when depressed	0.543		
	8. Weight loss	0.501		
	9. Initial insomnia	0.660		
	10. Middle insomnia	0.571		
	11. Decreased libido	0.620		
	12. Worthlessness	0.781		
	13. Guilt	0.694		
4 Mania dimension	1. Elated mood	0.635	0.528	0.918 ( $\geq 0.911$ )
	2. Irritable mood	0.570		
	3. Racing thoughts	0.799		
	4. Pressured speech	0.761		
	5. Distraction	0.721		
	6. Overactivity	0.838		
	7. Grandiosity	0.597		
	8. Over spending	0.778		
	9. Decreased Sleep	0.745		
	10. Hypersexuality	0.501		
5 Hallucinations dimension	1. Auditory Hallucination	0.547	0.528	0.918 ( $\geq 0.911$ )
	2. Hallucination quality	0.804		
	3. Internal Hallucination	0.806		
	4. Voices commenting	0.800		

Table 5 (continued)

SCIP dimension	SCIP items included in dimension	Item-rest correlation	Mean inter-item correlation	Cronbach's alpha for the dimension (one-sided 95% CI)
6 Schneiderian dimension	5. Second and third hallucination	0.759	0.574	0.890 ( $\geq 0.880$ )
	6. Visual hallucination	0.538		
	1. Voices commenting	0.497		
	2. Reading thoughts	0.523		
	3. Loud thoughts	0.557		
	4. Thought insertion	0.531		
7 A Delusions dimension (with bizarre delusions)	5. Thought broadcast	0.693	0.388	0.792 ( $\geq 0.773$ )
	6. Replacement of will	0.429		
	1. Paranoid delusion	0.729		
	2. Persecutory delusion	0.744		
	3. Conspiracy delusion	0.731		
	4. Delusion of reference	0.458		
7 B Delusions dimension (without bizarre delusions)	5. Bizarreness of delusion	0.309	0.447	0.802 ( $\geq 0.784$ )
	1. Paranoid delusion	0.745		
	2. Persecutory delusion	0.782		
	3. Conspiracy delusion	0.755		
8 Disorganized thoughts dimension	4. Delusion of reference	0.460	0.575	0.844 ( $\geq 0.830$ )
	1. Derailment	0.297		
	2. Flight of ideas	0.163		
	3. Tangentiality	0.378		
	4. Incoherent speech	0.042		
9 Disorganized behavior dimension	5. Illogical speech	0.093	0.107	0.375 ( $\geq 0.314$ )
	1. Agitation	0.463		
	2. Violence	0.603		
	3. Violence a day	0.430		
	4. Violence a period	0.436		
10 Negative dimension	5. Odd behavior	0.476	0.342	0.722 ( $\geq 0.697$ )
	1. Blunted affect	0.727		
	2. Avolition	0.666		
	3. Alogia	0.709		
	4. Psychomotor slowing	0.489		
11 Alcohol dimension	5. Poor self care	0.620	0.509	0.838 ( $\geq 0.824$ )
	1. Alcohol tolerance	0.741		
	2. Alcohol withdrawal	0.770		
	3. Drinking alcohol to avoid withdrawal	0.747		
	4. Unable to control alcohol	0.771		
	5. Unable to reduce or stop alcohol	0.827		
	6. Time spent to drink alcohol	0.785		
	7. Less time working	0.814		
	8. Work problem	0.589		
	9. Fighting when intoxicated	0.593		
	10. Family problem	0.775		
	11. Legal problem	0.636		
	12. Medical problem	0.419		
	13. Alcohol-induced depression	0.529		
	14. Continue alcohol with problem	0.874		
	15. Alcohol in hazardous situations	0.656		
	16. Alcohol binge	0.747		
17. Alcohol blackout	0.810			
12 Drug dimension	1. Drug tolerance	0.643	0.532	0.951 ( $\geq 0.947$ )
	2. Drug withdrawal	0.647		
	3. Using drug to avoid withdrawal	0.615		
	4. Unable to control drug use	0.690		
	5. Unable to reduce or stop drug use	0.596		
	6. Time spent to use drug	0.667		
	7. Less time working	0.687		
	8. Work problem	0.620		
	9. Family problem	0.708		
	10. Drug use with problem	0.796		
	11. Drug use in hazardous situations	0.563		
13 Attention dimension	1. Drug tolerance	0.926	0.480	0.910 ( $\geq 0.902$ )
	2. Poor concentration	0.902		
	3. Poor organization	0.537		
	4. Changing activities	0.855		

