The performance of clinical self-report screens amongst a non-clinical population: Examining the use of a popular screen for eating disorders

Michael Munro

Supervised by Sue Waring

March 2013
The performance of clinical self-report screens amongst a non-clinical population: Examining the use of a popular screen for eating disorders

**ABSTRACT**

Research has shown that there are problems within the diagnostic procedure for eating disorders. The first stage of this process involves administering a short self-report survey called the SCOFF test (Morgan, Reid & Lacey, 1999). It is commonly used in the NHS alongside screens for depression and anxiety, namely the PHQ-9 (Kroenke & Spitzer, 2002) and the GAD-7 (Spitzer, Kroenke, Williams & Lowe, 2006). Because these screens are often given together, this research aimed to test the performance of all three of them amongst a non-clinical population. The tests were given both in paper format and made available online to students only, in order to speed up the data collection process. The results suggest that the screens may not be valid measures amongst this population group, as they found many more people scoring significantly than had actually been diagnosed with the relevant disorder. There may be other factors that cause members of this group to experience some of the signs and symptoms associated with these disorders. The findings are discussed in relation to previous research, as well as implications for future study.

**KEY WORDS:** EATING DISORDERS, SELF REPORT, DIAGNOSTIC ISSUES, DEPRESSION, ANXIETY
Introduction

Background

Eating disorders have a profound impact on human life, with 6.4% of adults in the UK showing signs of one (Adult Psychiatric Morbidity Survey, 2007). Mortality rates are also a concern, with one study (Herzog, Greenwood, Dorer, Flores, Ekeblad, Richards, Blais & Keller, 2000) observing a mortality rate of 5.1% in subjects with anorexia nervosa. It is important that a diagnosis be made as early as possible; to avoid extra health risks being placed on the patient. Peebles, Wilson & Lock (2006) found this to be especially true with younger patients, who tend to lose weight more rapidly than older patients do and so could have problems with proper growth in later life. Furthermore, Yeo & Hughes (2011) found that prognosis amongst adolescents can be severely improved when the disorder is detected earlier, and that general practitioners are in a unique position to do this.

Eating disorders often present comorbidly with other disorders such as depression and anxiety. Depression is a major risk factor for eating disorders (Keel, 2006) and tests such as the Beck Depression Inventory are often used during the initial stages of the diagnostic process. McElroy, Frye, Helleman, Altschuler, Leverich, Suppes, Keck, Nolan, Kupka & Post (2011) found that 14.3% of patients with a diagnosis of bipolar disorder also met the criteria for an eating disorder. A review by Beebe (2004) highlights the link between depression and the binge-purge cycle associated with bulimia nervosa. Furthermore, anxiety disorders can predict disordered eating (Levinson & Rodebaugh, 2012) and the two disorders often share risk factors (Konstantellou, Campbell, Eisler, Simic & Treasure, 2011). Clearly, proper diagnosis can improve the prognosis of a person deemed to have an eating disorder. However, there are issues with the current diagnostic criteria. The ‘eating disorder not otherwise specified’ (EDNOS) category appears in the DSM-IV (American Psychiatric Association, 1994) as an ‘umbrella diagnosis’ for any disordered eating that does not fully match the criteria for anorexia nervosa or bulimia nervosa. EDNOS is actually the most common diagnosis made, with an average prevalence of 60% (Fairburn & Bohn, 2005), and yet like many NOS categories for other disorders listed in the DSM they have received very little study (Pincus, Wakefield Davies & McQueen, 1999).

Although the EDNOS category is thought to be a less severe eating disorder, several studies have shown that the psychopathology is often equal to cases of anorexia or bulimia (Key & Lacey, 2002). Nicholls, Charter & Lask (2000) also found that both the DSM-IV and the ICD-10 (World Health Organisation, 1994) were of little value, and the DSM in particular could not specifically classify over 50% of children who clearly had problems with eating behaviours. 42% of them were classified as EDNOS, whilst the other 16.25% were classified as EDNOS by one clinician but not by another.

