ORIGINAL ARTICLE

The sensitivity, specificity and reliability of the Malay version 30-item General Health Questionnaire (GHQ-30) in detecting distressed medical students.

Muhamad Saiful Bahri Yusoff

Medical Education Department, School of Medical Sciences, Universiti Sains Malaysia

Abstract

Objective: To determine the sensitivity, specificity and internal consistency of the Malay version GHQ-30 among medical student population. This study also determined the level of agreement between GHQ-30 and M-BDI.

Methods: The Malay version GHQ-30 and Malay version Beck Depression Inventory (M-BDI) were administered to 190 medical students. ROC curve analysis was applied to determine the sensitivity and specificity of the GHQ-30 by testing against the M-BDI diagnoses. Reliability and Kappa analysis were applied to test internal consistency of the GHQ and to determine the level of agreement between GHQ-30 and M-BDI respectively.

Results: 141 (74.2%) medical students participated in this study. The GHQ-30 sensitivity and specificity at cut-off point of 5/6 was 87.5% and 80.6% respectively with positive predictive value (PPV) of 70% as well as area under ROC curve was 0.84. The Cronbach's alpha value of the GHQ-30 was 0.93. The Kappa coefficient was 0.64 (p<0.001).

Conclusion: This study showed the Malay version GHQ-30 is a valid and reliable screening tool in detecting distressed medical students. The GHQ-30 score equal to or more than 6 was considered as significant distress. The GHQ-30 showed a good level of agreement with M-BDI in detecting distressed medical students.

Keywords: reliability, validity, medical students, General Health Questionnaire (GHQ), Beck's Depression Inventory (BDI)

Introduction

The General Health Questionnaire (GHQ) was widely used internationally and locally (1-7) to measure mental health status especially in detection of emotional disorders such as distress. Since Goldberg introduced the GHQ in 1978, it has been translated into 38 different languages, testimony to the validity and reliability of the questionnaire (4). Reliability coefficients of the questionnaire have ranged from 0.78 to 0.95 in various studies (4). It has four versions based on the number of items; GHQ-60, GHQ-30, GHQ-28 and the shortest version GHQ-12. Each item is accompanied by four responses, typically being 'not at all', 'no more than usual', 'rather more than usual' and 'much more than usual'. There are two recommended methods for scoring the GHQ. The first scoring method ranged from 0 to 3 respectively. The second scoring method was binary scoring method

(with the two least symptomatic answers scoring 0 and the two most symptomatic answers scoring 1 – i.e. 0-0-1-1). The total possible score on GHQ-28 ranges from 0 to 84 and allows for means and distributions to be calculated, both for the global total, as well as for the four sub-scales (somatic symptoms, anxiety/insomnia, social dysfunction and severe depression). Using the alternative binary scoring methods the 28- and 30-items versions classify any score exceeding the threshold of 4 as achieving 'caseness'. The caseness threshold is 3 for the 12-item version. The shortened version work was found to be as reliable as the long version in detecting distress (3). The GHQ-30 and GHQ-12 was commonly used as its validity is well-established internationally (8, 9) and locally (1, 2). It was also used because of its popular use in student sample (9) and young populations in the community (8). Furthermore it is simple, easy to understand, short and straightforward to complete. Most of the validation studies have been done in the western countries (7) and a few studies were done in Malaysia (1, 2). However, generally it is hardly found such study in medical student population.

Numerous studies have revealed that persistence stressful condition was associated with mental and physical health problems in medical students at various stages of their training (10-15). Studies have reported an association of excessive stress level with lowered medical students' self-esteem (16, 17), anxiety and depression (18; 19), difficulties in solving interpersonal conflicts (20), sleeping disorders (21), increased alcohol and drug consumption (22-23), cynicism, decreased attention, reduced concentration and academic dishonesty (24). It is also associated with inhibition of students' academic achievement and personal growth development (17). Excessive stress was also linked with medical student suicide (25). As a result, medical students may feel inadequate and unsatisfied with their career as a medical practitioner in the future (26). It is noteworthy that many researchers have stated the importance of early detection as well as effective intervention programme,

which can prevent possible future illnesses among medical students (11, 12, 15).

