

Comparative Study of the Reliability of an Egyptian Computerized Version of Beck Depression Inventory with Its Traditional Counterpart in a Sample of 268 Egyptian Depressed Patients

M. A. Owida

To detect the reliability of an Egyptian computerized version of Beck Depression Inventory (BDI) with its well known traditional counterpart, the present study was designed. A sample of 268 depressed patients were chosen along one year interval, half of them were males and half were females. All fulfilled the diagnostic criteria for depression according to the diagnostic criteria for research of ICD-10 and had a Hamilton Depression rating score above 18 through a semi-structured clinical interview. Half of the sample was tested by the traditional paper and pencil version of BDI and the other half was tested with an Egyptian computerized version of BDI developed by the author. Both groups were matched as regards sex, age, education, marital status, occupation and degree of severity of depression. Reliability was calculated through split-half technique with person correlation coefficient between the two halves of each group. The results indicated that the computerized version has a higher reliability degree than its traditional counterpart. The reliability of the computerized version was differentially affected by sex, age, education, marital status and occupation, being higher among females, younger in age, university graduated, unmarried and jobless.

(Egypt.J. Psychiat., 1999, 22: 231 -238).

INTRODUCTION

Every science develops methods for measuring and characterizing its phenomena. In comparison with most fields of medicine; psychiatry is at a disadvantage because the clinical manifestations of psychiatric disorders cannot be measured using physiological and biological parameters, such as blood pressure or bacteria in urine. As a result, the field has resorted to the use of rating scales (Marder, 1995). The rating scales are valuable instruments that are designed to translate and convert impressionistic and descriptive clinical phenomena into

objective and quantitative information. (Tyrrer et al., 1993). The information conveyed by a rating scale can be used to arrive at a diagnosis, to document the clinical state of a person at a point in time or to supplement the information about a patient that is provided by a clinical examination (Marder, 1995).

Hence, the importance of psychiatric rating scales in clinical practice and research has increased dramatically in recent years (Versavel et al, 1997).

However, the administration and scoring of the traditional paper and pencil rating scales present a host of problems to clinicians and researchers due to the human errors that may be encountered at any of these steps. And, that errors may compromise the reliability of

M. A. Owida, M.D. Assistant Professor in Psychiatry. Al-Azhar University.

the obtained data and obstruct its valid interpretation (Wysocki, 1990). Moreover, the time requirements associated with administering, scoring and interpreting rating scales as perceived by clinicians is relatively long, thence offer an additional obstacle for their routine daily use (Ball et al., 1994).

As a result of development, more efficient use of resources and the new technologists and to overcome the previously mentioned pitfalls, many automated i.e. computer-based versions of rating scales have been developed in the last two decades (Neal, 1994).

Validation and reliability studies of the newly developed automated rating scales have been an area of interest in psychiatric research along the present decade (Agnew et al., 1991 - Drak et al., 1995-Tien et al., 1996-Gonzalez et al., 1997).

In Egypt, the mid 1990's witnessed the appearance of few computerized versions of psychiatric rating scales. However, development and validation (standardization) of those tests have gone side-by-side. e.g. computerized version of Taylor Manifest Anxiety scale (Owida., 1995a) and a computerized version of Beck Depression Inventory (Owida., 1995b). Both versions were developed by the present author. In both versions test administrations, scoring and interpretation of test results were automated.

Beck Depression Inventory (BDI) is probably the most commonly used self-rating scale for depression. A number of studies have found that BDI correlates reasonably well with both the Hamilton Rating scale for Depression and global ratings by psychiatrists. In addition, the scale is sensitive to change and so is useful in clinical drug trials (Marder, 1995). Also, the traditional paper and pencil form is reliable, as it has been the subject of many studies concerning its reliability

and utilizing test-retest or split-half methods (Peck & Dean, 1988). In Egypt, reliability ranged between 0.77 to 0.91 (Gharied, 1990). To better understand the relationship of the newly developed computerized rating scales to their well known traditional counterpart, the present study is designed.

The present study aimed to detect the reliability of an Egyptian computerized version of Beck Depression Inventory in comparison with its traditional version in an Egyptian sample.

SUBJECTS AND METHODS

a-Sample:

The sample of the present study consisted of 268 patients representing all the attendants to a private hospital in Cairo and who met the following inclusion criteria:

.Egyptian nationality.

.Age above 18 years old.

.Fulfilled the diagnostic criteria for Depressive Episode in ICD-10 diagnostic criteria for research (WHO, 1993).

.Hamilton Rating scale for Depression score above 18.

.Half of the sample were males and the other half were females.

Half of the sample were tested by the traditional paper and pencil version of Beck Depression Inventory (BDI) and are representing the paper and pencil group and the other half were tested by the computerized version of BDI and are representing the computer group and they were chosen alliteratively. (i.e. one by one).

