Effects of a third party observer and anxiety on tests of executive function

Julie E. Horwitz*, Robert J. McCaffrey

University at Albany, State University of New York (SUNY), United States
Accepted 15 February 2008

Abstract

For the past 10 years, research on the effects of observer presence on test performance has expanded in the neuropsychological literature. Previous studies have shown that the presence of a third party observer is associated with poorer performance on tests of effort, attention, concentration, learning, and memory. The present study was designed to investigate whether performance on tests of executive function is similarly impaired by the presence of a third party observer. The study also sought to examine associations among examinee anxiety, observer presence, and performance. Seventy-nine college undergraduates were recruited for the study, and 70 were included in the final analyses. Participants were randomly assigned to either the observation or control condition, and were administered verbal fluency tests, the Trail Making Test (parts A and B), and the Tactual Performance Test, as well as the Fear of Negative Evaluation scale and State-Trait Anxiety Inventory. Multivariate analyses of variance revealed that performance on the combined dependent variables was significantly associated with observer presence. A significant observation condition by trait anxiety interaction was also found. Univariate analyses revealed that performances on semantic fluency and TPT-localization were most strongly associated with observation and trait anxiety, with performance being poorer in the presence of a third party observer. Additionally, effects of trait anxiety on performance in the presence of an observer appear to vary depending on task characteristics. Implications and suggestions for further research are discussed.

© 2008 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Third party observer; Social facilitation; Anxiety; Tactual performance test; Trail making test; Verbal fluency

In the past decade, considerable attention has been given to research on the effects of third party observers on neuropsychological test performance (McCaffrey, Lynch, & Yantz, 2005). The expanding repertory of studies demonstrating significant effects has led to the assertion by many neuropsychologists and neuropsychological associations that third party observation of neuropsychological examinations violates standardized testing procedures, jeopardizes test security, renders interpretation of norms less valid, and may be a breach of ethics and/or standards (e.g., AERA, APA, & NCME, 1999; Axelrod et al., 2000; Essig, Mittenberg, Petersen, Strauman, & Cooper, 2001; Hamsher, Lee, & Baron, 2001; McCaffrey, 2005; McSweeny et al., 1998). Issues surrounding third party observation are particularly salient for the field of forensic neuropsychology, given that attorneys often request to observe evaluations of their...
clients, and in some cases are legally allowed to do so (e.g., Essig et al., 2001; Lynch & McCaffrey, 2004; McSweeney et al., 1998).

Research in the area of third party observation first entered the neuropsychological literature when Binder and Johnson-Greene (1995) published a case study on a woman who demonstrated impaired performance on the Portland Digit Recognition Test, a test of effort, in the presence of her mother. Subsequent studies sought to examine the effects of a third party observer on performance on a variety of other tests assessing various neuropsychological domains. These studies have suggested that the presence of a third party observer may lead to impaired performance on tests of learning and memory, including number of perseverative errors on the Rey Auditory-Verbal Learning Test (Kehrer, Sanchez, Habif, Rosenbaum, & Townes, 2000) and number of words recalled at delay on the Verbal Paired Associates subtest from the Wechsler Memory Scale – Revised (Lynch, 2005). Impaired performance on tests of memory has also been found when the third party observer is an audiotape recorder (Constantinou, Ashendorf, & McCaffrey, 2002), a videotape recorder (Constantinou, Ashendorf, & McCaffrey, 2005), or a supervisor explicitly attending to the examiner rather than the examinee (Yantz & McCaffrey, 2005). Studies have also demonstrated impaired performance in the presence of a third party observer on tests of attention, sustained concentration, response inhibition, and verbal fluency, including digit span, the Paced Auditory Serial Addition Task, the Stroop color-word test, and the Controlled Oral Word Association Test (Kehrer et al., 2000). However, performance on the Trail Making Test (TMT) parts A and B, tests of attention and set-shifting, has not been found to be affected by the presence of a third party observer (Kehrer et al., 2000; Lynch, 2005), nor have tests of motor function including the Finger Tapping Test (FTT), Grooved Pegboard, and grip strength (Constantinou et al., 2005; Kehrer et al., 2000; Lynch, 2005). Interestingly, use of motor measures at the beginning of the testing session in attempt to facilitate adaptation to the testing process has been found to be effective only when a third party observer is not present, as unobserved examinees given an adaptation period performed better on a paired list learning task as compared with unobserved examinees not given an adaptation period and with observed examinees, regardless of whether an adaptation period was given (Gavett & McCaffrey, 2007).