Because of issues with diagnosis and the fact that the categories share such similar psychopathologies, Fairburn, Cooper and Shafran (2003) proposed a transdiagnostic model. It argued that patients move between the diagnostic criteria over time, and so there are common mechanisms involved in all disordered eating.
Therefore, a broader diagnosis may be sufficient to cover it. Cinton and Norring (2005) support this model, arguing that there are common psychological features amongst the various types of eating disorder and that endless sub-typing of disorders is not useful. Beumont, Garner & Touyz (1994) also agree with this, stating, “The solution of the diagnostic muddle is unlikely to come from tinkering with the criteria once again”.

Hewitt, Flett, Besser, Sherry & McGee (2003) have argued against this model because it downplays the importance of experience and instead focuses entirely on cognition. They argue from the standpoint of research into perfectionism, which Shafran, Cooper & Fairburn (2002) criticised for being multidimensional in the approach to their proposal of the transdiagnostic approach. Hewitt et al (2003) offer a great deal of empirical and theoretical support to a multidimensional model of perfectionism that makes a good case for sub-typing of criteria.

With the American Psychiatric Association proposing to release the DSM 5 in May 2013, it is predicted that the diagnosis of eating disorders will change. Walsh & Sysko (2009) proposed the Broad Categories for Diagnosis of Eating Disorders (BCD-ED) system, which would eliminate the EDNOS category but would only do so by creating a long list of new categories, as opposed to reducing the categorization as suggested by Fairburn et al (2003). Millar, Vaillancourt and Hanna (2009) argued for more of an emphasis on thoughts than behaviours when screening the non-clinical population, as these were more common amongst members of the general population who were later found to have an eating disorder.

According to the official website for the DSM 5, the Eating Disorder Work Group has chosen to add disorders but still keeping the EDNOS category, naming it Feeding or Eating Disorder Not Elsewhere Classified. This suggests that previous calls for a simplification of the diagnostic criteria (e.g. Fairburn et al., 2003) have largely been ignored and the structural approach is being kept. The criteria also appear to measure behaviour as opposed to thoughts, ignoring the proposal from Millar et al (2009). It would appear that the diagnostic paradigm for eating disorders is in a state of crisis. A full review of the proposed changes can be seen at http://www.dsm5.org.

The diagnostic process usually begins with a short screening measure, the most popular one in the UK being Morgan, Reid & Lacey’s (1999) SCOFF test, which at only five questions is a quick and easy screen (Leung, Lee, Lee, Leung, Hung, Lee, Leung, Li, Tse, Wong & Wong, 2009). SCOFF is an acronym, standing for Sick, Control, One stone, Fat and Food, with each letter representing one of the five questions. Each question requires a ‘yes’ or ‘no’ answer, and a score of 2 or more ‘yes’ answers represents a high risk of the patient having an eating disorder. This particular screening measure has a great deal of empirical support. Hautala, Junnila, Alin, Gronroos, Maunula, Karukivi, Liuksila, Raiha, Valimaki & Saarijarvi (2009) found that the SCOFF test was better at detecting an eating disorder than general health examinations, and 81% of students who self-reported an eating disorder in SCOFF remained undetected in the general exam. Cotton, Ball & Robinson (2003) found that the SCOFF was accurate in detecting eating disorders amongst both students and primary care patients. However, they also found that
the SCOFF was not as good as other measures at ruling out eating disorders when they are not the actual diagnosis. In other words, it cannot offer alternative explanations for the scores and could therefore contribute to a process of over-diagnosis.

Studies have also been conducted that have validated the SCOFF in several cultures around the world. For example, Feung et al (2009) found that the test was valid and reliable in China, and Duarte Garcia, Grigioni, Allais, Huoy-Durand, Thibaut & Dechelotte (2011) found similar results in a French patient population. Furthermore, Aoun, Azzam, Jabbour, Hleiss, Honein, & Dechelotte (2010) found that the arabic version of the SCOFF showed good psychometric properties for early detection of eating disorders in a high risk clinical setting.

