Original Article

Inter-Rater Reliability of the Short Arabic Form of The Present State Examination (PSE-10)

Noha Sabry.

Faculty of Medicine, Cairo University, Cairo, Egypt.

Date received: 1/1/2009
Date accepted: 4/3/2009

Background: The PSE-10 (WHO-SCAN, 1992) is a reliable and standard international instrument for the detection of symptoms of mental disorder. It is time consuming, requiring 2-3 hours to apply, and can only be used by trained clinicians, The Short Arabic version of the PSE (10) has been developed in conjunction with a large scale epidemiological survey of the prevalence of mental disorders in Southern Egypt. This was done in order to avoid a two stage assessment process, and shorten the assessment time.

Aim of the study: To assess the inter rater reliability of the Short Arabic Form of the PSE-10 so as to introduce it as a research instrument that can be used in epidemiological studies in Arab countries.

Methods: 160 subjects: 109 (68.1%) males and 51 (31.9%) females were each assessed by two trainees to make a total of 320 assessments by 32 trainees. Each trainee had a minimum of 2 years psychiatric training All trainees attended a 5 day course application of the PSE .At the end of training, trainees were divided into 6 groups each supervised by a senior supervisor. Each team of two trainees assessed 10 cases. Both raters made independent symptom ratings, rated the level of confidence as to the clinical severity of any detected symptoms, and recorded a diagnosis according to DSM IV at the end of each section when applicable. Each rater made his/her final diagnosis independently and no change was made in subsequent discussions. The degree of agreement concerning individual items, case status, and final diagnosis are calculated using Kappa. Internal consistency of the different subscales of the Short Arabic PSE is measured using Cronbach's coefficient alpha.

Results: .Inter-rater reliability of individual symptoms is very high levels for most symptoms. Only 4 symptoms namely sudden loss of movement, irritability, situational anxiety and conscious resistance of obsessive thoughts and acts have moderate degrees of reliability. Inter rater agreement on the degree of certainty about caseness: this diagnostic decision distinguishes between those with no or minimal symptoms, sub clinical cases that do not require intervention, likely and definite cases of mental disorder. Raters agreed on caseness in 48 subjects (30%). There were excellent degrees of agreement on diagnosis: 20 cases have anxiety disorders (k=0.91), 14 cases had mood disorders(k=0.86), 9 cases have psychotic disorders (k=0.88) , and 5 cases have a somatoform disorder (k=0.82).

Conclusion: The results of this study illustrate that the short Arabic PSE is a reliable interview with adequate levels of internal consistency. The Short Arabic version is less time consuming, and still retains the core symptoms required to reach a diagnosis of the common mental disorders

Keywords: Reliability, Inter rater, Short Arabic PSE, SCAN, Caseness.

Abbreviations:
PSE: Present State Examination
SCAN: Scheudules for Clinical Assessment in Neuropsychiatry
PSE/CATEGO-4: Present State Examination/Computer Assissted Versionof the Text
WhO:World Health Organization

Egypt. J. Psychiatry, June 2009; 29(2): 45-54

INTRODUCTION

The present State Examination (PSE-10) type of interview as administered by experienced clinicians can be considered as a "Standard", by which other interviews should be assessed, especially in real life situations which have the aim of detecting cases in the community (Farmer, et al. 2002).

PSE is the central component of the Schedules for Clinical Assessment in Neuropsychiatry (WHO, 1992). The English version of the SCAN (Schedule for Clinical Assessment in Neuropsychiatry) was developed by Wing, et al. (1990) as an updated form of PSE-9 (Cooper, et al. 1977) .The overall aim of SCAN is to provide comprehensive, accurate and standard means of describing and classifying clinical phenomena. The first, clinical aim is to promote the use of high-quality clinical observation. The PSE is designed to allow comparison of each respondent's experience and behavior against the examiner's glossary-defined concepts by a process of controlled clinical 'Cross-

examination'. The resulting output is in the form of single symptom ratings. Categories of disorders can be compared with each other by demonstrating the underlying symptoms in a reliable manner. The second, scientific aim is to accelerate the accumulation of knowledge. Using standard procedures in research projects makes the results more comparable, thus leading to more agreement on theoretical lines for further testing (Wing and Brugha, 2001). Several versions of SCAN and PSE were used in numerous studies and translations in many corners of the world, both in sample populations , in a number of community surveys, and in experimental research (e.g Henderson, et al. 1979; knights, et al. 1980; Okasha and Ashour, 1981; Sturt, 1981; Mavreas and Bebbington, 1988; Huxley, et al. 1989; Brugha, et al. 1990; Lehtinen, et al. 1990 and Jenkins, et al. 1997).

