A False Sense of Security: Safety Behaviors Erode Objective Speech Performance in Individuals With Social Anxiety Disorder

Karen Rowa
Anxiety Treatment and Research Centre, St. Joseph’s Healthcare, Hamilton, and McMaster University

Jeffrey R. Paulitzki
University of Waterloo

Maria D. Ierullo
Anxiety Treatment and Research Centre, St. Joseph’s Healthcare, Hamilton

Brenda Chiang
University of Waterloo

Martin M. Antony
Ryerson University

Randi E. McCabe
Anxiety Treatment and Research Centre, St. Joseph’s Healthcare, Hamilton, and McMaster University

David A. Moscovitch
University of Waterloo

In the current study, 55 participants with a diagnosis of generalized social anxiety disorder (SAD), 23 participants with a diagnosis of an anxiety disorder other than SAD with no comorbid SAD, and 50 healthy controls completed a speech task as well as self-reported measures of safety behavior use. Speeches were videotaped and coded for global and specific indicators of performance by two raters who were blind to participants’ diagnostic status. Results suggested that the objective performance of people with SAD was poorer than that of both control groups, who did not differ from each other. Moreover, self-reported use of safety behaviors during the speech strongly mediated the relationship between diagnostic group and observers’ performance ratings. These results are consistent with contemporary cognitive-behavioral and interpersonal models of SAD and suggest that socially anxious individuals’ performance skills may be undermined by the use of safety behaviors. These data provide further support for recommendations from previous studies that the elimination of safety behaviors ought to be a priority in cognitive behavioral therapy for SAD.

Keywords: social anxiety; social phobia; speech performance; safety behaviors; social skills
whether there are actual deficits present or not (Clark & Wells, 1995; Rapee & Heimberg, 1997). While studies consistently support the notion that adults and children with SAD underestimate their social performance and abilities (e.g., Ashbaugh, Antony, McCabe, Schmidt, & Swinson, 2005; Furukawa et al., 2009; Rapee & Lim, 1992; Schmitz, Krämer, & Tuschen-Caffer, 2011; Stopa & Clark, 1993; Voncken & Bogels, 2008), the extant research is less consistent about whether people with SAD show objectively more signs of anxiety and exhibit poorer performance than others on social tasks. Theoretical models suggest that objective social performance for people with SAD may be eroded by self-focused attention, self-monitoring, and the overuse of safety behaviors (Clark & Wells, 1995; Rapee & Heimberg, 1997). Understanding social performance by people with SAD is important because certain aspects of performance have an effect on how others perceive an individual (Alden & Taylor, 2004), including desire for further contact (Alden & Wallace, 1995; Meleshko & Alden, 1993; Papsdorf & Alden, 1998). Performance difficulties can also have an impact on school and career trajectories (Beidel, Turner, & Dancu, 1985). Understanding any existing social performance deficits and the contributors to these deficits can and should be addressed in treatment.

Research on social performance has spanned two main types of social tasks: interaction tasks and performance tasks. Interaction tasks generally involve having participants with SAD interact with one or more social partners in a conversation or “getting to know you” scenario. Performance tasks generally involve asking people with SAD to give an impromptu speech in front of a video camera and/or an evaluative audience. In past studies, participants with SAD or elevated social anxiety tend to receive poorer performance ratings on interaction tasks than healthy controls (Baker & Edelmann, 2002; Norton & Hope, 2001; Voncken & Bogels, 2008), individuals with other anxiety disorders (Fydrich, Chambless, Perry, Buergener, & Beazley, 1998), and individuals with dysthymia (Norton & Hope, 2001). In a recent study, people with high social anxiety were also rated as less “likable” than those with low social anxiety in a first-meeting conversational task (Voncken & Dijk, 2013). However, results are not as consistent for performance ratings on speech tasks. Some studies have found performance deficits on speech tasks for people with SAD compared to controls (Levitan et al., 2012; Moscovitch & Hofmann, 2007), while others have found no global differences in performance (Rapee & Lim, 1992; Voncken & Bogels, 2008). Voncken and Bogels argue that interaction tasks are less structured and require a more complicated set of skills as compared to speech tasks, which may contribute to greater difficulties for people with SAD on those types of tasks. On the other hand, speech tasks may be generally more difficult for all participants, including healthy control participants, thus helping explain the lack of consistency in performance deficits across studies.

