

Case Report

Further validation of Zagazig depression scale shortened form (ZDS-SF) and depression diagnosis in a United Kingdom (UK) student population

Ahmed. K. Ibrahim^{1, 2*}, Shona. J. Kelly³ and Cris Glazebrook⁴

¹Public Health Division, Faculty of Medicine, Assiut University, Assiut, Egypt.

²Division of Epidemiology, Community Health Sciences School, D Floor, West Block Queens Medical Centre, University of Nottingham, Nottingham, United Kingdom (UK).

³Centre for Intergenerational Health Research, University of South Australia, Division of Health Sciences, Social Epidemiology Unit, City East Campus, Adelaide, Australia.

⁴Division of Psychiatry, Institute of Mental Health, Jubilee Campus, Nottingham, United Kingdom (UK).

Received 09 July, 2012; Accepted 09 November, 2012

The Zagazig Depression Scale has been validated in Egyptian populations and the shortened form (ZDS-SF) found high rates of depression in Egyptian students. Preliminary research has supported the validity and reliability of the measure in a UK student but further work is needed. The study aimed to determine the criterion validity of the ZDS-SF against a clinical interview and its test-retest reliability in a UK student sample. Participants (n=20) completed online measures of the Patient Health Questionnaire and the ZDS-SF at time 1. At time 2 (median follow-up 15 days) they were interviewed using the Clinical Assessment in Neuropsychiatry (SCAN) to establish a clinical diagnosis of depression, followed by re-administration of the online measures. There was excellent (95%) agreement for diagnosis of depression between time 2 ZDS-SF and SCAN (Kappa = 0.89). The sensitivity of the ZDS-SF was 100% and the specificity 93.3%, giving an overall positive predictive value of 83.3%. The ZDS-SF symptom score also had good response stability over a two week interval (ICC = 0.66). ZDS-SF scale is a valid measure of depression for use in a UK university student cohort with good psychometric properties and can be used for cross-cultural comparison studies.

Key words: Depression, university student, Zagazig depression scale.

INTRODUCTION

Depression is recognized as the most common psychiatric disorder affecting adolescents and young adults (Birmaher et al., 2004). It has a multi-factorial

origin; hence many theories have been postulated to explain it (biological, genetic, environmental theories) (Hankin, 2006). Also, it is multi-faceted and can be presented by a mixture of psychiatric and/or physical symptoms (Weinberger et al., 2009). It is not an easy task to diagnose depression in adolescents and young adults as they report certain symptoms more frequently than adults (Gladstone et al., 2011). Although frequently a mild affective disorder depression can have serious complications in young adults and suicide is the leading cause of death in this age group (Mirjami et al., 2011). There is evidence that university students have high rates of depressive symptoms and early identification is important since depression is a preventable and treatable disease especially in children and young adults (Ibrahim et al., 2012; NICE, 2009).

*Corresponding author. E-mail: ahmed.khair@yahoo.com.

Abbreviations: ZDS-SF; Shortened form Zagazig depression scale, PHQ-9; patient health questionnaire-9, SCAN; schedule for clinical assessment of neuropsychiatry, DSM-IV; diagnostic and statistical manual of mental disorders, 4th edition, WHO; World Health Organization, ICD-10; International Classification of Disease, 10th revision, BDI; Beck depression inventory, SES; socio-economic status, GP; general practitioner, NHS; National Health Service, ICC; intra-class correlation, HDRS; Hamilton depression rating scale.

The Arabic version of the Zagazig depression scale (ZDS) has been previously used to screen for depressive symptoms in the Egyptian general population (El-Sayeh, 1991; Fawzi et al., 1982; Fawzy, 1982; Shaheen and Fawzi, 1985) as well as in Egyptian university student cohort (Ibrahim et al., 2011a). It has been translated to be used as a tool for cross-cultural comparison between Egyptian and UK students and preliminary research has supported its validity and reliability in this population (Ibrahim et al., 2010).

Few studies have explored test-retest reliability in depression scales in university students particularly. The beck depression inventory (BDI) multiple test-retest reliability was investigated in a university student sample over a two month period, concluding that BDI has a strong correlation over time ($r = 0.9$), however they did not look at the agreement with other scales (Ahava et al., 1998). This was further supported in another study, where the test-retest reliability was 0.96 over one month interval (Sprinkle et al., 2002). Test-retest reliability is considered good if the agreement between responses on separate administrations is high. Measures of stable constructs over time are expected to have high test-retest reliability. In contrast, since overall mood is anticipated to change over time, a mood scale with very high test-retest reliability may be less sensitive to mood changes (Valentin et al., 2002).