The original version of the PSE-10 (Wing, et al. 1990) has been translated into colloquial Arabic (Hamdi, et al. 1993). Revision and correction were made. The difficult items were checked out and rephrased by using Arabic expressions instead of English ones. Face validity and acceptability was checked. Back translation has been compared with original English for similarity of meaning and content by senior bilingual Psychiatrists. PSE-10, Arabic version fulfills the requirements of trans-cultural equivalence concerning its content, syntax, criteria and conceptual equivalence.

One of the main problems concerning the use of PSE 10 is the extended time of application (Hamdi, et al. 1994), which makes its use for screening in epidemiological surveys difficult. The second difficulty with its use is the requirement that it is applied by clinicians (Clinical psychiatrists) International projects are coordinated by World Health Organizations to develop versions suited to particular settings, such as the survey form and short form more suited to service delivery and screening purposes (Wing and Brugha, 2001).

The Short Arabic version of the PSE 10 (Huxley, et al. 1989) has been developed in conjunction with a large scale epidemiological survey of the prevalence of mental disorders in Southern Egypt. It is based upon the following principles:

1. Coverage of the common mental disorders that may be encountered in the community.

2. Satisfaction of the diagnostic criteria of each surveyed disorder.

3. Adoption of the template of similar English versions (Harris, et al. 2000).

STUDY AIM

To assess the inter rater reliability of the Short Arabic Version PSE-10 so as to introduce it as a research instrument that can be used in epidemiological studies in Arab countries. The Short version is less time consuming, but still retains the symptoms required to reach a diagnosis of the common mental disorders included in the original English- Arabic version.

Design of the Instrument:
Selected sections of PSE-10 are identified to represent the common psychiatric symptoms required to make a diagnosis of the common mental disorders found in the community. The respective PSE-10 sections are:

- Physical Health and somatoform disorder.
- Worry/Tension.
- Anxiety /phobias.
- Obsessions.
- Thinking and concentration.
- Depression.
- Bodily function.
- Expansive Mood and Ideation.
- Perceptual disorders.
- Thought disorder and experience of replaced will
- Delusions.

Sections related to cognitive disturbance and substances use disorders have been excluded. However verification of the impact of substance misuse and organic causality are assessed during the interview through the general rating scales applicable to all symptoms.

The PSE-ID-CATEGO program (Wing, 1983, 1994) is a computer derived process of deciding whether a particular interviewee is suffering from a mental disorder and arrives at a diagnosis by introducing the scores on the PSE 9-10 interview. The computerized process has been the subject of criticism in the literature after long experience, was that many users tended to regard the computerized output only in terms of a single diagnosis, rather than a rich and varied psychopathology profile (Wing and Brugha, 2001).

With the short Arabic version we substituted this process by clinical decision making. At the end of each section of the interview, covering a group of related disorders, the interviewer has to make two decisions. The first decision involves how certain does he/she feel about caseness (Case status) as indicated by the presence of key symptoms and their severity. Four levels are mentioned: no symptoms, sub clinical symptoms, probably a case, definitely a case. The second pathway gives the assessor the opportunity to make interim diagnoses concerning the disorders within a section. Four diagnostic boxes have been added to the interview following the sections of: health and somatoform disorder, anxiety disorders, mood disorders and psychotic disorders to evaluate the degree of certainty about caseness. Four boxes allow making interim diagnostic decisions according to DSM IV. Questions are added to the PSE-10 to assess the diagnostic criteria according to DSM-IVTR (APA, 2000).

These include the onset and course of the suspected disorder and any additional symptoms not present in the English version of the PSE-10. The questions concerning onset include the temporal relation with substance abuse and organic diseases. This addition has been permitted in the recent versions of the PSE. It is to be noted that the time limit for symptoms is not exclusive to the previous month as in the PSE-9.

SUBJECTS AND METHOD:

Description of the Sample:

160 subjects 109 males (68.1%) and 51 females (31.9%) were each assessed by two trainees to make a total of 320 assessments by 32 trainees who attended the training. Patients were recruited from the outpatient clinics and inpatient units of mental health hospitals, Kasr El-Aini, El Menia University and the general population. The aim of selection was to ensure a heterogeneous group of patients covering most symptom areas with different grades of severity. The patient characteristics are shown in (Table 1).