Thus, while it is clear that people with high social anxiety or SAD have consistent performance deficits in interaction tasks, more research using speech tasks is necessary to help clarify mixed results. It also seems important to include comparison groups beyond healthy controls to investigate possible performance deficits in SAD. Research using a dysthymic comparison group found that performance ratings by observers were most positive for the healthy comparison group and least positive for the SAD group, with ratings of the dysthymic group in between those of the other two (Norton & Hope, 2001). In a study investigating performance in individuals with SAD, anxious controls, and healthy controls, ratings of observable negative behaviors during speech and conversation tasks were less positive for the SAD group as compared to the other two groups, which did not differ from each other (Stangier, Heidenreich, & Schermelleh-Engel, 2006). However, sample sizes in this study were small and individuals with comorbid depression were excluded, thus reducing the generalizability of the results. Further research using an anxious control group with fewer exclusion criteria would help confirm that performance deficits are specific to social anxiety per se rather than characteristic of any types of clinical anxiety difficulties. Thus, the first aim of the current study was to examine objective performance on a speech task across a large sample of participants with a principal diagnosis of SAD, those with a principal diagnosis of another anxiety disorder (with no comorbid SAD), and healthy control participants. Participants with comorbid depression were included in this sample to increase the representativeness of the sample, and depressive symptoms were included as a covariate in analyses to ensure that any group differences found were not accounted for by comorbid low mood. Based on results of previous studies, reviewed above, we hypothesized, first, that people with SAD would receive poorer performance ratings than those with other anxiety disorders who, in turn, would receive similar performance ratings to healthy controls. Second, we predicted that results would remain consistent when depressive symptoms were covaried in analyses.

In addition to examining the nature of objective speech performance in a large sample of individuals
with SAD relative to anxious and healthy controls, the second aim of the current study was to investigate whether the use of safety behaviors might help to explain why objective performance tends to be impaired in SAD. Safety behaviors may be defined as overt or covert actions designed to help a person feel safe in an anxiety-provoking situation (Salkovskis, 1991). Although there is no consensus definition of safety behaviors, most researchers and clinicians agree that the function of safety behaviors is to prevent a person’s feared outcomes. Researchers have operationalized social anxiety-related safety behaviors in different ways. Some have categorized them according to their function into two broad categories: (a) avoidance behaviors (e.g., minimizing the amount of talking, low levels of self-disclosure) and (b) impression management behaviors (e.g., overrehearsing what one will say, excessive self-monitoring during social interactions; Plasencia, Alden, & Taylor, 2011). Others have divided social-anxiety-related safety behaviors into three categories: (a) inhibiting/restricting behaviors (e.g., remain silent), (b) active behaviors (e.g., rehearse what one will say), and (c) physical symptom management (e.g., wear clothes or makeup to hide blushing; Cumming et al., 2009). Moscovitch (2009) has emphasized that safety behaviors in social anxiety function primarily as self-concealment strategies designed to prevent one from having one’s perceived self-flaws exposed to evaluative others in social situations. Although safety behaviors may create perceptions of safety and are commonly used by people with elevated social anxiety symptoms (Alden & Bieling, 1998; McManus, Sacadura, & Clark, 2006), they appear to have several negative consequences, as emphasized by influential CBT models of SAD (Clark & Wells, 1995; Rapee & Heimberg, 1997). The use of safety behaviors by socially anxious individuals has been shown to elicit negative reactions from others (Plasencia et al., 2011) and to mediate the relationship between anticipatory self-portrayal concerns and elevated postevent negative affect (Moscovitch et al., 2013). Moreover, the elimination of safety behaviors leads to better performance ratings by objective observers (Furukawa et al., 2009; Taylor & Alden, 2011) or conversation partners (McManus, Sacadura, & Clark, 2008), greater symptom reduction across challenging social tasks (Kim, 2005), and more accurate self-appraisals (Taylor & Alden, 2010).