The interval between tests is therefore a key point to consider before interpreting the results of any test-retest analysis. Most depression scales ask participants about depression symptoms over the past two-four weeks. Administration of the depression scale after a longer interval might yield moderate test-retest reliability; because it is more likely that depression symptoms can change over time. If administered two to four weeks apart, test-retest reliability should be moderate to high (William Li et al., 2010).

It was proposed that a standardized and validated tool for depression screening is necessary as it enables comparisons of findings both nationally and internationally and enhances the reputation of the measure (Laake et al., 2007). However, there is no universal agreement on how to adapt an instrument for use in another cultural setting (Gjersing et al., 2010). Self-reported depression scales are cost-effective in time and resources, especially in case of mass screening surveys (Jenkins and Dillman, 1995). Clinical interview is considered the gold standard for diagnosis of depression as it is more objective than subjective and it is carried out by well-trained personnel who can rate the symptoms accurately (Wing et al., 1990).

Aim of this research

To determine the criterion validity of the ZDS-SF against

a clinical interview and the test-retest reliability of the ZDS-SF in a UK undergraduate student sample.

METHODOLOGY

Participants

Participants were university students who had taken part in an online survey of the socio-economic determinants of depression (Ibrahim et al., 2011b). Inclusion criteria were undergraduate students in completion of an online assessment of ZDS-SF and PHQ and attend either the University of Nottingham or Nottingham Trent University. Participants were excluded if they were EU or international students and if they did not provide contact details. Our minimum target sample size was 20 students which was sufficient to detect a Kappa of 0.7 with 90% power based on an estimate of 20% positive ratings (Sim and Wright, 2005).

Of the 923 students participating in the online survey, 564 met the inclusion criteria and were invited to take part in further study. A total of 184 students (32.6%) supplied contact details of which 96 met the inclusion criteria and were invited to participate in the validation study. Of the 34 who agreed to be interviewed 14 either failed to make an appointment or failed to attend for interview within the study period, giving a final sample of 20.

Study design

The validation study consisted of a cross-sectional standardized face to face psychiatric interview with repeated measures for ZDS-SF scores.

Measures

Zagazig depression scale shortened form (ZDS-SF)

The ZDS was derived from the Hamilton depression scale (HDRS) (Hamilton, 1967). The original Arabic version contained 52 items representing 17 domains ((Fawzi et al., 1982). In the current study the translated, modified 43-item version of the ZDS was used (Ibrahim et al., 2010, 2011a), which consists of 11 domains (depressed feelings, suicide ideation, guilt feelings, anxiety, insomnia, agitation/hypochondriasis, sleep maintenance, diminished ability to think, concentrate or slowness, lack of energy and motivation, weight loss, sexual symptoms). Participants rate symptoms as present (0) or absent (1) during the last two weeks giving a maximum score of 43. The reliability and validity of the ZDS-SF was tested in two studies (Ibrahim et al., 2010, 2011a). The first was a pilot study that used a sample of

275 UK University students to improve the questionnaire wording and layout. This pilot study (estimated 30 to 40% response rate) found a strong ($r = 0.8$) and statistically significant correlation between the Patient Health Questionnaire-9 (PHQ-9) (Spitzer et al., 1999) which has been well validated for use as a depression screening tool in UK adult population (Lowe et al., 2004; Spitzer et al., 2004). Internal consistency for the total ZDS-SF was excellent (Cronbach's $\alpha = 0.90$) (Ibrahim et al., 2010). The second was a reanalysis of a representative sample of Egyptian university students, revealed that the internal consistency of the revised ZDS-SF was excellent (Cronbach's $\alpha = 0.91$) as was the split-half correlation coefficient ($r = 0.81$, $p < 0.001$) (Ibrahim et al., 2011a). We used a cut-off of 10 proposed by the ZDS developers (Fawzi et al., 1982) to categorize participants as depressed or non-depressed. ZDS-SF scores were also used as a continuous variable in some analyses.