Table 1: Sample characteristics:

Sex		
• Male	109	68-1%
• Female	51	31-9%
Marital status		
• Married	86	54%
• Separated	1	0.6%
• Divorced	4	2.5%
• Widow	6	3.8%
• Single	61	38.6%
Education		
• Illiterate	68	42.5%
• Read & write	10	6.25%
• Undergraduate	44	27.5%
• University graduation	31	19.4%
• Post graduate	7	4.4%
Occupation		
• Full time	52	32.5%
• Part time	49	30.6%
• Not working	31	19.4%
• Retired	4	2.5%
• House wife	18	11.25%
• Student	6	3.75%

Training:

Thirty two trainees (With at least 2 years experience in practicing psychiatry) were trained in 3 training courses. Training was done for 5 days and it consists of revision of all sections of the short Arabic PSE and discussion of individual items. Training was conducted by a team of 5 trainers with experience in the PSE, The senior trainer (EH) attended two complete PSE training courses, and MR has been certified after one training course. The trainees were trained to rate the presence and severity of symptoms. This was followed by interviewing patients by a trainer, and scoring by all the trainees individually, with discussion of scoring. Trainees were also invited to practice interviewing in front of the training group. A Parallel process was done aimed at testing the understanding and the fluency of Arabic wording of symptoms.

At the end of training, trainees were divided into 6 groups each supervised by a senior supervisor. Each team of two trainees assessed 10 cases. One applied the interview and the second trainee was given the opportunity to ask clarifying questions at the end of each section covered. Both raters made independent symptom ratings, level of confidence as to the clinical severity of any detected symptoms, and recorded a diagnosis according DSM IV at the end of each section when applicable. Each rater made his diagnosis independently and no change was made in subsequent discussions.

Discussion between raters and the supervisor was done after the interview concerning scoring and diagnosis without alteration which was not allowed.

Data Management and Statistical Analysis:

The data were coded and entered on an IBM compatible computer using the statistical package SPSS vers.15.o.The 160 observed records by the two raters were assessed for inter-raters reliability. The degree of agreement concerning individual items, case status, and final diagnosis are calculated using Kappa. The Kappa value ranges from “0” in the case of absent agreement to “1” in case of perfect agreement. The kappa value was considered to be statistically significant at a p-value <0.05. Table (1): We used Fenistein, (1985) categorization of Kappa results.

0.21-0.4	Fair reliability
0.41-0.6	Moderate reliability
0.61-0.8	Strong reliability
0.81-1.0	Almost perfect

Internal consistency of the different subscales of the Short Arabic PSE is measured using Cronbach’s coefficient alpha. Essentially, alpha is a measure of the degree to which individual items that make up a score agrees with total score.

Cronbach’s alpha values varying between 0.6-0.8 are considered moderate but satisfactory and a score above 0.8 is seen as representing high internal consistency.

Cronbach’s alpha was calculated for the subscales of anxiety disorders, depressive disorders, and psychotic disorders. The individual items in the section on

physical, somatoform and dissociative disorders were heterogeneous and do not allow analysis because few variables are assigned to each diagnostic category

RESULTS

The results of the assessment of reliability of the short Arabic form of the PSE 10 can be divided into three sets; levels of inter-rater agreement on symptom presence and severity, levels of agreement on caseness, and levels of agreement on diagnosis.

Internal consistency measured by Cronbach's alpha of the subcategories of anxiety, depression, and psychotic disorders revealed that when the symptoms of anxiety and the diagnostic categories based upon these symptoms are studied Alpha is 0.81, for the mood symptoms and diagnoses Alpha is 0.76 ,and for psychotic symptoms and diagnoses it is 0.82. These are high levels of internal consistency for the different subscales.

Inter-rater reliability of individual symptoms:
Figure (1) reveals that the symptom inter rater reliability when measured by kappa was either high or very high for most symptoms. Obsessions have the lowest degree of agreement (Moderate).

In the somatoform and dissociative disorders and physical illness: average Kappa agreement for all items is about (0.78): The highest agreement is for 'Trance experiences' (0.92) followed by 'dissociative amnesia' (0.85) and degree of subjective feeling of physical health (0.85). The least agreements (0.54) were obtained in 'Conversion symptoms, and sudden loss of movement or abilities'.