Taken together, there is compelling evidence to suggest that the overuse of safety behaviors by socially anxious individuals during social interactions may significantly contribute to objective performance impairments. The demands of most safety behaviors would likely co-opt attentional resources helpful to a solid performance by focusing attention on the self rather than on the task at hand. The unintended side effects of safety behaviors may also mimic poor performance. For example, if someone is excessively rehearsing what he or she should say next, this would have a negative impact on how fluidly ideas are presented in a formal speech. Similarly, long pauses or stretches of silence during a presentation that are used by an individual to avoid making mistakes would have the unintended effect of becoming a perceived mistake. Further, the use of safety behaviors may erode perceived performance by creating an interpersonal barrier between the person who uses safety behaviors and his/her social partners, thereby preventing the type of interpersonal connection that is required for strong social performance across various tasks. One previous study has directly examined this question, finding that self-reported use of safety behaviors partially mediated the relationship between diagnostic status and objective ratings of “negative” behaviors (e.g., looking nervous) during a speech and conversation task (Stangier et al., 2006). As noted above, however, the Stangier et al. study was comprised of a small sample and excluded participants with comorbid depression. Our goal in the current study was to replicate Stangier et al.’s results and build upon this initial study by recruiting a larger sample of participants with SAD, other anxiety disorders, and healthy controls with minimal exclusion criteria, using a well-validated measure of safety behaviors, and using a well-validated observational coding scheme of objective performance. Thus, our third hypothesis was that self-reported use of safety behaviors during the speech task would mediate the relationship between diagnostic group and performance ratings, such that the presence of a diagnosis of SAD would contribute to more negative performance ratings by observers because of greater use of safety behaviors by participants.

Material and Methods

PARTICIPANTS

A total of 146 participants were recruited for the present study, including: (a) 55 individuals with a principal DSM-IV-TR (American Psychiatric Association, 2000) diagnosis of generalized SAD; (b) 23 individuals with a principal anxiety disorder diagnosis other than SAD; and (c) 50 healthy controls.

Participants were recruited from individuals referred to a specialty Anxiety Disorders Clinic at a Canadian public hospital. Diagnoses were made using the Structured Clinical Interview for DSM-IV Disorders (SCID; First, Spitzer, Gibbon, & Williams, 2002) administered by graduate-level clinicians who received thorough training in diagnostic
assessment. Training on SCID administration involved observing at least three SCID interviews conducted by experienced interviewers and then completing at least three SCID interviews while being observed by an experienced interviewer, with evidence of strong interrater reliability. The principal diagnosis was determined, in consultation between the clinician and the patient, to be the problem that was associated with the greatest interference and/or distress for the patient. Diagnoses were reviewed and confirmed during a weekly team meeting led by a psychologist with over 15 years of experience in SCID administration. The interrater reliability across two independent raters at our Anxiety Disorders Clinic for principal diagnosis was $k = .89$ for a subset of administrations of the SCID ($N = 13$). Exclusion criteria for this study included active symptoms of psychosis or mania, comorbid diagnoses of current substance abuse or dependence that were thought to interfere with the person’s ability to complete the study, and significant suicidality.

The nonclinical control group was recruited through advertisements posted in the community and on the Internet. All interested participants completed a phone interview about current and past mental health concerns using the SCID screening questions. Participants were invited to be part of the study if diagnostic criteria were not met for any current mental disorders and if they did not report a significant history of anxiety or mood difficulties or other psychological problems.

This study is based on data collected as part of a larger multiphase study investigating self-portrayal concerns in SAD. Data on self-appraisals including their relationship with safety behaviors are presented elsewhere (Moscovitch et al., 2013).

**PROCEDURE**

Participants provided written informed consent to participate in the study. The study was approved by both the Hamilton Integrated Research Ethics Board and the Human Research Ethics Board at the University of Waterloo. After completing questionnaires that were part of the larger study, participants were asked to give an impromptu 5-minute speech on three randomly chosen topics in front of a research assistant and a video camera. They were informed that their speeches were being video recorded and would be rated at a later time by research assistants who were blind to their diagnostic status. Participants were asked to talk on any or all of three speech topics for 5 minutes (My views on abortion; What I did on my last vacation; and What it means to be a “proud Canadian”). Most participants spoke on all three topics within their speech. Immediately following their speech, participants completed the Subtle Avoidance Frequency Examination (SAFE; Cumming et al., 2009).

**MEASURES**

The SAFE (Cumming et al., 2009) is a 32-item self-report measure of safety behavior use in social situations. Each item is rated on a 5-point scale from 1 (never) to 5 (always). Items cluster into three subscales including inhibiting/restricting behaviors, which are used to avoid attracting attention (e.g., “remain silent”; “avoid eye contact”), active behaviors, which are used in an attempt to enhance social performance (e.g., “rehearse sentences in your mind”), and physical symptom management behaviors, which are behaviors designed to minimize blushing and physical symptoms (e.g., “wear clothes or makeup to hide blushing”). The total score of the SAFE summed across these subscales was the primary focus in the current study. Total scores range from 32 to 160. The SAFE has strong psychometric properties, including good internal consistency and adequate convergent and divergent validity.