Patient health questionnaire-9 (PHQ-9)

PHQ-9 is the depression module of the PHQ (Spitzer et al., 1999). It is composed of 9 items each representing one of the 9 DSM-IV criteria for depression. It uses a 4-point scale; "not at all", "several days", "more than half the days" or "nearly every day". The maximum possible score for the PHQ-9 is 27, with a cutoff of 5 to indicate the presence of at least mild depression (Spitzer et al., 1999). The validity, feasibility, and ability to detect changes in depressive symptoms have been reported in several studies (Kroenke et al., 2001; Liu et al., 2011; Lowe et al., 2004; Spitzer et al., 1999; Wulsin et al., 2002). Additionally, the PHQ-9 is increasingly being used in research, and has demonstrated superior criterion validity with respect to the diagnosis of depression compared with other established screening instruments for depression (Kroenke et al., 2004; Lowe et al., 2004).

Schedule for clinical assessment in neuropsychiatry (SCAN)

Schedule for clinical assessment in neuropsychiatry (SCAN) is a set of instruments and manuals designed to assess, measure and classify psychopathology and behavior associated with major psychiatric disorders in adults. Developed under the aegis of the World Health Organization (WHO), it has a bottom-up approach where no diagnosis-driven frames are used to group the symptoms but rather each symptom is assessed in its own right. It has a proven stability and robustness to differentially assess psychotic and neurotic states (Wing et al., 1998). The validity, reliability and the psychometric properties of the SCAN to detect changes in a wide variety of neuropsychiatric disorders have been supported in several studies (Forsell, 2005;

Krisanaprakornkit et al., 2006, 2007; Piyavhatkul et al., 2008; Schutzwahl et al., 2007). To assess the subjects the interviewer conducts a semi-structured, standardized clinical interview. The order in which the sections are completed depends on the most important symptoms of the respondent. This lack of a fixed order makes it very flexible and versatile. Rating is done on the basis of matching the answers of the respondent against the definitions of the symptoms in the SCAN glossary (Wing et al., 1998). In the current study we used the depression section (6th section) in the manual. After completing the interview the data were entered into the laptop version of SCAN "Ishell". Subsequently the data were fed into algorithms for ICD-10 and DSM-IV diagnoses. These algorithms produce a diagnostic classification for depression and a list of symptoms. The severity of the condition is classified as mild, moderate or severe.

Procedures

At time 1 participants completed online versions of the ZDS-SF and PHQ sequentially and, together with demographic details. Participants meeting the study inclusion criteria and agreeing to follow-up clinical interview were offered a choice of interview dates (median follow-up 15 days). The diagnostic interviews (SCAN) were administered at time 2 in a quiet studio. Interviews ranged between 40 to 80 min in duration, with an average of 60 min. All diagnostic interviews were administered by the same trained and reliable clinician (AKI) who was blind to participants' time 1 ZDS scores. Immediately after the SCAN interview, participants were asked to complete the online version of ZDS-SF followed by the online PHQ.

Statistical analysis

A Kappa analysis (Cohen, 1960) was conducted to explore the degree of agreement between the ZDS-SF and PHQ and SCAN (concurrent validity). According to Fleiss a kappa over 0.75 is considered as excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor (Fleiss, 1981). Sensitivity and specificity and other validity measures were also calculated. All analyses were carried out using STATA version 10.1 software (STATA, 2008).

Ethical considerations

This study received approval from the Medical School Ethics Committee of Nottingham University Ref. No. N/9/2008 as a compensation each student participated in was offered a £10 gift voucher for a local department store. A signed written consent was obtained before starting the interview. If the student was

Table 1. Description of the interviewed students.

		N = 20 (%)
Age group	20y or less	10 (50)
	More than 20y	10 (50)
Sex	Male	10 (50)
	Female	10 (50)
Faculty	Arts	1 (5)
	Social Sciences	3 (15)
	Science	7 (35)
	Medicine	9 (45)
Year of study	1st	11 (55)
	2nd	4 (20)
	3rd	3 (15)
	4th or more	2 (10)
Father's occupation	Never worked/unemployed	3 (15)
	Intermediate occupations	8 (40)
	Managerial/professional occupations	9 (45)
Mother's occupation	Never worked/unemployed	5 (25)
	Intermediate occupations	9 (45)
	Managerial/professional occupations	6 (30)
Father's education	No higher education	7 (35)
	Higher education	13 (65)
Mother's education	No Higher education	10 (50)
	Higher education	10 (50)
FAS	Low	2 (10)
	Medium	6 (30)
	High	12 (60)
SCAN diagnosis	Mild depressive disorders	4 (20)
	Moderate depressive disorders	1 (5)
	Alcohol dependence	2 (10)
	Anxiety	1 (5)
	Obsessive compulsive disorders	1 (5)

diagnosed by SCAN as depressed, an e-mail was sent directing the participant to contact his or her GP, NHS direct, the University Counseling Services or the researcher to make the necessary arrangements.