However, Mond, Myers, Crosby, Hay, Rodgers, Morgan, Lacey and Mitchell (2008) found that the SCOFF slightly underperformed when compared to the Eating Disorder Examination Questionnaire (EDE-Q) (Fairburn & Beglin, 1994), mainly because of the latter's more detailed nature and the fact that it has a higher validity in regards to age and weight. The EDE-Q is a 36 item self-report questionnaire that assesses the specific psychopathology of eating disorders behaviour (Mond et al., 2008). It should be noted that this questionnaire is used to evaluate the nature of an already established eating disorder, not to predict the risk of one being present.

Screening for anxiety and depression can be more difficult, as the two are often misdiagnosed as one another (Ormel, Koeter, van der Brink & van der Willige, 1991). This is because the symptoms can often appear very similar, and so the initial screening may provide similar levels of prevalence for both disorders. It shall be interesting to observe whether this is the case amongst a student sample. Alternatively, the average North American child in the 1980s was more anxious than child psychiatric patients in the 1950s (Twenge, 2007), so it is possible that anxiety will be more prevalent than depression in this sample.

Despite the reports that these self-report measures have received, a clear problem is that most of the testing is performed using clinical populations made up of patients who are displaying signs of an eating disorder. Even when studies have been carried out using the general population (for example Cotton et al., 2003) the screens have been administered by a qualified clinician and most likely in a clinical setting, so a response bias may have been present. Diagnosis of eating disorders is clearly something that needs attention as it has many problems, and perhaps testing the screening measures on the general population, in a non-clinical setting, would be a good place to start.

It is important to assess whether or not the screening measures would produce any ‘false positives’, which is a high amount of people being diagnosed with an eating disorder when they do not actually have one. Although in this case this would be impossible, since the researcher is not in a position to offer diagnosis, this was achievable by asking the participants whether they had been diagnosed with, or felt that they had, a disorder. It is expected that the amount of eating disorder risks would be higher in a clinical setting, but if it is still high in a general population then maybe the validity of the test can be questioned.
Objectives of the Study

This study aimed to examine the use of a clinically validated self-report screen for eating disorders amongst a non-clinical student population. The SCOFF test was administered, as well as the PHQ-9 and GAD-7 (screening tools for depression and anxiety respectively). This was done because the three screens are prominent within the NHS, and are often given together. The aim was to examine how the amount of people from a general student population scoring highly on this type of screen compared to the amount who had either been diagnosed with a disorder, or felt that they had experienced one in the past.

Hypothesis

It was expected that a low number of participants would reach the clinical threshold score for any of the screens, and that this number would be similar to the amount of participants who have either had a previous diagnosis of a disorder, or in the case of eating disorders feel themselves to have met the criteria for one. Furthermore, the validity of each screen, using Cronbach’s alpha, should be at least 0.7 (the agreed score for a valid test), and preferably 0.9, as this is the score often required for a screen to be used in clinical settings (Bland & Altman, 1997).

Methodology

Design

This study utilised a survey, comprised of 3 established self-report screening tools, distributed amongst the student population at Manchester Metropolitan University (MMU) through opportunity sampling. The survey was both paper-based and available online through the free survey website obsurvey.com. The link to the survey was posted in Facebook groups restricted to MMU students. The results of the survey were analysed using the computer software program SPSS.

Survey

The survey consisted of three established self-report screens, some demographic information and questions regarding past diagnosis. The total number of questions in the survey was 27.

Part one consisted of the SCOFF screen, a five-item screen for eating disorders developed by Morgan, Reid & Lacey (1999). Each item scores one point for yes and two or more points would indicate a likelihood of an eating disorder being present.