The Perception of Speech Performance scale (PSP; Rapee & Lim, 1992) was used to assess objective performance. This is a 17-item measure that contains 12 specific items (e.g., “kept eye contact,” “stuttered”) and 5 global items (e.g., “generally spoke well”), each rated on a 5-point scale from 0 (not at all) to 4 (very much). Positive items are reverse coded so that higher scores indicate poorer performance. The PSP was used by two trained coders, who each rated all videotaped speeches. Coders were blind to diagnostic status of participants. There was some missing data (0.07%) from one coder, which was imputed using the expectation-maximization method (e.g., Schafer, 1997). To determine interrater reliability, intraclass correlations (ICCs) were computed separately for both the total of the specific items and the total of the global items. Given that all participants were rated by both coders, the reliability of mean ratings provided by both coders was the main focus. Also, because assessing the degree of consistency in the coder ratings was more important than agreement in absolute values of nominal ratings, a two-way, average-measures model was utilized to calculate a consistency-based ICC coefficient. The consistency ICC was .81 for specific items and .79 for global items. The correlation between the global and specific subscales was .86. For analyses, the raters’ scores were averaged to form one PSP specific rating and one PSP global rating.

The Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996) is a 21-item measure of depressive symptoms. It has strong psychometric properties (Beck et al., 1996).
**Data Analysis**

Group differences on coded performance ratings and self-report of the use of safety behaviors were analyzed using analyses of variance (ANOVAs) and analyses of covariance (ANCOVAs), wherein BDI-II total scores were entered as the covariate in order to statistically control for the impact of depression level on the main analyses of interest. Significant group differences were followed up using Tukey HSD posthoc tests for ANOVAs and Sidak-corrected posthoc tests on the covariate-adjusted means for ANCOVAs. We hypothesized that people with SAD would receive poorer performance ratings than those with other anxiety disorders, who would not differ from healthy controls, and this hypothesis was investigated using ANOVAs with diagnostic status as the categorical variable and both the global and specific ratings on the PSP as the dependent variables. These analyses were repeated with ANCOVAs to take into account the potential impact of depressive symptoms. We also hypothesized that people with SAD would report using more safety behaviors during a speech task when compared to healthy controls and that safety behavior use in the anxious control group would fall between that of the other two groups. This hypothesis was also investigated using ANOVA and ANCOVA. Given the known deleterious effects of safety behaviors on performance in social situations, we expected that people with SAD would engage in more safety behaviors, which would then contribute to poorer observed performance during the speaking task. In other words, we predicted that the relation between group membership and observed performance would be mediated by the extent to which safety behaviors were used during the performance itself. Separate models were generated for each type of performance indicator. The diagnostic status variable was effect-coded such that the healthy controls represented the group of reference. In order to test the hypothesis that the use of safety behaviors during the speaking task mediated the relation between group status and speech performance, direct and indirect effects of group status on performance scores were generated using the Monte-Carlo bootstrapping method in Amos 20.0 (Arbuckle, 2008). Separate analyses were carried out for each type of performance indicator. The sample size for each model was 126 after removing participants with any missing data from each analysis. For each respective model, the reported effects were generated as a result of resampling from the observed cases 2,000 times. The effect-coded group variables reported below effectively make use of the healthy control participants as the relevant reference group. As a preliminary step before analyzing the full model, the relation between group membership and the use of safety behaviors was examined using similar bootstrapping procedures.

**Results**

**Descriptive Statistics**

The mean scores of measures for each of the participant groups are presented in Table 1 and Figure 1. Skew and kurtosis statistics indicated that the data from all measures were normally distributed (see Cohen, Cohen, West, & Aiken, 2003).

**Group Comparisons on Performance Ratings**

The first ANOVA used mean global PSP scores as the dependent variable and indicated significant differences between diagnostic groups, $F(2, 125) = 14.66, p < .001, \eta_p^2 = .190$. Follow-up Tukey-HSD posthoc analyses further indicated that the global performance ratings of participants with a principal diagnosis of SAD ($M = 9.4, SD = 3.5$) were significantly worse when compared to anxious controls ($M = 7.3, SD = 2.4; p = .01$) and healthy controls ($M = 6.3; SD = 2.7; p < .001$). Anxious controls did not receive poorer performance ratings than did participants in the healthy control group ($p = .38$). An ANOVA using specific PSP ratings mirrored the same pattern, $F(2, 125) = 30.19, p < .001, \eta_p^2 = .326$. Post-hoc analyses suggested that SAD participants ($M = 8.9; SD = 3.0$) received poorer performance ratings as compared to anxious controls ($M = 6.3; SD = 1.8; p < .001$) and healthy controls ($M = 5.2; SD = 2.0; p < .001$). The performance of anxious controls was not rated any less favorably than the healthy controls ($p = .18$).