RESULTS

Description of the interviewed sample

Detailed socio-demographic characteristics of the sample are shown in Table 1. Males and females were equally represented in the sample and there was a reasonable range in terms of socio-economic background. The age

of the students sampled ranged from 18 to 34 with a mean (SD) of 20.7 years (6.9). There was an equal distribution in the mother's educational levels, but fathers were more likely to have higher education. Additionally, parental occupational distribution was more or less equal. As expected the majority of students (60%) were in the high affluent group as measured by the family affluence scale (Boyce et al., 2006) (Table 1).

SCAN diagnoses of the 20 interviewees

The prevalence of psychiatric disorders in the current sample at time 2 as ascertained by the clinical interview

Table 2. SCAN Diagnoses of the 20 interviewees and socio-demographic variables, ZDS and PHQ-9 scores.

SN	SCAN diagnosis	Age	Sex	FAS	ZDS score	PHQ-9 score
1	Mild Depressive Episode	19	Female	High	12	4
2	NO	19	Female	High	7	3
3	NO	21	Female	Medium	4	4
4	NO	19	Male	High	6	2
5	NO	19	Female	Low	8	3
6	Obsessive Compulsive Disorders	34	Female	Medium	4	0
7	NO	19	Male	High	9	2
8	NO	19	Male	Medium	6	3
9	Mild depressive episode and alcohol dependence	18	Male	High	18	7
10	NO	18	Male	High	0	3
11	NO	26	Female	Low	9	2
12	NO	23	Male	High	4	0
13	Anxiety disorders	26	Female	Medium	17	14
14	NO	21	Male	High	2	2
15	Alcohol dependence	47	Male	Medium	8	1
16	NO	20	Female	High	0	4
17	NO	19	Male	High	5	1
18	Moderate depressive episode	50	Male	High	28	18
19	Mild depressive episode	30	Female	Medium	17	4
20	Mild depressive episode	23	Female	High	19	10

was 40% (95% CI; 38.9-40.4) (n=8). The most common SCAN diagnosis in the interviewed students was depressive disorder (n=5), alcohol dependence (n=2), anxiety and obsessive compulsive disorders (n=1). One student had minor depressive disorder and alcohol abuse Table 2 shows the detailed description of the 20 interviewed students regarding their SCAN diagnosis, socio-demographic characters and their ZDS and PHQ-9 scores.

ZDS-SF reliability and validity

Depression scores

The mean ZDS-SF score for the 20 participants at time 1 was 11.80, (SD = 4.3). The 5% trimmed mean was 11.17, only 0.63 below the sample mean. Scores were essentially normally distributed (Skewness (SE) = 0.82 (0.36), Kurtosis (SE) = 0.94 (0.57)); with a median score of 8.5. Five participants (25%) scored above the cut-off for depression on the ZDS-SF at time 1, all of whom were later classified as cases of depression by the SCAN clinical interview at two-week follow-up (100% agreement, Kappa = 1). At time 2 the mean ZDS-SF score was 14.62, (SD = 4.5). The 5% trimmed mean was 14.36, only 0.26 below the sample mean. Scores were also normally distributed (Skewness (SE) = 0.69 (0.23), Kurtosis (SE) = 0.72 (0.47)); with a median score of 15. Of the 20 participants, six (30%) scored above the cut-off

for depression on the ZDS-SF at time 2. A quarter of the total sample (n=5) was diagnosed as depressed using the SCAN diagnostic classification. Four were classified as having mild depression and one was classified as moderate depression.

Concurrent validity

There was a strong positive correlations between ZDS-SF scores at time 2 and SCAN diagnostic symptom scores (Spearman's Rho = 0.88, $p < 0.001$). There was agreement between ZDS-SF and SCAN on whether the participant was depressed or not depressed for 95% of cases. In only one case (5%) participants were classified as depressed by the ZDS-SF (symptom score = 18) and not by the SCAN. The resulting Kappa score (0.89) indicates excellent agreement ($p < 0.001$) (Table 3). The sensitivity of the ZDS-SF was 100% and the specificity was 93.3% giving an overall positive predictive value of 83.3%. For the PHQ there was a moderate positive correlation between PHQ scores at time 2 and SCAN diagnostic results (Spearman's rho = 0.54, $p < 0.001$) with lower sensitivity and specificity (Table 4).