Although the aforementioned studies have investigated the effects of a third party observer on neuropsychological test performance at the group level, no published studies in the neuropsychological literature have examined the contribution of individual examinee characteristics. One feature that can be expected to have a role in modulating the effects of social facilitation is the examinee’s anxiety. However, comparisons of performance of individuals high and low in anxiety under observed and unobserved conditions in the social psychology literature have been somewhat inconclusive. For example, while Ganzer (1968) found that more highly anxious individuals demonstrated impairments particularly during the initial and later stages of learning on a nonsense syllable list-learning task, Martens (1969) found that high anxiety participants learned a complex motor task more quickly than those participants low in trait anxiety. Conflicting findings such as these suggest that a number of other variables may also contribute to the effects that an individual’s anxiety will have on his or her performance, including whether a given test is timed. For instance, Siegman (1956) found that participants high in anxiety performed significantly worse on the timed subtests of the Wechsler Adult Intelligence Scale (WAIS) as compared with the untimed subtests, whereas performance on the two types of tests among low anxiety participants did not differ. In addition, in a test of experimenter- versus self-pacing, Mayer (1977) found that participants low in trait anxiety performed similarly in both conditions, while those high in trait anxiety performed significantly better when allowed to self-pace.

While no published studies in the neuropsychological literature have examined the contribution of anxiety to social facilitation or third party observer phenomena, a number of studies have looked directly at the influence of anxiety on neuropsychological test performance. For example, Bucklew and Hannay (1986) found that while performance on a variety of neuropsychological tests was not affected by trait anxiety, those participants high in state anxiety performed more poorly on a simple word fluency test and the block design subtest from the WAIS as compared with participants low in state anxiety. These two tests were rated as being significantly more difficult than the other tests administered, including the Digit Symbol subtest from the WAIS and the FTT, suggesting that high levels of state anxiety may be associated with poorer performance on difficult but not necessarily easy tests.

In another study investigating anxiety and neuropsychological test performance, King, Hannay, Masek, and Burns (1978) found that for women only, higher trait anxiety as assessed with the State–Trait Anxiety Inventory (STAI) was associated with poorer performance on both the FTT and the dominant hand and both hand subtests of the form board, a precursor to the Tactual Performance Test (TPT). Additionally, in an investigation of the effects of various personality traits on performance on 13 neuropsychological tests among 57 subjects with toxic encephalopathy and 57 healthy
referents, Persson, Österberg, Karlson, and Ørbæk (2000) found that within the healthy group, high trait anxiety was associated with worse performance on measures of visual reaction time, visual search, and response inhibition. When collapsed across subject groups, high trait anxiety was also associated with poorer performance on the Digit Symbol subtest of the WAIS-R. Trait anxiety was not associated with performance on measures of verbal fluency, general knowledge, spatial ability, or verbal memory. Interestingly, while healthy subjects low in trait anxiety demonstrated better performance than the toxic encephalopathy subjects on 8 of the 13 tasks, those healthy subjects high in trait anxiety exhibited superior performance only on a test of verbal memory. These results suggest that trait anxiety may have a clinically as well as statistically significant impairing effect on performance on certain neuropsychological tasks.

As previously discussed, many social facilitation studies have suggested that performance on neuropsychological tests may be impaired in the presence of many types of third party observers across a number of different domains. However, there has been limited research specifically examining the effects of a third party observer on performance on non-computerized tests of executive function. The accurate assessment of executive functioning capabilities is important in neuropsychological testing, especially given associations found between executive functioning and quality of life in some populations (e.g., Alptekin et al., 2005; Fujii, Wylie, & Nathan, 2004). Therefore, one goal of the present study was to investigate the effect of a neutral observer on performance of neuropsychological tests of executive functioning, including phonemic (letter) and semantic (category) verbal fluency tests, the TMT, and the TPT. Given previous research demonstrating adverse effects of observer presence on performance of complex or novel tasks, it was hypothesized that the presence of a third party observer would be associated with impaired performance on tests of executive functioning. More specifically, it was predicted that impairment due to observer presence would be considerable on the verbal fluency tests, in accordance with those findings of Kehrer et al. (2000) on the COWAT and of Buckelew and Hannay (1986) on the simple word fluency test. Performance differences on the TMT between observation groups, however, were predicted to be small or nonexistent, given null findings of both Kehrer et al. (2000) and Lynch (2005) on this test. Given the complexity of and need for cognitive flexibility required on the TPT, it was predicted that impairments on this test in the presence of an observer would be large.