For the first ANCOVA, mean global PSP scores were entered as the dependent variable, BDI-II scores were entered as the covariate and results indicated significant differences among diagnostic groups, $F(2, 113) = 3.76, p = .03, \eta_p^2 = .062$. The depression covariate was not significant, $F(1, 113) = 1.61, p = .21, \eta_p^2 = .014$. Follow-up posthoc analyses further indicated that the global performance ratings of participants with a principal diagnosis of SAD (adjusted $M = 8.9, SE = .54$) tended to be somewhat worse when compared to anxious controls (adjusted $M = 7.1, SE = .67$;
p = .08) and were statistically worse compared to healthy controls (adjusted M = 6.7; SE = .54; p = .05). Anxious controls did not receive poorer performance ratings than participants in the healthy control group (p = .98). An ANCOVA using specific PSP ratings mirrored a similar group pattern, \( F(2,113) = 56.61, p < .001, \eta^2_p = .141 \). Post-hoc analyses suggested that SAD participants (adjusted M = 8.7; SE = .44) received poorer performance ratings as compared to anxious controls (adjusted M = 6.2; SE = .55; p < .001) as well as healthy controls (adjusted M = 5.6; SE = .44; p < .001). The performance of anxious controls was not rated any less favorably than the healthy controls (p = .79). BDI-II scores showed no significant main effect in the second ANCOVA, \( F(1, 113) = 1.83, p = .18, \eta^2_p = .016 \).

**GROUP COMPARISONS OF SAFETY BEHAVIOR USE ANOVA results with SAFE total scores as the dependent variable suggested significant differences among the groups, \( F(2, 123) = 26.20, p < .001, \eta^2_p = .299 \). Based on the follow-up Tukey-HSD posthoc analyses, those with SAD (M = 73.6, SD = 20.6) used safety behaviors considerably more than anxious controls (M = 59.0, SD = 20.7; p < .001), as well as individuals in the healthy control group (M = 47.9; SD = 13.1; p < .001). There was also evidence of more safety behavior use in the anxious control group, when compared to healthy controls (p = .02).

The ANCOVA with SAFE total scores as the dependent variable and BDI-II scores as the covariate suggested significant differences among

---

**FIGURE 1**  PSP global and specific performance ratings by group membership. Notes. PSP = Perception of Speech Performance; SAD = Social Anxiety Disorder. Higher scores represent poorer performance; **p < .01; ***p < .001.

### Table 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>SAD (a) N = 55</th>
<th>Anxious Controls (b) N = 23</th>
<th>Healthy Controls (c) N = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>33.9 (12.4)</td>
<td>35.9 (10.6) c</td>
<td>29.3 (12.3) b</td>
</tr>
<tr>
<td>Sex (% Female)</td>
<td>58.2%</td>
<td>73.9% ct</td>
<td>54.0% bt</td>
</tr>
<tr>
<td>Education (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Finish High School</td>
<td>10.9% bt</td>
<td>0.0% at</td>
<td>8.0%</td>
</tr>
<tr>
<td>Completed High School</td>
<td>12.7%</td>
<td>26.1% ct</td>
<td>10.0% bt</td>
</tr>
<tr>
<td>Some Post-Secondary</td>
<td>32.7%</td>
<td>26.1%</td>
<td>38.0%</td>
</tr>
<tr>
<td>Completed Post-Secondary</td>
<td>38.2%</td>
<td>39.1%</td>
<td>44.0%</td>
</tr>
<tr>
<td>Missing</td>
<td>5.5%</td>
<td>8.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Relationship Status (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>56.4% bc</td>
<td>17.4% ac</td>
<td>80.0% ab</td>
</tr>
<tr>
<td>Married</td>
<td>21.8% b</td>
<td>47.8% ac</td>
<td>12.0% b</td>
</tr>
<tr>
<td>Co-habiting</td>
<td>3.6%</td>
<td>13.0% ct</td>
<td>2.0% bt</td>
</tr>
<tr>
<td>Separated/Divorced</td>
<td>7.3%</td>
<td>4.3%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Widowed</td>
<td>0.0%</td>
<td>4.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Missing</td>
<td>10.9%</td>
<td>13.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>PSP Global (range 0-20)</td>
<td>9.4 (3.5) bc</td>
<td>7.3 (2.4) a</td>
<td>6.3 (2.7) a</td>
</tr>
<tr>
<td>Specific (range 0-48)</td>
<td>8.8 (3.1) bc</td>
<td>6.3 (1.8) a</td>
<td>5.2 (2.0) a</td>
</tr>
<tr>
<td>SAFE Total (range 32-160)</td>
<td>73.6 (20.6) bc</td>
<td>58.7 (20.7) ac</td>
<td>47.9 (13.1) ab</td>
</tr>
<tr>
<td>Active Behaviors</td>
<td>32.7 (9.3) bc</td>
<td>23.5 (9.4) ac</td>
<td>18.0 (6.3) ab</td>
</tr>
<tr>
<td>Restricting Behaviors</td>
<td>31.2 (9.8) bc</td>
<td>26.7 (10.6) ac</td>
<td>22.3 (6.2) ab</td>
</tr>
<tr>
<td>Managing Physical Symptoms</td>
<td>9.8 (4.6) c</td>
<td>8.5 (2.4)</td>
<td>7.6 (2.4) a</td>
</tr>
</tbody>
</table>