The purpose of this study is to determine the sensitivity, specificity and reliability of the Malay version GHQ-30 among medical student population particularly in Malaysia. It is also looking at the appropriate GHQ-30 score in detecting distress in the population as well as the level of agreement between GHQ-30 and M-BDI.

Methodology

The Malay version GHQ-30

The Malay version of GHQ-30 was used in this study. In Malaysia, the instrument has been validated in the local population using English and Malay versions (1, 2). It is a self-reporting questionnaire and it consists of broad symptoms of psychiatric disorders in the general population. The items of the Malay version GHQ-30 were rated under 4 categories of responses; *tiada langsung* (not at all) , *tidak lebih dari biasa* (no more than usual), *lebih dari biasa* (more than usual), *sangat lebih dari biasa* (much more than usual) for statement 2, 3, 11, 14, 15, 16, 18, 19, 21, 22, 23, 24, 28, 29 and 30 as shown in table 1, whereas, for the rest of statements the responses were *lebih dari biasa* (more than usual), *tidak lebih dari biasa* (no more than usual), *kurang dari biasa* (less than usual) and *sangat kurang dari biasa* (much less than usual). The recommended scoring method was a binary scoring method where the two least symptomatic answers score 0 and the two most symptomatic answers score 1 – i.e. 0-0-1-1 (27). The minimum GHQ-30 total score was 0 and the maximum GHQ-30 total score was 30. Similar GHQ scoring was used in this study.

The Malay version Beck's Depression Inventory (M-BDI)

The M-BDI is a translated and validated self-reporting questionnaire and it consists of broad symptoms of psychiatric disorders in the general population (28). The original BDI

was first introduced in 1961 by Aaron T Beck and it is one of the most widely used measures of depressive symptoms both in adolescents and in adults (28). Although the instrument was initially developed to measure intensity or severity of the depressive symptomatology in patients with psychiatric disorders (29), it has now been widely used as a goal standard for screening instrument to detect emotional disorder such as distress and depression both in clinical practice and in research projects (30, 31). It comprises of 21 items and was rated using a four-point scale ranging from 0 to 3. The reliability coefficient of the BDI in nonpsychiatric samples ranges from 0.73 and 0.92 (31); The M-BDI reliability coefficient was 0.89 (28). The minimum and maximum scores of BDI were 0 and 63 respectively. Recommended screening threshold score for nonpsychiatric samples was 9/10; it yielded 84.6% of sensitivity and 86.4% of specificity (31). In this study those who score equal to or more than 9 were considered as experiencing significant emotional disturbances (distress) (28, 31). The scores are also classified into mild-to-moderate (9 to 18), moderate-to-severe (19 to 29) and extremely severe depression (more than 29) (30, 31). Similar scoring method was used in this study.

Validation study

The Malay version GHQ-30 used for this study contained 30 items. Population of this study was 1065 medical students of 2008/2009 academic session in the School of Medical Sciences, Universiti Sains Malaysia. Proper instructions were given before the administration of the scale. The subjects were asked to respond to all the statements and no time limit was imposed. During the time of administration the investigator gave proper assistance and directions whenever and wherever necessary.

Sample size

Sample size calculated based on recommended ratio which was 5 subjects per item (32) with 20 percent dropout rate was 190 subjects. Convenient sampling method

was applied; 70 second year medical students and 120 fifth year medical students were asked to participate in this study.

Collection of data

The investigator requested 190 medical students from second year and fifth year to fill in the GHQ-30. Completion of the questionnaire was voluntary and would not affect the students' progress in the course. A face-to-face session was held with the students in a hall. Data was collected by guided self-administered questionnaire. The time taken by the students to fill in the questionnaire was around 15 minutes. The questionnaires were collected on the same day. Verbal consent was taken from the students. The investigator obtained permission and clearance from the School of Medical Sciences, Universiti Sains Malaysia.