Although previous research has suggested that anxiety may impair performance on various neuropsychological tests, interactions between anxiety and presence of a third party observer on neuropsychological tasks have not been investigated. Given previous research suggesting poorer performance of anxious individuals on complex non-motor tasks, it was hypothesized that there would be a main effect of anxiety on performance of tests of executive functioning such that high anxiety would be associated with poorer performance. However, given the simple nature of the TMT—part A, combined with findings that high anxiety is associated with greater speed (e.g., Leon & Revelle, 1985), it was predicted that performance on this test would be better among high anxiety individuals as compared with individuals low in anxiety. In addition, a significant observation condition by anxiety interaction was predicted such that the presence of a third party observer would be associated with greater decrements in performance for high state and trait anxiety individuals as compared with low anxiety subjects.

1. Methods

1.1. Participants

After gaining approval of the human subjects institutional review board, 79 college undergraduates were recruited from introductory psychology classes. Informed consent was obtained and subjects of each gender were randomly assigned to either the experimental or the control group, to ensure roughly equivalent ratios of males to females in each group. The experimental group was observed by a third party observer, while the control group was not observed during test administration. Data from five subjects in the experimental condition and four subjects in the control condition were discarded for the following reasons: two participants exhibited cheating behaviors on one of the tests, two withdrew before completing all tests, and two were under the obvious influence of intoxicating substances. Additionally, two cases were discarded due to missing data and one due to experimenter error. Thus, 70 subjects were included in the final analysis, 35 in each group.

Participants included 44 males and 26 females (comprising 63 and 37% of the sample, respectively), ranging in age from 18 to 38 (M = 19.56, S.D. = 3.08). There were no statistically significant differences between groups in terms of age or gender.
1.2. Measures

Each participant was administered five tests in the following order: (1) Fear of Negative Evaluation scale (FNE), (2) phonemic (letters F, A, and S) and semantic (animal naming) verbal fluency tests, (3) Trail Making Test (TMT), parts A and B (4) Tactual Performance Test (TPT), and (5) State–Trait Anxiety Inventory (STAI).

1.3. Procedures

The FNE, verbal fluency tests, TMT, TPT, and STAI were administered according to standard procedures (Benton, Hamsher, & Sivan, 1994; Reitan & Wolfson, 1993; Spielberger, 1983; Watson & Friend, 1969). The same examiner, who was blind to the hypotheses of the study, administered all tests while the same third party observer was present during testing for the experimental group. Both the examiner and observer were female.

In the experimental group, participants were informed that an observer would be in the room, although the reason for the observer’s presence was not divulged to either participants or the examiner. The observer sat approximately 1 m behind and to the left of the subject, facing the examiner. Although the observer took occasional notes during testing, she did not interrupt or directly interfere with the testing process, and took efforts to remain as unobtrusive as possible. The duration of the testing session for both groups was approximately 45 min.

1.4. Analyses

A 2 × 2 × 2 between-subjects multivariate analysis of variance (MANOVA) was performed to determine whether three independent variables (IVs: observation condition, state anxiety, and trait anxiety) were related to performance on tests of executive functioning, as measured by nine dependent variables (DVs): fluency-FAS, fluency-animals, TPT-localization, TPT-memory, TPT-dominant hand, TPT-nondominant hand, TPT-both hands, TMTB-time, and TMTA-time. Given the goal of investigating associations among performance, observation, and anxiety and not of evaluation apprehension per se, as well as in the interest of maintaining adequate power, FNE scores were omitted from the final analyses. This decision was reinforced by recent findings suggesting that the scale may lack sensitivity at higher levels of fear of negative evaluation (Rodebaugh et al., 2004). TPT-total time scores were also excluded so as to avoid singular correlations with other TPT time scores. Additionally, error scores on both parts of the TMT were excluded from the final analyses given likely ceiling effects revealed upon initial inspection and in the interest of preserving adequate power.