Reliability

The time interval between the test and the retest ranged from 10 to 30 days with a mean of 17 ± 5.5 days (median

Table 3. Agreement between ZDSSF, PHQ, and SCAN.

		ZDS-SF*		PHQ*		Total (%)
		Not depressed (%)	Depressed (%)	Not depressed (%)	Depressed (%)	
SCAN	Not depressed	14 (70)	1 (5)	14 (70)	1 (5)	15 (75)
	Depressed	0 (0)	5 (20)	2 (10)	3 (15)	5 (25)
Total		14 (70)	6 (30)	16 (80)	4 (20)	20 (100)

*ZDS-SF cutoff ≥ 10 , PHQ cutoff ≥ 5 .

Table 4. Other validity measures for ZDS-SF vs. PHQ and SCAN.

	Vs. PHQ	Vs. SCAN
Sensitivity	100%	100%
Specificity	87.5%	93.3%
Positive predictive value (PPV)	66.6%	83.3%
Negative predictive value (NPV)	100%	100%
False positive rate (FPR)	12.5%	6.7%
False negative rate (FNR)	0%	0%
Positive likelihood ratio (LR+)	8	14.9
Negative likelihood ratio (LR-)	0	0
Accuracy	90%	95%
Power	1	1
False discovery rate (FDR)	33.3%	16.7%
Kappa	0.74	0.89

15 days, IQR range 12 to 27 days). The 43-item version ZDS had good response stability over about two weeks interval ($n=20$, Spearman's correlation = 0.72; intra-class correlation coefficient (ICC) = 0.66 (95% CI=0.58-0.69). Depression scores increased over time from a mean of 11.8 (± 4.3) at time one to 14.6 (± 4.5) at time two, ($t = 3.7$, $df = 38$, $p < 0.001$). The internal consistency of the ZDS-SF was very good for both test and retest (Cronbach's alphas 0.88 and 0.89 respectively).

DISCUSSION

The current study found that ZDS-SF had good response stability over two week interval (ICC = 0.66). There was an excellent agreement between ZDS-SF and SCAN on whether the participant was depressed or not depressed for 95% of cases (Kappa = 0.89). The ZDS-SF showed a good sensitivity and specificity (100 and 93%), with a positive predictive value of 83%.

The completion of self-rated depression scales needs a good level of education, co-operation of the respondents and may be more likely to be affected by cultural bias or illness presentation. But, they are cost-effective in time and resources especially in case of mass screening

surveys (Jenkins and Dillman, 1995). On the other hand, clinical interview is considered the gold standard for depression diagnosis as it is more objective than subjective and it is carried out by well-trained physicians who can rate the symptoms accurately. However it is time consuming and costly (Wing et al., 1990). The aforementioned highlights the importance of validating self-rating scales using a gold-standard clinical interview (for example, SCAN). It has been proposed that ZDS-SF is a relatively constant over two week period. Thus, test robustness can be measured over a relatively short time (Ibrahim et al., 2011a).

As predicted the result of the test-retest reliability was moderate because ZDS-SF scale was primarily intended to measure current mental state (over the past two weeks). Symptoms are expected to change over time but these changes are more likely to be cyclic in some individuals. Moreover, test-retest analysis has some methodological bias that is, time interval was variable and totally dependent on the participant's availability. It has been assumed that shorter intervals will produce higher correlations than longer intervals. If there was more information about what happened to participants during the time interval then the test results should be better differentiated. In particular, adverse life events are

expected to introduce variability in the mood profile of participants and thus modify the test outcome (Valentin et al., 2002).

The concurrent validity of the scale was tested against the PHQ and the SCAN. ZDS-SF scores were strongly correlated with both PHQ scores and SCAN diagnostic results ($r_s = 0.76$, $p < 0.001$ and $r_s = 0.88$, $p < 0.001$ respectively). Also, using the Kappa analysis, ZDS-SF demonstrated a very good agreement with the SCAN (0.89, $p < 0.001$) which is a 'gold standard'. This was stronger than the agreement with PHQ (0.74, $p < 0.001$). Additionally, ZDS-SF as a screening tool for depression was tested against the SCAN and showed 100% sensitivity, nearly 93% specificity, and an overall accuracy of 95%. The SCAN interview probes persistent symptoms experienced over the past month which explains the perfect agreement between the ZDS-SF classification of depression at time 1 and the clinical diagnosis of depression two weeks later, providing further evidence of the ZDS-SF validity.