Reliability analysis

Reliability analysis was done using SPSS version 12 to determine the internal consistency of the items measured by using Cronbach's alpha coefficient. For an estimation of reliability, statistical reliability of individual item was done. Items with Cronbach's alpha value if item-deleted could determine which statement was highly contributed to the alpha value. If the Cronbach's alpha value for those items-deleted were decreased, it indicated that the items were highly contributed to alpha value. In contrast, if the Cronbach's alpha value for those items-deleted were increased, it indicated that the items poorly contributed to alpha value. An item was considered as having an acceptable correlation with other items if it had corrected-item total correlation more than +0.3 or -0.3; It represented on how well one specific item correlate with the whole items of an instrument (32). The items of the GHQ-30 were considered to represent a measure of high internal consistency if the Cronbach's alpha value was more than 0.7 (33).

Sensitivity and specificity analysis

Distress detection was made based on the M-BDI. The M-BDI score equal to and more than 9 was considered as positive score for significant distress (28). In order to determine the sensitivity and specificity, the GHQ-30 was tested against the distress diagnoses made by the M-BDI. The Receiver Operating Characteristics (ROC) curve analysis was done using SPSS version 12 to determine the sensitivity, specificity, and area under ROC curve. The sensitivity, specificity and area under ROC curve value more than 0.70 was considered as having an acceptable predictive and discriminative value (34). The negative and positive predictive values were calculated manually by Microsoft Excel software.

Degree of agreement between the GHQ-12 and GHQ-30

The Cohen's Kappa (often simply called Kappa) is used as a measure of agreement between two instruments. The degree of agreement between GHQ-30 and M-BDI was considered as having a good level of agreement if Kappa value more than 0.6 (35).

Results

141 (74.2%) medical students participated in this study. 99 (70.2%) were female students. 49 (34.8%) were second year medical students and 92 (65.2%) were fifth year medical students. 86 (61.0%) were Malay, 53 (37.6%) were Chinese, 2 (1.4%) were Indian and 3 (0.3%) were others. 86 (61.0%) were Muslim, 37 (26.2%) were Buddha, 15 (10.6%) were Christian, and 9 (2.1%) were others.

Reliability analysis

Table 1 showed all the items had corrected-item total correlation more than 0.3. Thus all the items were remained in the questionnaire. The Cronbach's alpha value for the Malay version GHQ-30 was 0.93. It suggested that the items of the GHQ-30 were reliable as having high internal consistency.

Sensitivity and specificity analysis

Table 2 and figure 1 showed that GHQ-30 score at 5/6 have the optimum sensitivity and specificity which were 87.5% and 80.6% respectively with positive predictive value of 70% as well as having area under curve more than 0.7. It was an evidence to suggest that the GHQ-30 has an acceptable predictive and discriminative value in detecting distressed medical students.

Degree of agreement between the GHQ-30 and M-BDI

The Kappa value for the GHQ-30 and M-BDI was 0.64 (p < 0.001) and it reflected a good level of agreement (35).

Discussion

Reliability generally is defined as consistency or reproducibility of measurement over time or occasions, whereas validity generally is defined as to what extent the measurement measures what it should measure (33, 36, 37). Sensitivity is defined as the proportion of persons with disease who test positive, whereas, specificity is defined as the proportion of persons without disease who test negative (34, 36, 37). Therefore sensitivity and specificity describe how well the test predicts and discriminates between patients with and without disease. The accuracy of a test depends on how well the test separates the group being tested into those with and without the disease in question. Accuracy is measured by the area under the ROC curve (34). The Cronbach's alpha coefficient was commonly used by researchers in determining the internal consistency of an instrument, whereas ROC curve analysis was used to determine the specificity and sensitivity. The Cohen Kappa coefficient was used to determine level of agreement between the GHQ-30 and M-BDI (35). In this study, the same analysis was applied to determine the sensitivity, specificity and internal consistency of the Malay version GHQ-30 as well as the level of agreement with M-BDI.