(continued on next page)



Table 5 (continued)

SCIP dimension	SCIP items included in dimension	Item-rest correlation	Mean inter-item correlation	Cronbach's alpha for the dimension (one-sided 95% CI)
14 Hyperactivity dimension	5. Distraction	0.715	0.631	0.932 ( $\geq 0.880$ )
	6. Misplacing things	0.537		
	7. Forgetting	0.843		
	8. Losing track of things	0.819		
	1. Impulsivity	0.559		
	2. Talking too much	0.527		
	3. Disturb others	0.860		
4. Difficulty being seated	0.784	0.528	0.887 ( $\geq 0.798$ )	
5. Leaving seats	0.523			
6. Fidgety	0.757			
7. Over activity	0.757			

### 5.2. Strengths

The most widely used Standardized Diagnostic Interviews (SDIs), the Composite International Diagnostic Interview (CIDI), the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Mini-International Neuropsychiatric Interview (M.I.N.I.), do not provide dimensional measures (Schutzwohl et al., 2007). The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) has the potential to provide dimensional measures. However, the complexity of the SCAN limits this advantage to research applications only (Schutzwohl et al., 2007). Finally, these interviews are mainly dependent upon the current diagnostic system. Because the DSM-5 was published in May 2013, all of these interviews will require major revisions and new testing for the reliability and validity of the new versions. The SCIP was designed to produce clinically useful dimensions built upon clinically useful questions. The SCIP questions were shown to be reliable, as were the SCIP dimensions. Another strength of the SCIP study is the large sample size (1,004 subjects), making the SCIP project the largest reliability study of a diagnostic assessment tool. In the SCIP study, no discussions of assessments or diagnoses were allowed among the interviewers. No funding was obtained from any pharmaceutical companies or agencies. The SCIP is comprehensive, covering the main domains in adult psychiatry, and can be easily tailored and used in specific studies. For example, if researchers want to study the efficacy of a new antidepressant, the 20 SCIP questions on depression and the SCIP depression dimension can be used. Similarly, researchers studying bipolar disorder, schizophrenia, or attention deficit/hyperactivity disorder can use the SCIP modules and the dimensions of these disorders.

### 5.3. Limitations

The SCIP study has several weaknesses. Although the sample size of the SCIP study is large (1,004 subjects), 78% of the sample were inpatients at one hospital (Sharpe Hospital) where the author of the SCIP mainly worked and designed the SCIP. The SCIP author interviewed 58% of all patients (Table 2) and 21% of patients tested for inter-rater reliability (Table 3). The SCIP method diagnoses the main psychiatric disorders. However, some psychiatric disorders (e.g. eating disorders, dementia, child psychiatry and personality disorders) were not tested in this study. Those sections can be the subjects of investigation in new studies. No community samples were used in the SCIP study, limiting the generalizability of the SCIP. The first author designed the SCIP for psychiatrists to use in clinical settings and he used the SCIP in his full-time clinical duties (the average SCIP assessment of a new patient was 30–40 minutes). Whether psychiatrists will use the SCIP or not remains to be seen. A formal study on the acceptability of the SCIP by psychiatrists is underway.