Part two consists of the PHQ-9, a nine-item screen for depression developed by Kroenke & Spitzer (2002). Each item gives the choice of answers as ‘not at all’, ‘several days’, ‘more than half of the days’ and ‘nearly all of the days’, scoring as 0, 1, 2 & 3 respectively. The scores are distributed as 5 for mild depression, 10 for moderate depression, 15 for moderately severe depression, and 20 for severe depression. The threshold score for clinical significance is 9.
Part three consists of the GAD-7, a seven-item screen for anxiety developed by Spitzer, Kroenke, Williams & Lowe (2006). This screen is distributed and scored in the same way as the PHQ-9 above, with the threshold for clinical significance being 8.

The survey also gathered data about the participants’ gender and age bracket, as it was decided that participants would be more comfortable placing their age into a pre-defined bracket than giving a numerical value. This also helped to increase the anonymity of the participant. Finally, the survey asked whether or not the participant had ever been thought by their GP to have an eating disorder, depression or an anxiety disorder, and as the initial focus of the research was eating disorders, it also asked whether the participant themselves felt that they had had an eating disorder. This was done to try to generate discussion points on the reliability of the screens.

Sample

A total of 107 people responded to the survey, above the original target of 100. Of these, 2 did not fully complete the survey. This lead to their results being discarded, leaving 105 participants and a response rate of 98%. All respondents were students at MMU, with 85% being female (n= 89 and 15% being male (n= 16). As would be expected in a student population, 66% (n= 72) were in the 18-22 age bracket.

Procedure

All participants were either approached at MMU campuses or gathered from MMU-only Facebook groups to complete the survey. A short pilot indicated that the survey took approximately 5 minutes to complete, and participants were informed of this relatively quick completion time as an added incentive to complete the survey. They were also informed that they were under no obligation to complete it, but that handing it back (paper form) or pressing ‘submit’ (online form) indicated that they gave their consent for their data to be used in the study. It is also important to note that the participants completing the paper copy were ‘left alone’ by the researcher, meaning that they were not likely to have been affected by the immediate presence of the person collecting the results. This was an attempt to control for response bias. The survey itself is included as appendix 1.

Ethics

This study was carried out in accordance with the British Psychological Society’s (BPS) ethical guidelines. None of the participants used were classed as vulnerable, and they did not come to any physical or psychological harm during the study. There was also no coercion or deception required to recruit participants. No other ethical guidelines were breached. For further information, please see the ECF and AEAF forms attached as appendix 3 and 4 respectively.
Results

105 participants completed the three screens, and the mean scores and standard deviations for each screen, as well as the necessary score for clinical significance, are shown in Table 1 below.

Table 1
Mean scores and standard deviations for each screen

<table>
<thead>
<tr>
<th>Screen</th>
<th>Clinical significance score</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOFF</td>
<td>2</td>
<td>0.94</td>
<td>1.11</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>9</td>
<td>5.37</td>
<td>4.59</td>
</tr>
<tr>
<td>GAD-7</td>
<td>8</td>
<td>4.90</td>
<td>4.83</td>
</tr>
</tbody>
</table>

None of the screens yielded a mean score that was equal to or above the one showing clinical significance, suggesting that the average participant on this study was not ‘at risk’ for any of the three disorders.

Table 2 below shows the distribution of scores between the genders.

Table 2
Frequency, mean scores and standard deviations for each gender.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>SCOFF Mean</th>
<th>S.D</th>
<th>PHQ-9 Mean</th>
<th>S.D</th>
<th>GAD-7 Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>16</td>
<td>0.50</td>
<td>0.82</td>
<td>4.25</td>
<td>4.09</td>
<td>4.31</td>
<td>4.06</td>
</tr>
<tr>
<td>Female</td>
<td>89</td>
<td>1.02</td>
<td>1.14</td>
<td>5.57</td>
<td>4.67</td>
<td>5.01</td>
<td>4.97</td>
</tr>
</tbody>
</table>

As can be seen from Table 2, there were many more females than males (F= 89, M=16) amongst the participants in this study. However, the mean score for each screen did not differ significantly between the genders, and this lack of significant difference was confirmed using a Spearman’s Rho calculation for each test (SCOFF: p= .18, N= 105, p > .05, PHQ-9: p= .12, N= 105, p > .05, GAD-7: p= .04, N= 105, p > .05.