This was in accordance with other studies; in general population studies where the validity of the ZDS-SF has been examined against the HDRS which found that it was a valid measure and could be used as a useful screening measure for both clinical and research studies (Fawzi et al., 1982; Fawzy, 1982; Shaheen and Fawzi, 1985). Moreover, in two student sample studies the ZDS-SF was tested against PHQ revealing good concurrent validity against this well-established screening tool (Ibrahim et al., 2010, 2011a). These results demonstrated that ZDS-SF is good screening tool for depression in university students. Although longer than the PHQ we believe that the more comprehensive and the wider range of symptom domains assessed makes it particularly useful for screening for emotional difficulties in well-educated populations across both developed and developing countries.

The strength of the present study was the use of a well validated self-administered scale (PHQ) in addition to a gold-standard depression semi-structured interview (SCAN), administered blind to ZDS-SF scores, in order to validate the ZDS-SF. However the study encountered the following limitations; the small number of participants ($n=20$), and the use of a convenience approach for student recruitment.

In conclusion, ZDS-SF scale is a valid measure of depression with very good psychometric properties in a university student cohort. Further research is planned to establish the psychometric properties of the ZDS-SF in a Chinese student population.

ACKNOWLEDGEMENTS

The authors are very grateful for the Ministry of Higher Education, Egyptian Government specially Assiut

University for sponsoring my whole studies. It is a pleasure to express my deepest gratitude and grateful appreciation to the University of Nottingham for supporting this study. Last but not least, my special thanks and gratitude to the students who took part in this study. It would not have been possible without their help.

REFERENCES

- Ahava G, Iannone C, Grebstein L, Schirling J (1998). Is the Beck Depression Inventory reliable over time? An evaluation of multiple test-retest reliability in a nonclinical college student sample. *J. Pers. Assess.*, 70(2): 222-231.
- Birmaher B, Williamson D, Dahl R, Axelson D, Kaufman J, Dorn L (2004). Clinical Presentation and Course of Depression in Youth: Does Onset in Childhood Differ From Onset in Adolescence? *J. Am. Acad. Child Adolesc. Psychiatry*, 43(1): 63-70.
- Boyce W, Torsheim T, Currie C, Zambon A (2006). The family affluence scale as a measure of national wealth: validation of an adolescent self-report measure. *Soc. Indic. Res.*, 78: 473-487.
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Measure.*, 20(1): 37-46.
- El-Sayeh A (1991). Epidemiology and symptomatology of depression in an upper Egyptian community., Assiut University, Assiut, Egypt.
- Fawzi M, El-Maghraby Z, El-Amin H, Sahloul M (1982). The Zagazig Depression Scale Manual. Cairo: El-Nahda El-Massriya (Arabic).
- Fawzy M (1982). Depression among the elderly patients of the outpatient psychiatric clinics (Arabic). The international conference of the geriatric mental health. Cairo, Egypt.
- Fleiss J (1981). Statistical methods for rates and proportions (2nd ed). New York: John Wiley.
- Forsell Y (2005). The Major Depression Inventory versus Schedules for Clinical Assessment in Neuropsychiatry in a population sample. *Soc. Psychiat. Epidemiol.*, 40(3): 209-213.
- Gjersing L, Caplehorn J, Clausen T (2010). Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med. Res. Methodol.*, 10(13): 1-10.
- Gladstone T, Beardslee W, O'Connor E (2011). The Prevention of Adolescent Depression. *Psychiatric Clinics of North America*, 34(1), 35-52.
- Hamilton M (1967). Development of a rating scale for primary depressive illness. *Br. J. Soc. Clin. Psych.*, 6: 278-296.
- Hankin B (2006). Adolescent depression: description, causes, and interventions. *Epilepsy Behav.*, 8(1): 102-114.
- Ibrahim A, Kelly S, Glazebrook C (2011b). Socioeconomic Status and the Risk of Depression