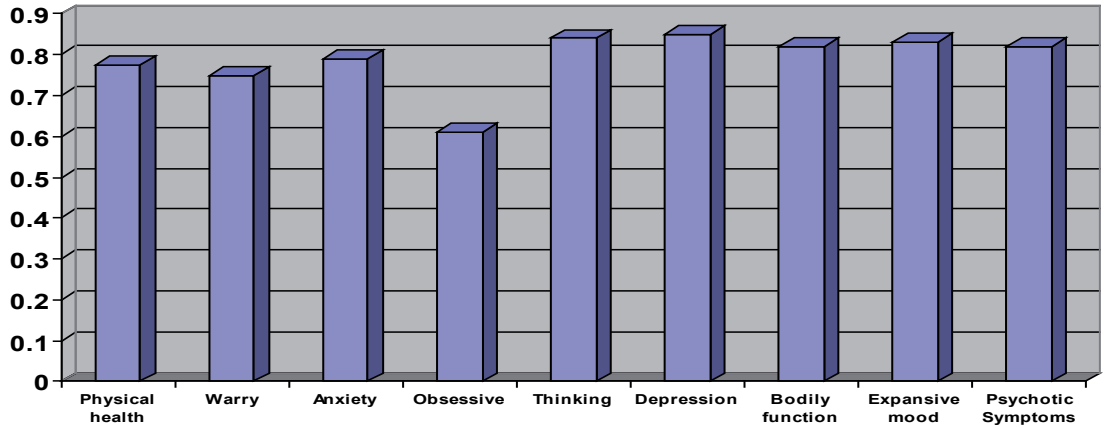

Fig 1: Average Kappa agreement for all symptom categories:

For the section on worry/tension: the average degree of agreement is 0.75, the highest degrees of agreement were found on 'interference in activity due to symptoms' (0.90) , followed by ideas of reference (0.85), 'Worrying' (0.83), and 'Tiredness and exhaustion (0.82).The lowest levels of agreement were detected in' calmness in the face of problems' (0.55) ,and 'Lrritability' (0.59).

Concerning symptoms of anxiety/phobia the average degree of agreement was 0.79 The highest degree of agreement were obtained in 'Panic attacks' (0.9), 'Enduring apprehension following panic attacks' (0.90), and 'Avoidance of situations of phobic experience' (0.90). The least agreements were in detecting 'Situational anxiety (0.61), and 'Specific phobia' (0.624).

As has been mentioned obsession showed the lowest Kappa (0.61): the highest degree of agreement detection of 'Obsessional checking' (1) and, the least agreement was in 'conscious resistance to obsessions' (0.64).

In section (5) concerned with thinking, concentration/ energy/interest: the average degree of agreement (0.84): The range of agreement about the different symptoms Kappa (0.93–0.78) the highest for 'Brooding' and least was for' loss of energy/drive.

Section (6) depression: the average degree of agreement (0.85). The highest degree of inter-rater reliability were for 'Suicide plans/acts' (0.94), 'Tedium vitae' (0.92), and 'Loss of self confidence' (0.92). The least agreement was about morning depression (0.57).

Concerning section (7) bodily functions: the average degree of agreement (0.9), the degree of agreement ranged between (0.91) for "Sleep terrors" and (0.75) for delayed sleep.

Section (8) expansive mood: the average degree of agreement (0.9): The highest degree of agreement was about 'Expansive mood' reported in (1), and the least was in the symptom of 'self reported over activity' (0.69).

Sections of psychotic symptoms (9,10,11) include perceptual disorders, subjectively described thought disorder and experience of replaced will and delusions scored (0.9) as an average degree of agreement: the maximum degree of agreement (1) was obtained in diagnosing cases ,and delusions of persecution, 'Auditory hallucination' reported in with kappa (0.8), while the least obtained in symptom of 'Replacement of will by external force' (0.66), 'delusions of being spied upon' (0.63) and 'Visual hallucination' (0.66).

The overall comment is that the great majority of items demonstrate high or very high levels of interrater agreement.

Table (2) outlines the same information mentioned above with the added presentation of the symptom reliability comparisons across categories. Only 4 symptoms namely sudden loss of movement, irritability, situational anxiety and conscious resistance of obsessive checking have moderate degrees of reliability and may require some form of revision of their definitions and scoring. On the other hand , very high levels of agreement in most symptoms in all domains are obtained.

Table 2: Inter rater reliability between different symptoms:

	Moderate (0.41-0.6)	Strong (0.6-0.8)	Almost Perfect (0.81-1.00)
Somatoform	Sudden loss of movement	Physical health (Dr)- Possession experience, Dissociative convulsions, Derealisation(people)	Hypochondriasis, physical health (subjective), Dissociative amnesia, Trance experience
Worrying	Irritability	Headaches, Muscular tension, Restlessness, Nervous tension, Sensitivity to noise	Worrying, Exhaustion, Ideas of reference
Anxiety	Situational anxiety	Free floating anxiety, Anxiety-meeting people, Specific phobias	Anxious foreboding, Panic attacks, Enduring apprehension, Action taken to prevent panic attack, Avoidance, interference
Obsessions	Conscious resistance of obsessive check	Obsessional hesitation and indecision, Ruminations ,obsessional actions associated with excessive orderliness	Obsessional checking
Thinking		Concentration, Nature of brooding, Retardation, Loss of energy	Inefficient thinking, Brooding, loss of interest, Feeling of being overwhelmed
Depression		Capacity for enjoyment, feeling of loss of feeling, Loss of self-esteem, Delusion of reference, Delusion of persecution, Guilty ideas of reference, Pathological guilt	Depressed mood, Suicide, Tedium vitae, Interference with activities.
Bodily functions		Delayed sleep, Early waking, Poor quality sleep	Weight loss, Weight gain, Excessive sleep, Middle insomnia, Nightmares, Sleep terrors, Loss of libido, Intercourse unpleasant, Premenstrual exacerbation
Expansive mood		Self reported over activity, Decreased need for sleep	Expansive mood, Over talkativeness
Psychotic symptoms		Auditory hallucinations, Visual hallucinations, Del. mood, Replacement of will, Del. of being spied, Del. of reference, del. Misinterpretation, Del. of jealousy	Del. that thoughts are being read, Thought insertion, Thought broadcast, Thought withdrawal