As would be expected with data gathered from undergraduate students, the vast majority of participants in this study fell into the ‘18-22’ age bracket (n=72). Table 3 below shows the distributions of scores for each age bracket.
Table 3  
Frequency, mean scores and standard deviations for each age range.

<table>
<thead>
<tr>
<th>Age Range</th>
<th>Frequency</th>
<th>SCOFF Mean</th>
<th>S.D</th>
<th>PHQ-9 Mean</th>
<th>S.D</th>
<th>GAD-7 Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-22</td>
<td>72</td>
<td>0.94</td>
<td>1.16</td>
<td>5.71</td>
<td>4.43</td>
<td>4.89</td>
<td>4.93</td>
</tr>
<tr>
<td>23-26</td>
<td>11</td>
<td>0.73</td>
<td>1.10</td>
<td>4.73</td>
<td>4.29</td>
<td>5.09</td>
<td>3.08</td>
</tr>
<tr>
<td>27-30</td>
<td>6</td>
<td>1.17</td>
<td>1.17</td>
<td>3.67</td>
<td>7.09</td>
<td>4.83</td>
<td>7.57</td>
</tr>
<tr>
<td>31-35</td>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
<td>7.33</td>
<td>5.13</td>
<td>6.33</td>
<td>3.51</td>
</tr>
<tr>
<td>36-40</td>
<td>1</td>
<td>2.00</td>
<td>0.00</td>
<td>12.00</td>
<td>0.00</td>
<td>9.00</td>
<td>0.00</td>
</tr>
<tr>
<td>41+</td>
<td>12</td>
<td>0.92</td>
<td>0.90</td>
<td>3.75</td>
<td>4.16</td>
<td>4.17</td>
<td>4.95</td>
</tr>
</tbody>
</table>

Although the ‘36-40’ age bracket appears to show scores at or above the level of clinical relevance for each test, it should be noted that this group contained only one participant and so it is impossible to label this participant as typical of the age group. Furthermore, a one way between-subjects ANOVA showed that there was no significant correlation between age range and scores on any of the screens (SCOFF: $F(5,99) = 0.31, p>0.05$, PHQ-9: $F(5,99) = 1.12, p>0.05$, GAD-7: $F(5,99) = 0.25, p>0.05$).

It was expected that experience of diagnosis would correlate with scoring highly on the screen relevant to the past diagnosis. Table 4 below shows the distribution of scores for each screen between those who have and haven’t experienced diagnosis from their General Practitioner (GP) in the past.

Table 4  
Frequency, mean scores and standard deviations for whether or not the participant had been diagnosed by their GP.

<table>
<thead>
<tr>
<th>GP Opinion</th>
<th>SCOFF</th>
<th>PHQ-9</th>
<th>GAD-7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f</td>
<td>Mean</td>
<td>S.D</td>
</tr>
<tr>
<td>Yes</td>
<td>6</td>
<td>1.67</td>
<td>1.37</td>
</tr>
<tr>
<td>No</td>
<td>99</td>
<td>0.90</td>
<td>1.08</td>
</tr>
</tbody>
</table>

Although the difference in the means was reasonably large, it was still the case that the difference in diagnosis history was not a significant predictor of score, as shown in the independent t-tests described below. Also of note is that, even when previously diagnosed with an eating disorder, the mean score on the SCOFF screen was below the clinically significant score of two. This was not the case for the other screens.

An independent t-test showed that the difference between the GP having thought the participant had an eating disorder and not having thought this was not statistically significant ($t = 1.67$, df = 103, $p > .005$, one-tailed). The magnitude of the differences in the means (mean difference = .768, 95% CI: -.148 to 1.684) was medium ($d = 0.69$).
An independent t-test showed that the difference between the GP having thought the participant had depression and not having thought this was not statistically significant (t = 2.95, df = 103, p > .005, one-tailed). The magnitude of the differences in the means (mean difference = 3.54, 95% CI: 1.158 to 5.929) was medium (d = 0.77).