B-Method:

The current study was conducted along one year interval starting from 2

nd of July 1996 to 1st of July 1997. All the patients of the study were subjected to:

.Semi-structured clinical psychiatric interview guided by the diagnostic criteria for research of ICD-10 (WHO, 1993), with medical including neurological examination.

.Hamilton Rating scale for Depression which is a clinician rated scale. The scale is rated on the basis of the patient's condition in the previous few days. It contains 17 variables. The items are not defined but detailed information is given about scoring (Peck & Dean, 1988).

Half of the patients were subjected to the traditional version of Beck Depression Inventory (BDI) and they represented the paper and pencil group. The other half of the patients were subjected to a portable computer (Note book computer) loaded with necessary modifications to suite* Egyptian culture (Owida, 1996 b) these modifications did not affect the validity of the test as it is based on the original Egyptian version of BDI previously tested for validity by Gharieb (Gharieb, 1990).

Moreover, it matched the traditional paper and pencil version and this half represented the computer group.

Reliability was tested by split-half method comparing the results of the two halves of the test (i.e. odd/even items) in each group of patients and calculating the Person correlation coefficient between those two halves. Then the correlation coefficient was corrected using Spearman-Brown equation (Abu el Neil, 1984).

RESULTS

The study was carried out on 268 pa-

tients. Half of them i.e. 134 are males and the other half are females. Socio-demographic variables are presented in table (1) The mean age of the computer group was 31.7 ± 5.65 years, while the mean age of the paper and pencil group was 31.5 ± 6.08 year with no statistical significance to the difference between the two means. Most of the sample are married, 64.9% of the computer group and 61.9% of the paper and pencil group. 55.2% of the computer group and 56% of the paper and pencil group are average educated (secondary school graduates and its equivalents). Administrative works are representing the relatively commonly met occupations in both groups; 39.6% of the computer group and 41.8% of the paper and pencil group. There were no statistical significance differences between the two groups regarding the previously mentioned sociodemographic variables.

Table (2) shows the mean scores of Beck depression Inventory (BDI) of the two groups, which were 69.4 ± 5.87 for the computer group and 68.9 ± 6.11 for the paper and pencil group with no statistical significance to difference between the two mean scores. Table (3) shows that reliability of the computerized version of BDI was 0.883, which is relatively better than the reliability of the traditional paper and pencil version of BDI (0.821) But both versions had a very high statistical significance ($p < 0.001$).

Table (4) shows comparison between the reliability of the two versions according to socio-demographic variables of the patients. Results showed that the computerized version of BDI had a better reliability among females (0.932) than males (0.833) while the traditional paper and pencil version of BDI had better reliability among males (0.834) than females (0.818). The

Table 1
Socio-Demographic Characteristics of the Sample

Date	Computer group		Paper and Pencil group	
	No	%	No	%
Less than 20	10	7.5	14	10.4
20 - 39	85	63.4	79	59.0
40 and more	39	29.1	41	30.6
Mean	31.7		31.5	
Standard deviation	± 5.65		± 6.08	
T-Test	0.28		P> 0.05	
Married	87	64.9	83	61.9
Un-married	47	35.1	51	38.1
*C.R	0.51		P> 0.05	
Below average	22	16.4	27	20.1
Average	74	55.2	85	56.0
University and higher	38	28.4	32	23.9
x ²	1.07		P> 0.05	
Administrative	53	39.6	56	41.8
Technical	46	34.3	38	28.4
Jobles	35	26.1	40	29.8
x ²	1.21		P> 0.05	

* C. R. = Critical Ratio
* Number of each group = 134 patients.

Table 2
Comparison Between the Two Groups According to the Scores of Beck Depression Inventory (degree of severity of depression)

Variables	Computer group	Paper and Pencil group
No. of Patients	134	134
Mean scores	69.4	68.9
Standard Deviation	±5.87	±6.11
T. Test	0.68	
P	>0.05	

Table 3
Comparison Between the two Versions of BDI According to the Reliability as Measured by the Split-Half Technique

Variables	Computer group version	Reliability version
Reliability	0.883	0.821
P	<0.001	<0.001

Table 4
Comparison Between the Reliability of the Two Versions of BDI
According to Some Socio-Demographic Variables of the Patients

Variables	Computer group		Paper and pencil group	
	No	Reliability	No	Reliability
Males	67	0.833	67	0.824
Females	67	0.932	67	0.818
Less than 20	10	0.925	14	0.713
20 - 39	85	0.885	79	0.838
40 and More	39	0.859	41	0.629
Married	87	0.873	83	0.867
Un-married	47	0.908	51	0.712
Below average	22	0.868	27	0.661
Average	74	0.897	75	0.843
University and higher	38	0.961	32	0.692
Administrative	53	0.869	56	0.865
Technical	46	0.877	38	0.683
Jobless	35	0.922	40	0.744

youngest in age i.e. less than 20 years were doing better than the other age groups with the computerized version (0.925). While the middle ages i.e. 20-39 years were doing better (0.838) than the other age groups with the traditional paper and pencil version. The reliability of the computerized version of BDI was better among the un-married (0.908) than among the married patients (0.873) while the reverse occurred among the second group using the traditional version of BDI. Reliability of the computerized version of BDI was clearly better among university graduated patients (0.961) than among patients with other levels of education. Reliability of the computerized version of BDI was better among jobless patients in this sample (0.922) than among other occupations. While reliability of the traditional paper and pencil version was better among patients who had administrative jobs (0.865) than other occupations.