There are several factors that can explain the high Kappa of the SCIP items. The reliability results shown were based upon an inter-rater

design that gives too rosy an impression of reliability because all variation in conducting the interview was eliminated. Reliability is lower in homogenous populations such as community samples and higher in heterogeneous populations such as clinical samples (Shrout et al., 1987). The SCIP was designed and tested in clinical populations only (inpatient and outpatient) and that also contributed to high reliability results. The SCIP method of navigation between the screening questions and module questions reflects what seasoned psychiatrists do in actual diagnostic interviews using the “Bottom First Then Top (BFTT)” approach described earlier. For example, during the screening part of questioning, the patient may deny depressed mood, anhedonia, euphoric mood, irritable mood, labile mood or any mood swings. However, the interviewer may observe and collect collateral data to indicate mood problems (e.g. pressured speech, impulsive behavior or past suicidal attempts). In this case, the interviewer proceeds to the mood module and explores manic and depressive symptoms and signs. If the patient's responses to screening questions on mood are negative (happy and stable mood) and there is no collateral information or any history to indicate mood problems, the interviewer does not need to proceed to the mood module. In that case, mood questions are not asked, but assumed to be “not present” and coded “0” in the SCIP data. This assumption is based upon the clinician's certainty that the patient does not have mood problems. Only an experienced clinician can make that clinical judgment. This technique helps the clinicians to utilize clinical skills to conduct efficient interviews. The disadvantage is that it contributes to an overestimation of reliability indices.

To conclude, the SCIP is a reliable instrument for assessing psychological symptoms, signs and dimensions of the main psychiatric disorders among adults. The computer software designed to help clinicians administer, score, and aggregate SCIP data was developed. The use of modern software technology to create a reliable psychological database is a necessary match to the advances of neurosciences.

#### Role of funding source

This work was not funded by any pharmaceutical companies or any private or governmental agencies.

#### Contributors

Dr. Aboraya is the author of the SCIP and he designed the study protocol. All authors participated in interviewing patients and have approved the final manuscript.

#### Conflict of interest

All authors confirm they have no conflicts of interest.

#### Acknowledgement

None.

#### References

- Aboraya, A., 2009. Use of structured interviews by psychiatrists in real clinical settings: results of an open question survey. *Psychiatry (Edgemont)* 6 (6), 24–28.
- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth edition*. American Psychiatric Association, Arlington, VA.