An independent t-test showed that the difference between the GP having thought the participant had an anxiety disorder and not having thought this was not statistically significant (t = 8.22, df = 103, p > .005, one-tailed). The magnitude of the differences in the means (mean difference = 9.52, 95% CI: 7.221 to 11.812) was large (d = 1.97).

In summary, it was found that having a previous diagnosis of a disorder was not significantly linked to scoring above the clinical threshold on a screen for that disorder.

Table 5 below shows the distribution and mean score on the SCOFF test for those who feel or do not feel themselves that they have had an eating disorder.

**Table 5**
Frequency, mean scores and standard deviations of patients who either did or did not feel that they have had an eating disorder

<table>
<thead>
<tr>
<th>Own Opinion</th>
<th>Frequency</th>
<th>Mean Score</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>14</td>
<td>1.86</td>
<td>1.51</td>
</tr>
<tr>
<td>No</td>
<td>91</td>
<td>0.80</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 5 shows that the mean score on the SCOFF test was slightly higher for participants who felt that they have had an eating disorder. However, an independent t-test showed that this difference was not statistically significant (t= 3.49, df= 103, p > 0.05). The magnitude of the differences in the means (mean difference = 1.06, 95% CI: 0.46 to 1.65) was large (d= 0.96).

Table 6 below shows the number, and percentage, of participants who scored at or above the clinical threshold for each screen. For ease of reference, the threshold score for each screen is shown within the table.

**Table 6**
Amount of participants meeting clinical significance in each screen

<table>
<thead>
<tr>
<th>Screen</th>
<th>Clinical Threshold Score</th>
<th>Amount</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOFF</td>
<td>2</td>
<td>31</td>
<td>29.5</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>9</td>
<td>24</td>
<td>22.9</td>
</tr>
<tr>
<td>GAD-7</td>
<td>8</td>
<td>22</td>
<td>18.1</td>
</tr>
</tbody>
</table>
Table 6 shows that the percentage of participants meeting clinical significance for the SCOFF test is 29.5%. For the PHQ-9 it is 22.9%, and for the GAD-7 it is 18.1%. The average rate of clinical significance in this study is 23.5%.

15 people met clinical significance on both the SCOFF and the PHQ, and 15 people met clinical significance on the SCOFF and the GAD-7. This makes up 14% of the participants in each case.

Table 7 below compares the amount of people showing clinical significance on each test, together with the amount who had been diagnosed with each disorder, and in the case of eating disorders the amount who felt that they had a disorder.

Table 7
Comparison between those who show clinical significance and those who have been diagnosed with each disorder

<table>
<thead>
<tr>
<th>Screen</th>
<th>Amount of people showing clinical significance</th>
<th>Amount of people diagnosed with relevant disorder</th>
<th>Amount of people who feel they have or have had an eating disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOFF</td>
<td>31</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>24</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>GAD-7</td>
<td>22</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Table 7 shows a stark contrast between the amount who have shown clinical significance and those who have actually been diagnosed with a disorder relevant to the screen. Furthermore, less than 50% of the participants who met clinical significance for the possible presence of an eating disorder actually felt themselves that they had one.

Table 8 below shows the Cronbach’s alpha (α) score for each separate screen. Whilst a score of 0.7 is generally valued as a reliable score, use of a screen in clinical settings usually demands a score of at least 0.9.