Interrater Agreement on Case Status:

Table (3) illustrates the degree of certainty about the existence of mental disorder between two clinicians assessing a person in the community. The inter rater reliability concerning the certainty about caseness which differentiate between the symptom free, sub clinical case and definite case of mental disorder were 0.88 i.e. within the strong correlation. The degree of certainty in diagnosing the somatoform disorders was 0.70, for the anxiety disorders Kappa=0.72 which is also considered strong correlation. Concerning the degree of certainty in diagnosing mood disorders and psychotic disorders, they showed similar degree of agreement (Kappa=069).It is apparent from these findings that certainty about caseness improved at the end of the interview compared to its beginning ie when the assessment was completed.

Table 3: Degree of certainty about caseness:

	Not a case		Sub clinical cases		Probable		Definite		Under treatment		Kappa
	N	%	N	%	N	%	N	%	N	%	
Somatoform disorders	116	72.5	9	5	2	1.25					0.70
Anxiety	60	37.5	85	11.87	53.12	11.87	16	10			0.72
Mood disorder	100	62.5	20	12.5	6	3.75	3	1.87			0.69
Psychotic disorder	131	81.87	2	2.5	3	1.87	4	2.5			0.69
Final Case Status	88	55	25	21.8	10	6.25	7	4.37	6	3.75	0.88

Table (4) shows the inter rater agreement on the main diagnostic categories. Raters agreed on the presence of a definite disorder in only 48 cases (30%) of the subjects. There is excellent agreement:on 20 cases (12.5%) with anxiety disorders (kappa=0.91), 14 cases (8.75%) with mood disorders (kappa=0.86), 9 cases (5.62%) having a psychotic disorders (kappa=0.88), and 5 cases (3.12%) gave a diagnosis of somatoform disorder (kappa=0.82). It is to be noted that raters could obtained more specific diagnoses according to DSM-IV. This aspect is not detailed further in this study due to the small number of cases in each category.

Table 4: Inter rater reliability concerning main diagnostic categories:

Diagnosis	N =valid cases	%	Kappa	p
Physical & somatoform disorders.	5	3.12%	0.82	0.000
Anxiety disorders	20	12.5%	0.91	0.000
Mood disorders.	14	8.75 %	0.86	0.000
Psychotic disorders.	9	5.62%	0.88	0.000

DISCUSSION

This study demonstrates high levels of inter-rater reliability of the short Arabic form of the PSE. Psychiatrists in training were able to agree on the majority of diagnostic decisions, the level of confidence in clinical severity, and the absence/presence of key symptoms of mental disorder. Decisions tend to give limited rates of definite cases. In addition a spectrum of different levels of certainty about caseness which improved further as the final decision about caseness is reached at the end of the interview. Most important is that the time of the interview has been reduced from an average of two hours in the original unabbreviated version to an average of 30-45 minutes with the short one.

The design of the study requires a number of comments. Thirty two raters participated in the study, creating 16 pairs of raters working together, with each team making around 10 subject interviews. This suggests that the high inter-rater agreement arrived at is not due to a single pair of raters who work closely together. Inter rater reliability reflects consistency of the measure when used with different raters (Wittenborn, 1972) All raters in this study were psychiatrists of at least 2 years experience. The level of agreement that can be achieved with raters who have less experience in psychiatry is not well known. It is also likely that the standardized interview guide facilitates the reliable use of this scale by less experienced raters because it standardizes the questions used to elicit the information necessary to rate the items

Both raters assessed the subject in the same live interview. This reduces the differences in symptoms ratings arising from differences in test circumstances. These include changes in individuals' memory, and different responses in different interviews, subject fatigue or desire to hasten completion of the test, and changes in the subject condition in different interviews. This approach may eliminate several problems that confound the assessment of this form of reliability (Farmer, et al. 2002). In other words, rating one interview decreases information variance, which refers to the quality and quantity of the originally collected psychopathological information (Jansson and Parnas, 2007). The significance of information variance is illustrated by higher kappa found in rating live or videotaped interviews than in rating hospital charts (Helmes, et al. 1983) .