Reliability analysis suggested that all the 30 items have corrected-item total correlation value more than 0.3 as shown in table 1; therefore all the items were correlated with each other well and they were remained in the GHQ-30. The GHQ-30 has shown a measure of high internal consistency as having Cronbach's alpha value more than 0.7 as shown in table 1; it reflected the consistency and reproducibility of the GHQ-30 measurements. The findings were evidence to support that the GHQ-30 was a reliable instrument that could be used in the future to detect distressed medical students.

ROC curve analysis has shown that the optimum cut-off point to detect distressed was 5/6 as shown in table 1. The analysis has also shown that at the GHQ-30 score of 5/6 have acceptable predictive and discriminative values as the sensitivity, specificity and area under ROC curve was more than 0.7; which were 0.87, 0.81, and 0.84 respectively. It reflects the ability of GHQ-30 to discriminate between distressed and non-distressed medical students. It is noteworthy to highlight that, present study finding is comparable with the Goldberg et al (3) finding which yielded sensitivity and specificity about 83.7% and 79.0% respectively at cut-off point of 4/5 and with the study by Abdul Hamid and Hatta (1), which yielded the sensitivity and specificity of the GHQ-30 was 96% and 93.3% respectively at cut-off point of 7/8. The findings were evidence to support and suggest that the GHQ-30 was a valid instrument to detect distressed medical students.

This study found that the GHQ-30 and M-BDI had a good level of agreement in detecting distressed medical students as the Kappa value was more than 0.6 (35). This is another evidence to support the validity of GHQ-30; it measured what it should measure.

It is noteworthy that, this study has its limitations which must be taken into consideration in the future studies. The sample size in this study was not representing the actual distribution of the study population in term of gender, ethnic groups, years of study, and religion. The convenient sampling method used in this study may lead to sample bias hence may affect accuracy of the results. Furthermore, the cut-off point of GHQ-30 was determined by M-BDI diagnoses which are not the gold standard diagnoses of psychiatrist, thus the accuracy of the result may be questioned. Considering the limitations, the present study results should be interpreted cautiously. However, this study has provided useful information on this area for future studies.

Conclusion

This study showed the Malay version GHQ-30 is a valid and reliable screening tool in detecting distressed medical students. The optimum cut-off point of the GHQ-30 score to detect distressed medical students was 6 and above. The GHQ-30 showed a good level of agreement with M-BDI in detecting distressed medical students.

Table 1: Reliability analysis on the questions of Malay version GHQ-30.

No.	Question	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	*Adakah anda boleh menumpukan perhatian kepada apa sahaja yang dibuat*? (Been able to concentrate on whatever you are doing?)	0.581	0.931
2	*Adakah anda kekurangan tidur kerana risau*? (Lost much sleep over worry?)	0.522	0.932

3	*Adakah anda mengalami resah dan ganguan tidur pada waktu malam*? (Been having restless, dirturbed night?)	0.432	0.933
4	*Adakah anda dapat memenuhi masa anda dengan sewajarnya*? (Been managing to keep yourself busy and occupied?)	0.441	0.933
5	*Adakah anda keluar dari rumah sekerap biasa*? (Been getting out of the house as much as usual?)	0.394	0.933
6	*Adakah anda dapat mengendalikan urusan seperti kebanyakan orang dalam keadaan yang serupa dengan anda*? (Been managing as well as most people would in your shoe?)	0.428	0.933
7	*Adakah secara keseluruhannya anda merasa yang anda telah melakukan segala-galanya dengan baik*? (Felt on the whole you were doing things well?)	0.632	0.930
8	*Adakah anda berpuas hati dengan cara anda melakukan tugas anda*? (Been satisfied with the way you're carried out your task?)	0.568	0.931
9	*Adakah anda dapat merasa kemesraan dan kasih sayang terhadap mereka yang rapat dengan anda*? (Been able to feel warmth and affection for those near to you?)	0.455	0.932
10	*Adakah anda dapati senang bergaul sengan orang lain*? (Been finding it easy to et on with other people?)	0.457	0.932
11	*Adakah anda meluangkan banyak masa berbual-bual dengan orang lain*? (Spent much time chatting with people?)	0.150	0.936
12	*Adakah anda rasa yang anda memainkan peranan yang berguna dalam banyak perkara*? (Felt that you are playing useful part in things?)	0.568	0.931
13	*Adakah anda merasa mampu membuat keputusan tentang sesuatu*? (Felt capable of making decisions about things?)	0.596	0.931
14	*Adakah anda sentiasa merasa tegang*? (Felt constantly under strain?)	0.665	0.930
15	*Adakah anda rasa yang tidak dapat mengatasi kesukaran/ masaalah anda*? (Felt you couldn't overcome your difficulties?)	0.682	0.930
16	*Adakah anda merasakan hidup ini penuh perjuangan/ pergelutan sepanjang masa*? (Been finding life a struggle all the time?)	0.500	0.932
17	*Adakah anda dapat menikmati kegiatan harian anda*?	0.606	0.931