Table 8
Cronbach’s Alpha score for each screen

<table>
<thead>
<tr>
<th>Screen</th>
<th>Cronbach’s alpha score (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOFF</td>
<td>0.53</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.85</td>
</tr>
<tr>
<td>GAD-7</td>
<td>0.91</td>
</tr>
</tbody>
</table>

The GAD-7 was the only screen showing clinical significance in this study, with a score of 0.91. Whilst not clinically relevant, the PHQ-9 scored 0.85 which is above the minimum reliability threshold of 0.7. However, the SCOFF test scored a mere 0.53 which does not meet the acceptable Cronbach’s alpha score for reliability.
Discussion

Summary of Findings

The results of the study show that the mean score on each screen was not above the clinical relevance threshold for that particular screen, and that age and gender were not significant predictors of the score on any of the three screens. Furthermore, there was no significant relationship between past diagnosis for each condition and the final score on the appropriate screen. However, the number of people scoring at or above the clinical relevance threshold for each screen was much higher than the amount of people who had been diagnosed. More people met clinical relevance on the SCOFF than the PHQ-9 or GAD-7, and a relatively small number of people met clinical relevance on both the SCOFF and one other screen. Because the scoring and thresholds are different for each screen, it is not possible to statistically test this relationship.

This may point to the screens not being valid measures of eating disorders, depression, or anxiety disorders. Cronbach’s Alpha scores show that the SCOFF is not at an acceptable level of validity, and whilst the PHQ-9 does reach a generally acceptable score, only the GAD-7 reaches a score necessary to be deemed valid enough for clinical use.

Interpretation of Findings

The SCOFF test has received a great deal of empirical support from a variety of studies, for example Cotton, Ball & Robinson (2003) and Leung et al. (2009). However, it performed very poorly when examined using the Cronbach’s alpha calculation, failing even to achieve the minimum acceptable score. This suggests that the SCOFF test is not a reliable screen for eating disorders, and casts doubt on the results.

According to research such as Keel (2006) and McElroy et al. (2011), there is a definite link between eating disorders and depression. Furthermore, Levinson and Rodebaugh (2012) claimed that eating disorders were linked with anxiety disorders. However, the results of this study suggest that neither of these links is strong, with only 14% of participants showing comorbidity for either eating disorders and depression, or eating disorders and anxiety. This is not a particularly significant finding, as it means that 86% of the participants did not show any comorbidity.

The poor performance of the SCOFF test may be linked to wider issues with the current diagnostic procedure for eating disorders. There are multiple categories for eating disorders, and yet the SCOFF test does not indicate which category the person is ‘at risk’ for, making the overall diagnosis much harder. Many have argued against the nature of categories, since they over-complicate the issue and often feature similar pathologies (see Key & Lacey (2002) or Fairburn et al. (2003) for more detailed discussion of this). The SCOFF test would therefore likely be a more valid measure if it did not have to contend with the issue of categorisation, as it would simply point out whether or not there is a likely eating disorder present. However, as this is unlikely because recommendations for DSM V seem to
propose even more categories (for example Walsh & Sysko, 2009), the SCOFF test will remain challenged because of its overall simplicity.

The wider issue of diagnosis is relevant to this study. Since the presence of a ‘disorder’ is simply decided by the presence of signs and symptoms as decided by ‘experts’, the actual nature of a disorder can change over time as research and the public Zeitgeist shifts. This can be illustrated in the way that disorders can change, grow or even disappear between editions of the DSM or ICD. This in turn means that someone who was diagnosed as having a disorder suddenly stops being diagnosed with it, as the criteria changes. They may well still be suffering greatly and in need of treatment, but they will no longer qualify for it.

This means that all screening measures for disorders should be looked at critically, as they may not be permanent. It is not implausible that in years to come, the SCOFF, PHQ-9 or GAD-7 may be replaced by another screen purporting to do the same thing, yet comprised of different items. This may mean that the results of this study would not transfer over and people who were seen as clinically relevant in this study would not be in the next, regardless of how they feel regarding their symptoms.