The relatively high inter rater reliability may be explained by the small number of cases in diagnostic subcategories. This took place in spite of the reasonable number of subjects included in this study. The majority of subjects are mentally healthy individuals with only few symptoms. Including community subjects was deemed more appropriate for this study which is part of a larger community survey of mental illness. Jansson and Parnas, (2007) stated that the study of reliability in clinical samples cannot be un-problematically extrapolated to community samples.

Questions about false positive agreements that may arise due to the relatively high inter rater reliability can be elaborated by reporting certain comments. The same degree of agreement within different symptoms was found with different frequencies. As an example, in the section of physical and somatoform kappa=0.84 was similar for items suggest subjective problems in physical health in 38 (23%) of the cases and in dissociative amnesia in 5 subjects (3.125%). Also this means that the exact significance of quantifying reliability is not unequivocal, as the magnitude of kappa coefficient may reflect differences in prevalence rates (Spitzer

and Fleiss, 1974). Excellent degree of Kappa obtained in inter rater reliability for some psychotic symptoms e.g. delusions of persecution, delusions that thoughts are being read, thought insertion, thought broadcasting and thought withdrawal. All these symptoms are easy to elicit especially with psychiatrists working in mental hospitals. Another point was that the degree of rating was similar even on the degree of severity in the rating scale. In fact this was similar to the Chinese cross cultural inter rarer reliability study (Cheng, et al. 2001), but rather different from Hamdi, et al. (1994).

On the other hand, another areas like obsession cleanliness, loss of hope, social withdrawal, actions based on expansive mood and exaggerated self esteem, the degree of agreement can't be calculated as one of the raters was extremely doubtful to make the rating (In spite of the possibility of using (8) rating which we considered the most appropriate from the rater point of view as it indicates indecisiveness. Most of these symptoms required through objective evaluation so it seemed to be difficult to judge.

When exploring the question of reliability, criterion variance, which refers to the differences in the raters use and interpretation of the diagnostic criteria, may be considered one source of diagnostic disagreement (Jansson and Parnas, 2007). In this study the high degree of kappa agreement among different symptoms can be understood as the interviewers were psychiatrists who possibly have the scope for probing the subject understanding of questions and being able not to take answers at face value (Brugha, et al. 1999). This is one of the privileges of the SCAN method of interviewing namely being flexible and based on detailed questioning of the subject to enable an interviewer to judge symptom presence and severity. This characteristic of the PSE is maintained in the short version by providing the interview with different alternatives of wording for each item. The trial to present the translation in a vernacular language ensures better understanding of the patients and helps the assessor go beyond the apparent meaning of words. This provides "Optimally valid questions" (Wittchen, et al. 1999) to the interview.

Certain symptoms such as depersonalization and calmness in face of problems were rather difficult to explore. This can be explained by the difficulty of grasping the concept in case of depersonalization. When asking about calmness in facing problems the subject is probed about positive healthy coping which may reduce understanding. The same finding was reported in a Chinese study of inter-rater reliability (Cheng, et al. 2001).

The PSE was not originally designed as a diagnostic instrument, Rose, (1992) has made the point that diagnosis splits the world in two those who have, and those who do not have a disorder. The short version incorporates the principles of making a decision about caseness adopted by the PSE-Index of Definition (Wing, et al. 1990), giving a more dimensional view of the existence of mental disorders that agrees with clinical reality, and may be more appropriate for community surveys.

The cross-cultural testing of an earlier version of the SCAN at the level of diagnostic categories shows satisfactory levels of agreement (Jenkins, et al. 1997; Wing, et al. 1990), but at the important level of individual symptom items there are as yet no published reports of cultural equivalence or inter rater reliability study demonstrates high levels of inter-rater reliability (Cheng, et al. 2001).

As a rule, a diagnosis based on a few simple items becomes easily reliable compared with the diagnostic algorithms defined by many and interacting features. Yet, if a creation or a revision of diagnostic criteria is motivated by reliability concerns, the emphasis should be focused on the criterion variance because the information variance is basically related to the comprehensiveness of the assessment.

Restriction of the categories of the symptoms might limit the utility of the tool. It is difficult to make some diagnoses e-g eating disorder, gender identity disorders, impulse control, substance abuse and organic mental disorders probably. However the scale provides some questions to detect the history of substance abuse (Item: 0.8) in the introductory section,. In addition, if disorders have been suspected at the end of each section, questions to identify the onset include the possible substance abuse and general medical conditions that could be etiologically related.