(Been able to enjoy your normal day-to-day activities?)

18	*Adakah anda terlalu mengambil berat/ mengambil tahu tentang hal-hal biasa?* (Been taking things hard?)	0.337	0.934
19	*Adakah anda mudah takut atau cemas tanpa apa-apa sebab?* (Been getting scared or panicky for no good reason?)	0.589	0.931
20	*Adakah anda dapat mengatasi masaalah-masaalah anda?* (Been able to face up with your problem?)	0.684	0.930
21	*Adakah anda dapati bahawa segala-galanya semakin membebani diri anda?* (Found everything getting on top of you?)	0.704	0.929
22	*Adakah anda merasa tidak gembira dan sedih?* (Been feeling unhappy and depressed?)	0.691	0.930
23	*Adakah anda telah hilang kepercayaan pada diri anda sendiri?* (Been losing confidence in yourself?)	0.653	0.930
24	*Adakah anda memikirkan diri anda seorang yang tidak berguna?* (Been thinking of yourself as a worthless person?)	0.595	0.931
25	*Adakah anda rasa bahawa hidup ini tiada harapan langsung?* (Felt that life is entirely hopeless?)	0.559	0.932
26	*Adakah anda menaruh harapan terhadap masa depan anda?* (Been feeling hopeful about your own future?)	0.440	0.933
27	*Adakah anda rasa cukup gembira dalam segala hal yang difikirkan?* (Been feeling reasonably happy, all things considered?)	0.596	0.931
28	*Adakah anda rasa cemas dan resah di sepanjang masa?* (Been feeling nervous and strung-up all the time?)	0.682	0.930
29	*Adakah anda rasa tiada guna hidup?* (Felt that life isn't worth living?)	0.571	0.932
30	*Adakah kadang-kadang anda tidak dapat berbuat apa-apa kerja kerana terlalu cemas dan gentar?* (Found at times you couldn't do anything because your nerves were too bad?)	0.681	0.930
Total Cronbach's Alpha = 0.934			

Table 2: The area under ROC curve, sensitivity, specificity, positive and negative predictive values of different cut-off points for significant distress (detection based on M-BDI).

*GHQ score	The Area Under ROC curve	Sensitivity %	Specificity %	PPV %	NPV %
0/1	0.67	100	34.4	44.0	100
1/2	0.75	95.8	53.8	51.7	96.2
2/3	0.79	93.8	64.5	57.7	95.2
3/4	0.81	91.7	71.0	62.0	94.3
4/5	0.82	89.6	74.2	64.2	93.2
5/6	0.84	87.5	80.6	70.0	92.6
6/7	0.82	79.2	83.9	71.7	88.6
7/8	0.81	75.0	87.1	75.0	87.1
8/9	0.80	70.8	89.2	77.3	85.6
9/10	0.79	68.8	89.2	76.7	84.7
10/11	0.79	66.7	91.4	80.0	84.2
11/12	0.78	60.4	95.7	87.9	82.4