- among UK Higher Education Students (pp. 1-30). Nottingham: The University of Nottingham.
- Ibrahim A, Kelly S, Glazebrook C (2011a). Reliability and validity of an Arabic version of Hamilton Depression Scale in an Egyptian University student sample. *Comp. Psych.*, 53(5): 638-647.
- Ibrahim A, Kelly S, Glazebrook C (2012). Analysis of an Egyptian study on the socioeconomic distribution of depressive symptoms among undergraduates. *Soc. Psychiatry Psychiatr. Epidemiol.*, 47(6): 927-937.
- Ibrahim A, Kelly S, Challenor C, Glazebrook C (2010). Establishing the reliability and validity of the Zagazig Depression Scale in a UK student population: an online pilot study. *BMC Psychiatry*, 10(107), doi:10.1186/1471-1244X-1110-1107.
- Jenkins C, Dillman D (1995). *Towards a theory of self-administered questionnaire design*. New York: Wiley-Interscience.
- Krisanaprakornkit T, Paholpak S, Piyavhatkul N (2006). The validity and reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Mood Disorders Section. *J. Med. Assoc. Thai.*, 89(2): 205-211.
- Krisanaprakornkit T, Rangseekajee P, Paholpak S, Khiewyoo J (2007). The Validity and Reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Anxiety Disorders Section. *J. Med. Assoc. Thai.*, 90(2): 341-347.
- Kroenke K, Robert L, Spitzer M (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Int. Med.*, 16(9): 606-613.
- Laake P, Olsen B, Benestad H (2007). *Research methodology in the medical and biological sciences* Amsterdam; (1st edition). New York: Elsevier, Academic Press.
- Liu S, Yeh Z, Huang H, Sun F, Tjung J, Hwang L (2011). Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatry*, 52(1): 96-101.
- Lowe B, Kroenke K, Herzog W, Grafe K (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J. Affect. Disord.*, 81(1): 61-66.
- Lowe B, Spitzer R, Grafe K (2004). Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J. Affect. Disord.*, 78(2): 131-140.
- Mirjami P, Linnea K, Mauri M (2011). Adolescent Suicide: Epidemiology, Psychological Theories, Risk Factors, and Prevention. *Curr. Pediatr. Rev.*, 7(1): 52-67.
- NICE (2009). *Depression: the treatment and management of depression in adults*. NICE Clinical Guideline 90. London: National Institute for Health and Clinical Excellence.
- Piyavhatkul N, Krisanaprakornkit T, Paholpak S, Khiewyoo J (2008). Validity and reliability of WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN)-Thai version: Cognitive Impairment or Decline Section. *J. Med. Assoc. Thai.*, 91(7): 1129-1136.
- Schutzwohl 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. *European Psychiatr.*, 22(4): 229-238.
- Shaheen O, Fawzi M (1985). Further assessment of a new self-rating scale for depression: Zagazig Depression Scale (Arabic). *Egypt. J. Mental Health*, 26(1): 73-91.
- Sim J, Wright C (2005). *The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements*. *Phy. Therapy*, 85(3): 257-268.
- Spitzer R, Kroenke K, Williams J (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ Primary Care Study. *JAMA*, 282(18): 1737-1744.
- Sprinkle S, Lurie D, Insko S, Atkinson G, Jones G, Logan A (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *J. Couns. Psychol.*, 49(3): 381-385.
- STATA (2008). *Data analysis and Statistical Software (Ver.10.1)*: Copyright © StataCorp LP, 2008-2009.
- Valentin R, Theo G, Burkhardt S (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat. Med.*, 21(22): 3431-3446.
- Weinberger M, Mateo C, Sirey A (2009). Perceived barriers to mental health care and goal setting among depressed, community-dwelling Patient Prefer Adherence, 3: 145-149.
- William LH, Chung O, Yan K (2010). Center for Epidemiologic Studies Depression Scale for Children: psychometric testing of the Chinese version. *JAN*, (66): 11.
- Wing J, Babor T, Brugha T, Burke J, Cooper J, Giel R (1990). Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Arch. Gen. Psych.*, 47(6): 589-593.
- Wing J, Sartorius N, Ustun T (1998). *Diagnosis and clinical measurement in psychiatry, a reference manual for SCAN/PSE-10*. In: Cambridge University Press, ISBN: 0 521 43477 7
- Wulsin L, Somoza E, Heck J (2002). The Feasibility of Using the Spanish PHQ-9 to Screen for Depression in Primary Care in Honduras Prim Care Companion. *J. Clin. Psych.*, 4(5): 191-195.