DISCUSSION

To better understand the relationship of the newly developed Egyptian computerized version of BDI to its traditionally administered counterpart, the present study was done. The two groups of the study i.e. the group of patients who were tested by the computerized version of BDI and the second group of patients who were tested by the traditional paper and pencil version, were well matched as regards: sex, age, marital status, educational level, occupation and the degree of severity of depression.

The results indicated that the computerized version was at least as reliable as the traditionally-administered version or better. This finding gave support to Agnew et al study in 1991 who compared a newly developed computerized neurobehavioral tests to their examiner-administered counterparts using a sam-

ple of 183 subjects (Agnew et al, 1991). Also, the results of the present study is consistent with Stokes et al in 1991 who found that the prototype version of automated (computerized) Mini-Mental Examination was more reliable than the traditional paper and pencil version (Stokes et al, 1991). Also, the present results went with the results of Neal et al (1994) who did a comparative study between a computer-administered version of Post traumatic Stress Disorder Scale and its clinician-administered version utilizing a sample of 40 subjects and found that the mean score difference between the two versions testing the same subjects was non-significant (Neal et al, 1994). Also, the present results agreed with the results of Tien et al in 1996 who compared the performance between the standard manual version of the Wisconsin card sorting test and the computerized version in a group of 33 psychiatrically ill subjects and concluded that the reliability of the computerized version is better than the reliability of its traditional version. (Tien et al, 1996). Based upon the results of the present study and the literature, we can conclude that the reliability of the computerized psychiatric test is better than the reliability of its traditionally administered version, whatever it is patient administered as the case in the present study and in Stokes et al study or clinician-administered as in Agnew et al and Neal et al studies and the results may be explained on the basis of the more reliable administration and accuracy in data acquisition and scoring in the computerized version.

Although the overall reliability was high, the results were differentially affected by sex, age, education, marital status and occupation. The reliability of the computerized version of BDI was relatively higher among females, younger in age, university graduates, unmarried and with the unemployed. Associa-

tion of relatively higher reliability with age and females agreed with Agnew et al results (Agnew et al, 1991), but no clues were found in the literature to support the rest of the present results. However, it seems that the other results i.e. university graduated, unmarried and jobless were all related to age, as the younger in age were university graduated or still in the university level and because of the outgoing socio-economic changes in the Egyptian society, most of the younger age groups are still unmarried and jobless. Also, it seems that the cornerstone of the present results was education: the younger, more educated took the opportunity to be accustomed with computers since their early years in schools, thanks to the policy of our Ministry of Education and / or it seems that usage of computer applications in our daily routine still new and this recent revolutionary trend goes with the naturally rebellants who were suppressed in the society i.e. the females, jobless, university educated and younger in age.

In conclusion, the new computerized version of BDI is at least as reliable as the traditional paper and pencil version. The advantages of the new test are that it is easy to operate, time saving, results are easy to be scored, analyzed, recorded and recalled with the least-if any-human errors. In addition, because it can be used with a computer which is easy to transport (portable forms), this version is suitable for use in all clinics especially in areas where trained psychologists are not available.

However, it is important to mention that the results overall indicate that the computerized version is not a substitute for a human examiner (as it lacks observations of the process by which the subject responds to the test situation) rather, the computer can function as a

reliable partner, carrying out the mechanics of test presentation and scoring, so freeing the examiner to more fully support the subject in taking the test and to observe the non-quantitative aspects of test performance.