**Note.** PSP = Perception of Speech Performance; SAD = Social Anxiety Disorder; SAFE = Subtle Avoidance Frequency Examination. Groups are labeled a-c. Within each column, superscripts indicate significant differences (p < .05 – based on Tukey’s HSD or Chi-square goodness of fit; \( ^* \) p < .10) between the group in that column and the group(s) whose label(s) correspond(s) with that/those superscript(s). Scores are not adjusted based on including the BDI-II as a covariate.
the groups, $F(2, 112) = 5.36, p = .01, \eta^2_p = .087$, and higher depressive scores were related to increased use of safety behaviors, $F(1, 112) = 18.78, p < .001, \eta^2_p = .144$. Based on the follow-up posthoc analyses, those with SAD (adjusted $M = 67.0, SE = 2.9$) used safety behaviors considerably more than anxious controls (adjusted $M = 53.8, SE = 3.6; p = .01$), as well as tending to use somewhat more safety behaviors than individuals in the healthy control group (adjusted $M = 55.8; SE = 2.9; p = .06$). There was no evidence of more safety behavior use in the anxious control group, when compared to healthy controls ($p = .97$).2

SAFETY BEHAVIOR AS A MEDIATOR OF THE RELATION BETWEEN DIAGNOSTIC GROUP AND PERFORMANCE

Results of mediation analyses included a significant relation between the SAD group and SAFE scores ($\beta = .57, p < .001$), but no evident relation between group status and SAFE scores for those in the anxious control group ($\beta = -.05, p = .62$). A second preliminary analysis involved investigating the direct effect of group membership on PSP Global scores (i.e., without SAFE scores in the model). In this case, there was a significant relation between group membership and Global PSP scores for those in the SAD group ($\beta = .47, p < .001$), but again no relation for anxious controls ($\beta = -.09, p = .41$). In the first complete mediation model, global PSP scores were predicted by both the Group and SAFE Total variables (see Figure 2a). In this full model, there was a significant, standardized direct effect of group status on global PSP total scores for those in the SAD group ($\beta = .29, p = .01$) whereas this was not observed for the anxious control group ($\beta = -.07, p = .47$). Furthermore, the standardized indirect effect of group membership on global PSP scores through the reported use of safety behaviors during the speech was significant for the SAD group ($\beta = .17, p = .01$), and was estimated to be within a nonzero range between .07 and .31 based on the 95%, bias-corrected confidence interval. In contrast, there was no significant indirect effect found for those in the anxious control group ($\beta = .04, p = .55$).

Similar mediational analyses were carried out using specific PSP scores. First, excluding the SAFE variable, there was a direct effect of group membership on specific PSP scores observed for the SAD group ($\beta = .63, p < .001$), but no significant relation for anxious controls ($\beta = -.12, p = .19$). The full model, including both the Group and SAFE total scores as predictors of PSP specific scores, is depicted in Figure 2b. A significant, direct effect of group status on specific PSP total scores remained for those in the SAD group ($\beta = .48, p < .001$), but there was no direct effect for the anxious control group ($\beta = -.11, p = .22$). Mirroring the results of the global PSP scores, the standardized indirect effect of group membership through SAFE scores was significant for the SAD group ($\beta = .17, p < .001$), somewhere between .05 and .27 based on the 95% confidence interval. No significant indirect effect was found for those in the anxious control group ($\beta = -.01, p = .53$). Based on the validation ratio (Freedman, 2001) or the mediational ratio (Ditlevsen, Christensen, Lynch, 2 Separate ANOVAs were also completed using each of the three SAFE subscale totals as dependent variables. These results are not reported here in detail given that we did not make any a priori hypotheses about subscale results; however, they were substantively similar to those reported for the SAFE total score. See Table 1 for mean SAFE subscale scores across groups. When depressive scores were included as part of an ANCOVA, group differences were found in the expected pattern for the active behaviors SAFE subscale, suggesting that the active behaviors subscale may be the aspect of safety behaviors with the most unique relevance for SAD. Specific questions about the types of safety behaviors that are most relevant to SAD and performance are a valuable area for future research.
the indirect effect of safety behaviors constituted 38% and 24%, respectively, of the total effect of group membership status on performance ratings for each of the respective models assessing global and specific performance ratings on the PSP.