The amount of people showing clinical relevance is much higher than those who have actually been diagnosed with the disorder before, or in the case of eating disorders have felt that they have had one. One possible explanation of this is that the screens were producing ‘false positives’, meaning that they indicated the likelihood of a disorder when it was not actually there. It is important to note here the distinction between screening and diagnosis, as a score on a screen does not mean that the individual actually meets all of the necessary criteria for a disorder. Therefore, scoring highly may simply mean that the individual is at risk of developing the disorder, or that they show some of the traits associated with that disorder but not enough of them to warrant diagnosis. This can be linked back to the earlier argument regarding the diagnostic process, as potentially these individuals may meet the criteria for diagnosis at a later time depending on how the categories shift.

**Evaluation**

The sampling used in this study may have created some problems. As the sample was entirely comprised of students, the majority of participants were in the '18-22' age bracket and so this group may have been over-represented in the sample. However, this is the most common age group for students, and given that the study aimed to investigate students only, this is perhaps not a relevant problem. However, females were significantly more prominent in the sample than males. This is because much of the data collection was carried out on a campus where the main subjects are Psychology and Nursing, which are widely studied by females. Visiting other campuses more often may have produced a more balanced sample in terms of gender.

Final-year psychology students were also highly represented in the data, as the researcher knew many of them personally and so asked them to participate first to ease the data collection process. This may have skewed the results, as final-year
students will likely show more signs of being stressed and anxious due to mounting coursework deadlines, financial worries and a lack of certainty regarding the future. In fact, all students may fall prey to these worries to a certain extent, and so it must be considered that these signs and symptoms may be indicative of the student population rather than indicating the presence of a disorder. As an aside, it may have been incorrect to label all participants as part of a non-clinical population, because if they were receiving treatment at the time of the study, or had done in the past, then they would be part of a clinical population at the time. Ideally, the study would have recruited a larger sample in order to try and attach more significance to the findings, with a more even spread of age and gender. However, the researcher acknowledges that time constraints did not allow for this, as adequate time needed to be allocated to write up this report, and so it was decided not to collect any more data after 31st January 2013.

It would be desirable to have a clinical population sample that had completed the same self-report screens, in order to compare the two groups. However, this is impossible for an undergraduate student to do due to issues of access and ethics, and whilst studies involving a clinical group may already exist, it would be more prudent to control for sample size, age range and any extraneous variables to make the comparison as fair as possible. To the researcher’s knowledge, the student population has never been examined in relation to these screens before, and so these results could be considered a ‘baseline’ and a starting point for future research.

In general, self-report screens should also be examined cautiously. It is always possible that the participant does not indicate their true feelings, either through not wanting to appear like anything is ‘wrong’ with them (social desirability bias) or simply just to get through the survey quickly and stop the researcher pestering them. The researcher attempted to counter these issues by giving the participants plenty of space and time to complete the survey, and assuring them that all responses will remain anonymous.

**Further Research**

This study could be adapted in three ways in the future, if adequate time and funding were secured. Firstly, a much larger and more representative student sample could be recruited in order to lend more significance to the statistical results. Secondly, a sample from a clinical population could also be recruited in order to make it a comparative study. Both samples would contain the same amount of participants if possible, and as even a spread of age and gender as could be realistically achieved. Thirdly, participants who were scored as clinically relevant on any of the screens could be invited for a follow-up interview to assess whether or not there are factors other than the possible existence of a disorder that could have contributed to their score.

It is acknowledged that this would be an expensive and time-consuming piece of research, but it is the view of the researcher that it may produce a lot of knowledge about the efficacy of these self-report screens and prompt discussion about the nature of screening, diagnosis and disorders.
Conclusion

This study found a prevalence rate for either eating disorders, depression or anxiety of 23.5% amongst a student population. However, it has been acknowledged that there may be other factors influencing the results of this study, such as the nature of the group studied and the poor reliability found in the screens used. Further research would look to using a greater sample size and comparing this non-clinical population with a clinical one, as well as following up with high-scoring participants in order to assess how many people deemed ‘at risk’ actually are.
References


Cotton, M.A., Ball, C. & Robinson, P. (2003). The SCOFF questionnaire was less sensitive but more specific than the ESP for detecting eating disorders. *Evidence Based Nursing*, 6, 118.