REFERENCES

- Abu el Neil, M. (1984) Psychological, social and educational statistics Cairo, el-Khanki bookshop.
- Agnew, J. Schwartz, N. Bolla, K. Ford, D. & Elecker, M. (1991) Comparison of computerized and examiner-administered neurobehavioral testing techniques *J. of Occup. Med.*, Vol 33 (11). P: 1156-1162.
- Ball, J. Archer, R. & Imhof, F. (1994). Time requirements of psychological testing: a survey of practitioners *J. Pers-Assess* Oct. Vol. 63 (2), P: 239-249.
- Drak, R. Mc Hugo, G. & Biesantz, K. (1995) The test-retest reliability of standardized instruments among homeless persons with substance use disorders *J of Stud. on Alcohol*. Vol. 56 (2), P: 161-167.
- Gharieb, A. (1990) Beck Depression Inventory-Instruction, studies of reliability and validity Cairo. Al-Nahda Bookshop, 2 nd edition.
- Gonzalez, G. Costello, C. Joyce, L. & Valenzuels, M. (1997). Bilingual telephone-assisted computerized speech-recognition-assessment.
- Marder, S. (1995) Psychiatric rating scales in: Harold I. Kaplan and Benjamin j. Sadock: Comprehensive textbook of psychiatry Baltimore, Williams & Wilkins. 6 edition. Vol (1) P: 619-635
- Neal, L. Busuttil, W. Herapath. R. & Strike. P (1994) Development and validation of the computerized clinician administered Post-traumatic stress disorder scale *Psychol. Med.* Vol. 24 (3). P: 701-706.
- Owida, M (1995a). Computerized version of Taylor Manifest Anxiety Scale and its standardization on a Saudi sample from the Eastern area *J. Psychol.* Vol. 35, P: 94-103 (in Arabic)
- Owida, M. (1995 b) Computerized version of Beck Depression Inventory-and its standardization on a Saudi sample from the Eastern are *J. Psychol.* Vol. 36, P: 70-80 (in Arabic)
- Peck, F. and Dean, C. (1988) Measurement in psychiatry in R.E. Kendell and A.K. Zealley Companion to psychiatric studies Edinburgh, Churchill livingstone, 4 th edition, P: 224-241
- Stokes, A. Banich, M. & Elledge, C. (1991) Testing the tests-an empirical evaluation of screening tests for the detection of cognitive impairment in aviators *Aviat-Space-Environ. Med. Age.* Vol 62(8) P: 783-788.
- Tien, A. spevack, T. Jones, D. person, G. & Strauss, M. (1996) Computerized Wisconsin Card Sorting Test: comparison with manual administration Kao Hisung I. Hsuch KO Hsuch Tso-Chil, Vol 12 (8) P: 479-485 (abstract).
- Tyrer, P. Henderson, F. and Mc Dermott, U. (1993) Validation of rating scales in psychiatry *Brit. J. of Hosp. Med.* Vol. 49 (6), P: 434-437.
- Versavel, M. Vanlack, D. Evertz, C. Unger. S. Meier, F. & Kuhlmann, J. (1997) Test-retest reliability and influence of practice effects on performance in a multi-user Computerized psychomotor test system for use in clinical pharmacological studies *Arzneimittel foscjung*, June Vol 47 (6), P : 781-786 (Abstract).
- World Health Organization (1993) ICD-10 Classification of mental and behavioral disorders-Diagnostic criteria for research Geneva. WHO.
- Eysocik, I. (1990) Review of Q-fast interactive testing system *Res-Dev-Disabil*-Vol. 11 (1), P: 124-127.

دراسة لثبات نسخة مبرمجة مصريا على حاسوب من إختبار بك للاكتئاب مقارنة بنسخة مناظرة تقليدية على عينة من ٢٦٨ مريضاً مصرياً مصابين بالاكتئاب

هدفت الدراسة الحالية إلى معرفة درجة ثبات نسخة من مقياس بك للاكتئاب مبرمجة مصرياً على حاسوب بالمقارنة مع نظيرتها النسخة التقليدية المألوفة، تكونت عينة الدراسة من ٢٦٨ مريضاً بالاكتئاب الذين تم إختيارهم على مدى عام كامل بحيث يكون نصفهم من الذكور ونصفهم من الإناث وأستوفت العينة مواصفات تشخيص الإكتئاب الواردة في تصنيف الامراض الدولي العاشر وبلغت درجاتهم على مقياس هاملتون للاكتئاب أكثر من ١٨. وذلك خلال مقابلة إكلينيكية شبه مقننة. وقد قسمت العينة لنصفين بالتبادل بحيث أختبر نصفهم باختبار بك التقليدي أى بالورقة والقلم. واختبر نصفهم الآخر بالنسخة المبرمجة على حاسوب والمعدة بواسطة الباحث وكانت المجموعتان متشابهتين من حيث الجنس، السن، مستويات التعليم، الحالة الزوجية، المهنة، ودرجة شدة الاكتئاب، وقيست درجة الثبات بواسطة طريقة التقسيم النصفى، تم حساب معامل الارتباط بين نصفى الاختبار لكل مجموعة على حدة، وقد اشارت النتائج أن درجة ثبات النسخة المبرمجة على حاسوب اعلى من نظيرتها التقليدية، وكانت درجة ثبات النسخة المبرمجة على حاسوب اعلى ما يكون لدى الاناث ولدى الاصغر فى السن والجامعيين وغير المتزوجين والذين بدون عمل.