- Andreasen, N.C., Flaum, M., Arndt, S., 1992. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. *Arch. Gen. Psychiatry* 49 (8), 615–623.
- Bastiaens, L., 2011. Poor practice, managed care, and magic pills: have we created a mental health monster? *Psychiatr. Times* 28 (2), 1–4.
- Busner, J., Kaplan, S.L., Greco, N., Sheehan, D.V., 2011. The use of research measures in adult clinical practice. *Innov. Clin. Neurosci.* 8 (4), 19–23.
- Carmines, E.G., Zeller, R.A., 1979. *Reliability and Validity Assessment*. SAGE, London.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20 (1), 37–46.
- Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16 (3), 297–334.
- Duffy, F.F., Chung, H., Trivedi, M., Rae, D.S., Regier, D.A., Katzelnick, D.J., 2008. Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatr. Serv.* 59 (10), 1148–1154.
- Fleiss, J., 1981. *Statistical Methods for Rates and Proportions*. John Wiley & Sons Inc., New York.
- Gilbody, S.M., House, A.O., Sheldon, T.A., 2002. Psychiatrists in the UK do not use outcomes measures. National survey. *Br. J. Psychiatry* 180, 101–103.
- Gliem, J.A., Gliem, R.R., 2003. Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Midwest Research-to-Practice Conference in Adults, Continuing, and Community Education. Ohio State University, Columbus, Ohio, pp. 82–88.
- Helzer, J.E., Kraemer, H.C., Krueger, R.F., Wittchen, H.U., Sirovatka, P.J., Regier, D.A., 2008. *Dimensional Approaches in Diagnostic Classification: Refining the Research Agenda for DSM-V*. American Psychiatric Association, Arlington, VA.
- Kraemer, H.C., Periyakoil, V.S., Noda, A., 2002. Tutorial in biostatistics: Kappa coefficient in medical research. *Stat. Med.* 21, 2109–2129.
- McHugh, P., Slavney, P., 2012. Mental illness—comprehensive evaluation or checklist? *N. Engl. J. Med.* 366 (20), 1853–1855.
- Morrison, J., 2008. *The First Interview*. The Guilford Press, New York.
- Nasrallah, H., 2009. Long overdue: measurement-based psychiatric practice. *Curr. Psychiatry* 8 (4), 14–16.
- Nienhuis, F.J., Van De Willige, G., Rijnders, C.A., De Jonge, P., Wiersma, D., 2010. Validity of a short clinical interview for psychiatric diagnosis: the mini-SCAN. *Br. J. Psychiatry* 196 (1), 64–68.
- Nunnally, J.C., Bernstein, I.H., 1994. *Psychometric Theory*, Third edition. McGraw-Hill, New York.
- Picardi, A., Adler, D.A., Rogers, W.H., Lega, I., Zerella, M.P., Matteucci, G., Tarsitani, L., Caredda, M., Gigantesco, A., Biondi, M., 2013. Diagnostic accuracy of the Primary Care Screener for Affective Disorder (PC-SAD) in primary care. *Clin. Pract. Epidemiol. Ment. Health* 9, 164–170.
- Pinna, F., Deriu, L., Lepori, T., Maccioni, R., Milia, P., Sarritzu, E., Tusconi, M., Carpiello, B., 2013. Is it true remission? A study of remitted patients affected by schizophrenia and schizoaffective disorders. *Psychiatry Res.* 210 (3), 739–744.
- Rettew, D.C., Lynch, A.D., Achenbach, T.M., Dumenci, L., Ivanova, M.Y., 2009. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int. J. Methods Psychiatr. Res.* 18 (3), 169–184.
- Rodgers, B., Mann, S.A., 1986. The reliability and validity of PSE assessments by lay interviewers: a national population survey. *Psychol. Med.* 16 (3), 689–700.
- Schene, A.H., Koeter, M., Wijngaarden, B., Knudsen, H.C., Leese, M., Ruggeri, M., White, I.R., Vazquez-Barquero, J.L., 2000. Methodology of a multi-site reliability study. *EPSILON Study 3. European Psychiatric Services: Inputs Linked to Outcome Domains and Needs*. *Br. J. Psychiatry (Suppl. 39)*, s15–s20.
- Schutzwahl, M., Kallert, T., Jurjanz, L., 2007. Using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN 2.1) as a diagnostic interview providing dimensional measures: cross-national findings on the psychometric properties of psychopathology scales. *Eur. Psychiatry* 22 (4), 229–238.
- Sheehan, D., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 (Suppl. 20), 22–33.
- Shrout, P.E., Spitzer, R.L., Fleiss, J.L., 1987. Quantification of agreement in psychiatric diagnosis revisited. *Arch. Gen. Psychiatry* 44 (2), 172–177.
- Spitzer, R.L., Williams, J.B., Gibbon, M., First, M.B., 1992. The Structured Clinical Interview for DSM-III-R (SCID). I: history, rationale, and description. *Arch. Gen. Psychiatry* 49 (8), 624–629.
- STATA, 2009. *Stata Statistical Software*. StataCorp LP, College Station, TX.
- Ustun, T.B., Tien, A.Y., 1995. Recent developments for diagnostic measures in psychiatry. *Epidemiol. Rev.* 17 (1), 210–220.
- WHO, 1993. *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. World Health Organization, Geneva, Switzerland.
- Williams, J.B., Gibbon, M., First, M.B., Spitzer, R.L., Davies, M., Borus, J., Howes, M.J., Kane, J., Pope Jr., H.G., Rounsaville, B., Wittchen, H., 1992. The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Arch. Gen. Psychiatry* 49 (8), 630–636.
- Wing, J.K., Babor, T., Brugha, 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 (6), 589–593.