CONCLUSION

The results of this study illustrate that the short Arabic version of the PSE is a reliable interview with adequate levels of internal consistency. Its main advantage is the assessment of the presence of common forms of mental disorder in a relatively short period of time. However, there are certain requirements if the instrument is to be successfully applied. Clinicians using the Short Arabic PSE for the first time should undergo satisfactory theoretical and practical training. This should be followed by pilot field application and review of scores with senior supervisors. The PSE should only be used as a method of eliciting clinical signs and symptoms that the clinician should then integrate to arrive at a diagnosis of the presence and nature of mental disorder. At present the PSE remains an instrument that cannot be used except by a clinician with sufficient previous experience in the assessment and diagnosis of mental disorders as it uses the interviewer's judgement to score symptoms.

REFERENCES

American Psychiatric Association. 2000. Diagnostic and statistical manual of mental disorders.4th ed. American Psychiatric Association.

Brugha, T. S., Bebbington, P. E., and Jenkins, R. 1999. A difference that matters: Comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. Psychological Medicine 29(5):1013-1020.

Brugha, T. S., Bebbington, P. E., MacCarthy, B., et al. 1990. Gender, social support and recovery from depressive disorders: A prospective clinical study. Psychological Medicine 20(1):147-156.

Cheng, A. T., Tien, A. Y., Chang, C. J., et al. 2001. Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. The British Journal of Psychiatry; The Journal of Mental Science 178:567-572.

Cooper, J. E., Copeland, J. R., Brown, G. W., et al. 1977. Further studies on interviewer training and inter-rater reliability of the Present State Examination (PSE). Psychological Medicine 7(3):517-523.

Farmer, A., McGuffin, P., and Williams, J. 2002. Measuring psychopathology.1st ed. USA: Oxford University Press.

Fenistein, A. R. 1985. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Saunders. Quoted from: Mental health outcome measures, edited by M. Tansella and G. Thornicroft. London: Gaskell (Royal College of Psychiatrics).

Hamdi, E., Askar,M., Halim,Z., Shafik,H. and El-Nagdi,A. 1994. Reliability of Present State Examination (PSE) in its Arabic version. Royal College Conference, Cairo.

Hamdi, E., et al. 1993. The Present State Examination (PSE-10) in its Arabic version.10th ed. (Unpublished manuscript – available from authors).

Harris, T. O., Borsanyi, S., Messari, S., et al. 2000. Morning cortisol as a risk factor for subsequent major depressive disorder in adult women. The British Journal of Psychiatry; The Journal of Mental Science 177:505-510.

Helmes, E., Landmark, J., and Kazarian, S. S. 1983. Inter-rater reliability of twelve diagnostic systems of schizophrenia. The Journal of Nervous and Mental Disease 171(5):307-311.

Henderson, S., Duncan Jones, P., Byrne, D. G., et al. 1979. Psychiatric disorder in Canberra. A standardised study of prevalence. Acta Psychiatrica Scandinavica 60(4):355-374.

Huxley, P., Raval, H., Korer, J., and Jacob, C. 1989. Psychiatric morbidity in the clients of social workers: clinical outcome. Psychological Medicine 19(1):189-197.

Jansson, L. B., and Parnas, J. 2007. Competing definitions of schizophrenia: What can be learned from polydiagnostic studies? Schizophrenia Bulletin 33(5):1178-1200.

Jenkins, R., Lewis, G., Bebbington, P., et al. 1997. The National Psychiatric Morbidity surveys of Great Britain--initial findings from the household survey. Psychological Medicine 27(4):775-789.

Knights, A., Hirsch, S. R., and Platt, S. D. 1980. Clinical change as a function of brief admission to hospital in a controlled study using the present state examination. The British Journal of Psychiatry; The Journal of Mental Science 137:170-180.

Lehtinen, V., Lindholm, T., Veijola, J., and Vaisanen, E. 1990. The prevalence of PSE-CATEGO disorders in a Finnish adult population cohort. Social Psychiatry and Psychiatric Epidemiology 25(4):187-192.

Mavreas, V. G., and Bebbington, P. E. 1988. Greeks, British Greek Cypriots and Londoners: A comparison of morbidity. Psychological Medicine 18(2):433-442.

Okasha, A., and Ashour, A. 1981. Psycho-demographic study of anxiety in Egypt: The PSE in its Arabic version. The British Journal of Psychiatry; The Journal of Mental Science 139:70-73.

Rose, G. 1992. The strategy of preventive medicine. USA: Oxford University Press.