Median splits were performed on the state and trait anxiety inventory scores to divide subjects into high and low state (Mdn = 37.0) and trait (Mdn = 38.0) anxiety groups. The median state anxiety score for this sample was 0.5 points lower and the median trait anxiety score 3.0 points higher than those of the sample of 102 undergraduates used in Leon and Revelle (1985). Additionally, mean trait anxiety scores for the low and high trait anxiety groups were within a standard deviation of those reported for low and high trait anxiety, respectively, in Buckelew and Hannay (1986). These similarities suggest that the median split used resulted in an appropriate and normative classification of subjects into low and high anxiety groups.

2. Results

Table 1 lists the multivariate F values for the combined DVs for each of the IVs, as well as for the condition by state anxiety and condition by trait anxiety interactions. Additionally, univariate F values are listed for those individual DVs considered likely to reach significance in a univariate context. Roy–Bargmann stepdown F’s are also listed for the individual DVs, as the use of these may be warranted in order to reduce the effects of shared variance given the presence of correlations greater than .30 (in a positive or negative direction) between DVs. However, given the lack of adequate previous research on associations of observation and anxiety with these particular DVs and the consequent difficulty in predicting which variables are likely to be most influenced by the IVs, both univariate and stepdown F’s were examined. DVs were entered into the analysis in the above-listed order. Table 1 also lists eta squared (η²) values as indications of effect sizes for all multivariate main effects and interactions as well as for each statistically significant univariate comparison. Table 2 lists η² values for all univariate comparisons regardless of statistical significance. In accordance with convention, cutoffs of
Table 1
Tests of main effects and interactions of observation condition, state anxiety, and trait anxiety

<table>
<thead>
<tr>
<th>IV</th>
<th>DV</th>
<th>η²</th>
<th>Wilks' value (F)</th>
<th>Univariate F</th>
<th>Stepdown F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation condition</td>
<td>Fluency-animals</td>
<td>.28†††</td>
<td>.722 (2.31), p = .03*</td>
<td>8.10²</td>
<td>7.69</td>
</tr>
<tr>
<td>State anxiety</td>
<td>TMTA-time</td>
<td>.12†</td>
<td>.777 (1.72), p = .11</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>TMTB-time</td>
<td>.22†††</td>
<td>.771 (1.78), p = .09</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Condition by state anxiety</td>
<td>TPT-localization</td>
<td>.23†††</td>
<td>.792 (1.58), p = .15</td>
<td>11.79⁹</td>
<td>10.15*</td>
</tr>
<tr>
<td>Condition by trait anxiety</td>
<td>Fluency-animals</td>
<td>.16†††</td>
<td>.735 (2.16), p = .04*</td>
<td>9.48⁹</td>
<td>7.64</td>
</tr>
</tbody>
</table>

*Significant at the p ≤ .05 level, †small effect size, ††medium effect size, †††large effect size.

Significance level cannot be evaluated but would reach p < .05 in univariate context.

Significance level cannot be evaluated but would reach p < .01 in univariate context.

Using the Wilks’ criterion to evaluate the main effects of each of the three IVs, the combined DVs were found to be significantly affected by observation condition, with approximately 28% of the variance in performance accounted for by condition. Examination of univariate F’s reveals that performance on the fluency-animals test appeared to be significant at the p ≤ .05 level, with the application of a Bonferroni correction for a final p value of .006. The stepdown F approaches significance, suggesting that some of the variance shared with observation condition may be accounted for through overlapping variance with other DVs. However, these results still suggest that performance on the fluency-animals test was particularly influenced by the presence of a third party observer. Inspection of means and standard deviations (see Table 3) reveals that performance on this test was poorer in the observation condition as compared with the control condition.

Main effects of state anxiety and trait anxiety were not found to be significant. However, informal inspection of univariate and stepdown F’s for each DV reveals that performance on the TPT-localization subtest appeared to be significantly related to trait anxiety. Examination of group means for this subtest reveals superior performance for the high anxiety group (M = 5.34, S.D. = 2.13) as compared with the low anxiety group (M = 3.91, S.D. = 1.99). Inspection of η² values reveals that effect sizes for all multivariate main effects and for the effect of trait anxiety on TPT-localization score are large. A medium effect size was found for the association of observation condition with the fluency-animals test (see Table 1).