ROC = Receiver Operating Characteristics, **PPV** = Positive Predictive Value,

NPV = Negative Predictive Value
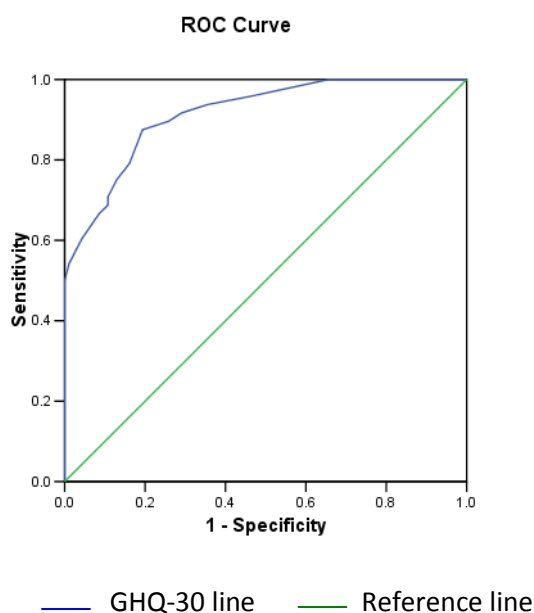
*Total score range from 0 to 30

Figure1: The GHQ-30 ROC area at the cut-off point of 5/6 for significant distress (diagnoses were based on the M-BDI).

References

1. Abdul Hamid AR, Hatta SMA. Validation study of the Malay language translation of General Health Questionnaire (GHQ-30), Malaysian Psychiatry, 2006; **2(4)**: 34-37.

2. Maniam T. Validation of the General Health Questionnaire (GHQ-30) for a Malaysian population, Mal J Psych, 1996; **4(2)**: 25-31.

3. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O *et al*. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med, 1997; **27(1)**: 191-197

4. Jackson C. The General Health Questionnaire, Occupational Medicine, 2007; **57**: 59.

5. Montazeri A, Harrichi AM, Shariati M, Garmaroudi G, Ebadi M, Fateh A. The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version, Health and Quality of Life Outcomes, 2003; **1(66)**.

6. Quek KF, Low WY, Razack AH, Loh CS. Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian Study, Psychiatry and Clinical Neurosciences, 2001; **55**: 509-513.

7. Nagyova I, Krol B, Szilasiova A, E Stewart R, P van Djik J, van den Heuval WJA. General Health Questionnaire-28: psychometric evaluation of the Slovak version , Studio Pysholoiga, 2000; **42 (4)**: 351-361.

8. Banks MH. Validity of the General Health Questionnaire in a young community sample, Pyschological Medicine, 1983; **13**: 349-54.

9. Radanovic Z, Eric LJ. Validity of the General Health Questionnaire in Yugoslav student population, Psychological Medicine, 1983; **13**: 205-7.

10. Zaid ZA, Chan SC, Ho JJ. Emotional disorders among medical students in a Malaysian private medical school. *Singapore Med J,* 2007; **48(10):** 895-899

11. Sherina MS, Lekhraj R, Nadarajan K. Prevalence of emotional disorder among medical students in a Malaysian university, *Asia Pacific Family Medicine*, 2003; **2**: 213-217.

12. Aktekin M, Karaman T, Senol YY, Erdem S, Erengin H, Akaydin M. Anxiety, depression and stressful life events among medical students: a prospective study in Antalya, Turkey. Medical Education, 2001; **35(1)**: 12-17.

13. Guthrie EA, Black D, Shaw CM, Hamilton J, Creed FH, Tomenson B. Embarking upon a medical career: psychological morbidity in first year medical students. *Med Educ*, 1995; **29(5)**: 337-341.

14. Miller PM, Surtees PG. Psychological symptoms and their course in first-year medical students as assessed by the Interval General Health Questionnaire (I-GHQ). *Br J Psychiatry,* 1991; **159**: 199-207.