Discussion

In the current study, we compared objective performance ratings on a speech task between individuals with SAD, other anxiety disorders, and healthy controls. Our first hypothesis, that people with SAD as a principal diagnosis would receive poorer performance ratings than those with other anxiety disorders or healthy controls, was supported. Using both global (e.g., made a good impression) and specific (e.g., stuttered) indicators, coders blind to diagnostic status rated the performance of individuals in the SAD group as poorer than anxious controls and healthy controls, with anxious and healthy controls receiving similar ratings. These results add to a body of research that suggests that people with SAD demonstrate observable difficulties with performance on stressful social tasks and that these objective performance difficulties are specific to SAD rather than generally characteristic of anxiety disorders per se. While this pattern has been consistently demonstrated using interaction tasks (Baker & Edelmann, 2002; Norton & Hope, 2001; Voncken & Bögels, 2008), some studies have found no performance deficits for people with SAD using a speech task (Rapee & Lim, 1992; Voncken & Bögels). Voncken and Bögels hypothesized that speech tasks might be more manageable for people with social anxiety given the high level of structure in a speech as compared to an interaction. Our results add further support to the studies that have found performance deficits even on a structured task such as a speech (Levitan et al., 2012; Moscovitch & Hofmann, 2007; Stangier et al., 2006).

It is difficult to know why our study found performance deficits for people with SAD on a speech task while other studies have not. Perhaps our study, which asked participants to give a speech in front of both a small audience (a research assistant) as well as a video camera for later review, provoked high enough anxiety in participants with SAD that they felt compelled to use safety behaviors, thus leading to more objective deficits as compared to previous studies that have not included video recordings (e.g., Rapee & Lim, 1992; Voncken & Bögels, 2008).

The main pattern of results remained generally consistent when scores on the BDI-II were included as a covariate in analyses, suggesting that the results were not fully explained by comorbid depressive symptoms. However, depressive symptoms did appear to influence one particular result: Global performance ratings on the speech task were no longer statistically different between individuals in the SAD and anxious control groups once depression scores were included as a covariate in analyses. However, results for specific performance ratings were unaffected by the inclusion of depression scores. It is possible that negative affect had a more salient effect on the global performance ratings because of the nature of the items (e.g., generally spoke well) whereas when raters were asked to code performance using very specific indicators (e.g., stuttered), these items were less affected by negative affect or general anxiety symptoms.

Given the growing consensus in the literature that people with SAD demonstrate greater performance deficits than people with other anxiety disorders and healthy controls, it becomes important to consider the factors that likely contribute to poorer performance ratings. In addition to the contribution of state anxiety (Beidel et al., 1985) and negative self-appraisals (Norton & Hope, 2001), one such variable that has received theoretical and research attention is the overuse of safety behaviors. Cognitive behavioral models heavily emphasize the deleterious use of safety behaviors in SAD which are thought to be the result of underestimating one’s abilities (Clark & Wells, 1995), highly valuing the perceptions of “audience” members in social encounters (Rapee & Heimberg, 1997), and attempting to conceal perceived negative self-attributes (Moscovitch, 2009).

We hypothesized that people with SAD would report greater use of safety behaviors during a speech task as compared to anxious and healthy controls (even when controlling for depressive symptoms), and that the use of safety behaviors would mediate the relationship between diagnostic group and performance ratings. The first part of this hypothesis was partially supported: People with SAD reported using more safety behaviors during the speech as compared to the anxious and healthy control groups. However, we also found an association between greater depressive symptoms and more self-reported use of safety behaviors. When depression scores were included as a covariate in analyses, the difference between the SAD group and healthy controls on use of safety behaviors decreased. It appears that low mood, in addition to the presence of social anxiety symptoms, may be driving the use of safety behaviors. Exploratory analyses of the SAFE subscales (see Footnote 2) suggested that the Active

3 These ratios were calculated as 1 – (c/c'), where c is the total effect of group membership on PSP scores (with no other predictors in the model) and c' is the effect of group status on PSP performance, controlling for SAFE total scores.
Behaviors subscale appeared most relevant to SAD once depressive symptoms were included as a covariate. Further research could explore the unique contributions of social-anxiety-relevant motivators (e.g., fear of negative evaluation) and depression-related motivators (e.g., self-criticism; “I don’t think I can do this”) in understanding the need for safety behaviors, including motivations to use particular kinds of safety behaviors.