The Wilks’ criterion was also used to investigate separate interactions between observation condition and state and trait anxiety. Although the interaction between condition and state anxiety was not significant, performance on the combined DVs was found to be significantly affected by the interaction between observation condition and trait anxiety, with 27% of the variance in performance accounted for by the interaction. Examination of univariate F’s reveals that

Table 2
Effect sizes (η²) for univariate comparisons within each multivariate comparison

<table>
<thead>
<tr>
<th>DV</th>
<th>Observation condition</th>
<th>State anxiety</th>
<th>Trait anxiety</th>
<th>Condition by state anxiety</th>
<th>Condition by trait anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluency-FAS</td>
<td>.01†</td>
<td>.01†</td>
<td>.01†</td>
<td>.05†</td>
<td>.03†</td>
</tr>
<tr>
<td>Fluency-animals</td>
<td>.12††</td>
<td>.08††</td>
<td>.04†</td>
<td>.11††</td>
<td>.13††</td>
</tr>
<tr>
<td>TMTA-time</td>
<td>.00</td>
<td>.08††</td>
<td>.06††</td>
<td>.02†</td>
<td>.02†</td>
</tr>
<tr>
<td>TMTB-time</td>
<td>.00</td>
<td>.07††</td>
<td>.03†</td>
<td>.03†</td>
<td>.00</td>
</tr>
<tr>
<td>TPT-dominant</td>
<td>.00</td>
<td>.07††</td>
<td>.02†</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>TPT-nondominant</td>
<td>.00</td>
<td>.01†</td>
<td>.02†</td>
<td>.00</td>
<td>.01†</td>
</tr>
<tr>
<td>TPT-both</td>
<td>.00</td>
<td>.05†</td>
<td>.04†</td>
<td>.00</td>
<td>.01†</td>
</tr>
<tr>
<td>TPT-memory</td>
<td>.00</td>
<td>.00</td>
<td>.07††</td>
<td>.02†</td>
<td>.02†</td>
</tr>
<tr>
<td>TPT-localization</td>
<td>.08††</td>
<td>.02†</td>
<td>.16††</td>
<td>.00</td>
<td>.03†</td>
</tr>
</tbody>
</table>

†small effect size, ††medium effect size, †††large effect size.
Table 3
Group means and standard deviations for observation and trait anxiety (TA) groups

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th></th>
<th>Observed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low TA</td>
<td>High TA</td>
<td>Total</td>
<td>Low TA</td>
</tr>
<tr>
<td>Fluency-FAS: total words</td>
<td>37.21 (10.26)</td>
<td>40.00 (14.63)</td>
<td>38.49 (12.33)</td>
<td>36.56 (8.27)</td>
</tr>
<tr>
<td>Fluency-animals: total words</td>
<td>19.37 (5.36)</td>
<td>22.25 (6.98)</td>
<td>20.69 (6.23)</td>
<td>18.88 (3.54)</td>
</tr>
<tr>
<td>TMTA: time (s)</td>
<td>28.41 (11.74)</td>
<td>25.06 (4.14)</td>
<td>26.88 (9.13)</td>
<td>27.15 (7.87)</td>
</tr>
<tr>
<td>TMTA: errors*</td>
<td>.21 (.42)</td>
<td>.25 (.58)</td>
<td>.23 (.49)</td>
<td>.13 (.34)</td>
</tr>
<tr>
<td>TMTB: time (s)</td>
<td>55.03 (17.32)</td>
<td>54.66 (15.74)</td>
<td>54.86 (16.38)</td>
<td>56.39 (22.74)</td>
</tr>
<tr>
<td>TMTB: errors*</td>
<td>.37 (.60)</td>
<td>.31 (.60)</td>
<td>.34 (.59)</td>
<td>.19 (.40)</td>
</tr>
<tr>
<td>TPT-dominant: time (min)</td>
<td>5.78 (2.87)</td>
<td>5.44 (2.16)</td>
<td>5.63 (2.54)</td>
<td>5.83 (3.11)</td>
</tr>
<tr>
<td>TPT-nondominant: time (min)</td>
<td>3.87 (1.43)</td>
<td>4.29 (3.53)</td>
<td>4.06 (2.58)</td>
<td>4.86 (3.44)</td>
</tr>
<tr>
<td>TPT-both: time (min)</td>
<td>2.07 (1.08)</td>
<td>2.01 (.97)</td>
<td>2.04 (1.02)</td>
<td>2.22 (1.06)</td>
</tr>
<tr>
<td>TPT-total: time (min)*</td>
<td>11.72 (4.69)</td>
<td>11.74 (6.17)</td>
<td>11.73 (5.33)</td>
<td>12.91 (6.66)</td>
</tr>
<tr>
<td>TPT-memory: total shapes</td>
<td>7.53 (1.22)</td>
<td>8.00 (1.21)</td>
<td>7.74 (1.22)</td>
<td>7.06 (1.81)</td>
</tr>
<tr>
<td>TPT-localization: total shapes</td>
<td>4.79 (1.81)</td>
<td>5.44 (2.22)</td>
<td>5.09 (2.01)</td>
<td>2.88 (1.71)</td>
</tr>
</tbody>
</table>

* Variable not included in final analyses.