Spitzer, R. L., and Fleiss, J. L. 1974. A re-analysis of the reliability of psychiatric diagnosis. The British Journal of Psychiatry; The Journal of Mental Science 125(0):341-347.

Sturt, E. 1981. Hierarchical patterns in the distribution of psychiatric symptoms. Psychological Medicine 11(4):783-792.

Wing, J. K. 1983. Use and misuse of the PSE. The British Journal of Psychiatry; The Journal of Mental Science 143:111-117.

Wing, J. K. 1994. Relevance of psychiatric epidemiology to clinical psychiatry. International Review of Psychiatry 6(4):259-264.

Wing, J. K., Babor, T., Brugha, T., et al. 1990. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Archives of General Psychiatry 47(6):589-593.

Wing, J. K., and Brugha, T. S. 2001. The schedules for clinical assessment in neuropsychiatry and the tradition of the present state examination. In Mental health outcome measures, edited by M. Tansella and G. Thornicroft. London: Gaskell (Royal College of Psychiatrics).

Wittchen, H. U., Ustun, T. B., and Kessler, R. C. 1999. Diagnosing mental disorders in the community. A difference that matters? Psychological Medicine 29(5):1021-1027.

Wittenborn, J. R. 1972. Reliability, validity and objectivity of symptom-rating scales. The Journal of Nervous and Mental Disease 154(2):79-87.

World Health Organization (WHO). 1992. Schedules for clinical assessment in psychiatry (SCAN-1). World Health Organization (WHO); Geneva.

الملخص العربى

درجة الثبات بين المقومين للنسخة العربية المختصرة لمقياس الحالة الحاضرة (الطبعة العاشرة)

نها أحمد صبرى
أستاذ مساعد الطب النفسي (كلية الطب-جامعة القاهرة)

الهدف من هذا البحث هو تقييم درجة الثبات بين المقومين للنسخة العربية المختصرة لمقياس الحالة الحاضرة ليتم تقديمها كأداة بحثية يمكن استخدامها فى الدراسة الوبائية فى الدول العربية. وقد تم تقييم 160 حالة. عدد 109 من الذكور و51 من الأناث، بواسطة اثنين من المقومين بحيث يتم فى النهاية عمل 320 تقييم بواسطة 32 مقيم ذو خبرة لا تقل عن عامين فى ممارسة الطب النفسى، وفى نهاية التدريب تم تقسيم المتدربون على 6 مجموعات كل تحت قيادة مشرف أكثر خبرة (أكبر)، ثم قسمت المجموعات إلى فرق ويتكون كل فريق من اثنين من المقومين يقوم بتقييم 10 حالات كل على حدة من حيث درجة الثقة فى تشخيص العرض ودرجة شدته، ثم تشخيص الحالات تبع لدليل التشخيص والإحصاء الأمريكي الرابع. ويقوم كل مقيم بعمل تشخيص مستقل وغير مسموح عمل أى تغيير فى التقييم. وقد تم تحليل المعلومات باستخدام مقياس (كابا) لكل بند على حدة وكذلك تحديد الحالات وايضاً التشخيص النهائى. كما تم قياس درجة التماسك الداخلى بمقياس كرونباخ.

وقد قسمت نتائج تقييم درجة الثبات إلى ثلاث مجموعات من حيث مستوى الثبات فى الاتفاق بين المقومين على وجود الأعراض وشدتها، وتحديد الحالات والإتفاق على التشخيص. وقد ظهرت درجة عالية من الاتفاق بين المقومين فى مستوى الأعراض فى جميع الأقسام الأساسية، وقد سجلت درجة متوسطة الثبات فى الاتفاق علي الأعراض فى الفقد المفاجئ للحركة (التحولى) والقلق الموقفى والاستثارة والمقاومة الواعية لوسواس التأكد. كما ثبت وجود درجة عالية من الثبات بين المقومين من حيث تقييم وجود الحالات والتى يمكن أن تفرق بين عدم وجود الحالات ووجود حالات دون المرضية أو وجود حالات مؤكدة، وكانت درجة الثبات علي مقياس 0,88 فى النهاية على اتفاق عالية وكانت على درجات الاتفاق على اضطرابات القلق. حالة 12,5% من العينة كما تم الاتفاق على تشخيص 14 حالة اضطراب للمزاج وكانت كابا 0.86 وشخصت 9 حالات ذهانية بدرجة ثبات 0.88 و5 حالات اضطرابات جسمانية بدرجة كابا 86. وقد أستخلصنا أن درجة الثبات للمقومين للنسخة العربية المختصرة الطبعة العاشرة درجة عالية ويسمح بتقديمه كأداة بحثية يمكن استخدامها فى الأبحاث الوبائية فى مصر والدول العربية.