15. Firth, J. Levels and sources of stress in medical students. *Br Med J (Clin Res Ed),* 1986; **292 (6529):** 1177-1180.

16. Silver HK, Glicken AD. Medical student abuse. Incidence, severity, and significance. *Jama,*1990; **263(4):** 527-532.

17. Linn BS, Zeppa R. Stress in junior medical students: relationship to personality and performance. *J Med Educ,* 1984; **59(1)**: 7-12.

18. Shapiro SL, Shapiro DE, Schwartz GE. Stress management in medical education: a review of the literature. *Acad Med,* 2000; **75(7):** 748-759.

19. Rosal MC, Ockene IS, Ockene JK, Barrett SV, Ma Y, Hebert JR. A longitudinal study of students' depression at one medical school. *Acad Med,* 1997; **72(6):** 542-546.

20. Clark EJ, Rieker PP. Gender differences in relationships and stress of medical and law students. *J Med Educ,* 1986; **61(1):** 32-40.

21. Niemi PM, Vainiomaki PT. Medical students' distress - quality, continuity and gender differences during a six-year medical

programme. *Med Teach,* 2006; **28(2):** 136-141.

22. Newbury-Birch D, White M, Kamali F. Factors influencing alcohol and illicit drug use amongst medical students. *Drug Alcohol Depend,* 2000; **59(2):** 125-130.

23. Pickard M, Bates L, Dorian M, Greig H, Saint D. (2000). Alcohol and drug use in second-year medical students at the University of Leeds. *Med Educ,* 2000; **34(2):** 148-150.

24. Liselotte N. Dyrbye, Matthew R. Thomas and Tait D. Shanafelt. Medical students distress: causes, consequences, and proposed solutions, *Mayo Clin Proc*, 2005; **80 (12)**: 1613-1622.

25. Hays LR., Cheever T, Patel P. Medical student suicide, 1989-1994. *Am J Psychiatry,* 1996; **153(4):**553-555.

26. Saipanish R. Stress among medical students in a Thai medical school. *Med Teach,* 2003; **25**(5)**:** 502-506.

27. Goldberg D. Manual of the General Health Questionnaire. NFER Publishing Company, 1978.

28. Wan Mohd Rushidi WM, Amir A, Iran H, Mahmood Nazaar M. Analysis of the psychometric properties of the Malay version of Beck Depression Inventory-II (BDI-II) among postpartum women in Kedah, north west of peninsular Malaysia, *Medical Journal of Medical Sciences*, 2004; **11(2)**: 19-25.

29. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Archives of General Psychiatry. 1961; **4**: 561 – 571.

30. Beck AT, Steer RA, Brown BK. Psychometric properties of Beck Depression Inventory: Twenty five years of evaluation. Clinical Psychology Review. 1988; **8**: 77 – 100.

31. McDowell I. *Measuring health: A guide to rating scales and questionnaires*. 3rd ed. New York; Oxford University Press, 2006.

32. Costello, Anna B, Jason O. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, Practical Assessment Research & Evaluation, 2005; **10 (7)**. http://pareonline.net/getvn?v=10&n=7. Accessed on August 2008.

33. Downing SM. Reliability: on the reproducibility of assessment data, Medical Education, 2004; **38**: 1006-1012.

34. Thomas GT. Interpreting diagnostic test, University of Nebraka Medical Center. Available online: http://gim.unmc.edu/dxtests/Default.htm. Accessed on March 2009.

35. Altman DG. Practical Statistics for Medical Research. London England; Chapman and Hall, 1991.

36. Indrayan A, Satyanarayana L. *Biostatistics for medical, nursing and pharmacy students*. New Delhi: Prentice-Hall of India, 2006.

37. Thomas KC. Fundamentals of educational research (2nd ed.). McGraw-Hill: United States of America, 1996.

Corresponding Author: Muhamad Saiful Bahri bin Yusoff, Lecturer, Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia. **Email**: msaiful@kb.usm.my