While previous research has found support for partial mediation of the relationship between diagnostic status and performance ratings by safety behaviors (Stangier et al., 2006), our results suggested that for both global and specific indicators of performance, self-reported safety behaviors strongly mediated the relationship between diagnostic status (SAD group membership) and performance ratings. In other words, the data from the present study were convincingly consistent with the notion that individuals with SAD exhibited objective performance deficits in their speech because they used more safety behaviors. This pattern of results was consistent whether using diagnostic status (SAD group membership) or self-reported social anxiety symptom severity (SPIN scores) in analyses.

These results complement a body of literature describing the deleterious effects of safety behavior use in SAD (see Alden & Taylor, 2004) and the positive impact that the elimination of safety behaviors has on symptom reduction (e.g., McManus et al., 2009). Given that the use of safety behaviors mediated the relationship between diagnostic status and performance ratings for people with SAD in the current study, our results suggest that safety behavior use may be one of the most important predictors of poor performance for people with SAD. Thus, it is possible that social performance skills for people with SAD are at least partially intact, but are masked or undermined by the use of safety behaviors and other concealment strategies. Research supports the notion that eliminating safety behaviors yields better interpersonal outcomes in a “getting to know you” task (Taylor & Alden, 2011), and better performance ratings from observers (Furukawa et al., 2009) and conversation partners (McManus et al., 2008). However, it is unclear to what degree social performance behaviors are generally intact for people with SAD once the confounding influence of safety behavior use is removed. This is an important question for future innovative research, including studies that can measure objective performance in the presence of active SAD symptoms (i.e., before successful treatment efforts) but in the absence of any safety behaviors.

These results are also consistent with cognitive behavioral models of SAD, which suggest that objective performance deficits are not a fundamental difficulty for people with SAD. Instead, observed performance deficits or difficulties with social skills may arise as a result of processes driven by high levels of anxiety, distorted thinking, and an investment in making a good impression (e.g., overuse of safety behaviors, self-focused attention; Clark & Wells, 1995; Rapee & Heimberg, 1997).

One possible limitation of the current study is the use of self-reported safety behaviors vs. observer-coded safety behaviors. Although it would be useful to also have objective indicators of safety behavior use, there may be limitations and challenges associated with measuring these personal and sometimes subtle behaviors via observation and, therefore, assessing safety behavior use via self-report appears to be a reasonable means of collecting this information. For example, it is difficult to imagine an observer reliably being able to ascertain when someone is excessively rehearsing what they will say in their head. Another limitation of the current study is that our research design does not allow us to definitively conclude that safety behavior use caused poorer performance ratings, as both safety behaviors and performance were technically measured at the same time point during the experimental sequence. Longitudinal or true experimental designs in which safety behaviors are manipulated and downstream impact on performance is observed would allow for more definitive conclusions about causal relationships. There is also some overlap between our well-validated measure of safety behaviors, on one hand, and performance, on the other. For example, both measures include items about long pauses or remaining silent. However, for the most part there is not overlap between items and many of the items with some overlap were tapping into different aspects of the safety behavior. To this end, the PSP performance measure generally assessed an observable behavior such as “long pauses,” while the SAFE assessed both the behavior and the motivation behind the behavior (for example, “remained silent” or “rehearsed sentences in your mind”).

The results of the current study have important clinical implications. Our findings support models of treatment of SAD where elimination of safety behaviors is a key goal (Clark, 2005; Moscovitch et al., 2013; Taylor & Alden, 2011). They also support studies that clearly demonstrate improved outcomes of exposure with safety behaviors versus exposure without safety behaviors for social anxiety symptoms (e.g., Morgan & Raffle, 1999) and that the removal of safety behaviors yields more positive social response (Taylor & Alden, 2011). Our results suggest that safety behaviors should be eliminated through treatment efforts before any conclusions are drawn about actual social skills. Indeed, safety behaviors
may ironically disrupt what would otherwise be a solid social performance.

References


Received: October 10, 2013
Accepted: November 14, 2014
Available online 22 November 2014