Performance on the fluency-animals test appeared to be significantly associated with the observation condition by trait anxiety interaction. The stepdown $F$ approaches significance, again suggesting that some of the variance shared with the condition by trait anxiety interaction may be accounted for through overlapping variance with other DVs. Further examination of the interaction suggests that the presence of a third party observer is associated with greater impairment on performance of the fluency-animals test among individuals high in trait anxiety as compared with those low in trait anxiety (see Fig. 1). Inspection of $\eta^2$ values reveals that effect sizes for both multivariate interactions are large. Additionally, a medium effect size of the condition by trait anxiety interaction on the fluency-animals test was found (see Table 1). Table 3 lists group means and standard deviations on each DV for low and high trait anxious groups within both observation conditions.

![Fig. 1. Observation condition by trait anxiety interaction on fluency-animals test.](image)
3. Discussion

The findings from this study suggest that the presence of a third party observer is associated with poorer performance on certain tests of executive functioning, thereby adding to the existing neuropsychological literature on the negative effects of third party observation on test performance. Performance on the fluency-animals test seems to be particularly impaired by observer presence, although the mean score for the observed group was within a standard deviation of that for age- and education-matched norms \((M = 19.8, \text{ S.D.} = 4.2; \text{Tombaugh, Kozak, & Rees, 1999})\).

Results from this study also imply that performance on tests of executive functioning may be influenced by trait anxiety. Although the multivariate analysis was not significant, univariate analysis suggested that low trait anxiety may be associated with worse performance on the TMT-localization subtest. Inspection of norms for TPT performance reveals that subjects classified in the present study as low in trait anxiety performed more than a standard deviation below that of age- and education-appropriate norms \((M = 6.47, \text{ S.D.} = 2.44; \text{Yeudall, Reddon, Gill, & Stefanyk, 1987})\). This finding is somewhat surprising given the difficult nature of the task and the more common finding that high anxiety is associated with worse performance on complex or novel tasks. However, this analysis did not account for third party observer effects.

Another aim of the present study was to examine interactions among observation condition and state and trait anxiety. Although there were no significant findings associated with state anxiety, the multivariate observation condition by trait anxiety interaction was found to be significant, suggesting that the presence of a third party observer differentially affects performance among individuals high and low in trait anxiety. Results additionally suggested that this disparity is particularly pronounced for performance on the fluency-animals test. Although group means for high and low trait anxiety subjects in the control and observed conditions were all within a standard deviation of age- and education-appropriate norms \((\text{Tombaugh et al., 1999})\), high anxiety subjects in the control condition performed over a half-standard deviation above the normed mean while those in the observed condition performed over a half-standard deviation below this mean. In contrast, low anxiety subjects performed within a quarter-standard deviation below the normed mean in both observation conditions. These results suggest that while high trait anxiety may have a facilitating effect on semantic fluency in less stressful situations, individuals high in trait anxiety may be especially vulnerable to the potential negative impact of an added stressor such as a third party observer.

Interestingly, investigation of group means on the localization subtest of the TPT, performance on which was also shown to be influenced by trait anxiety, reveals an opposite pattern. While performance among subjects high in trait anxiety was within a half-standard deviation below an age- and education-appropriate normed mean \((\text{Yeudall et al., 1987})\) in each condition, performance among subjects low in trait anxiety differed to a greater degree between observation groups. Whereas the mean performance of low trait anxiety subjects in the control condition was within a standard deviation of the normed mean, performance in the observed condition was more than a standard deviation below the appropriate normed mean. Additionally, in examining how these scores would be interpreted in computing the general neuropsychological deficit scale (NDS) score of the Halstead-Reitan Neuropsychological Test Battery \((\text{Reitan & Wolfson, 1993})\), individuals scoring within a standard deviation of the low anxiety control group mean on the TPT-localization subtest would be considered to be performing either within normal limits (NDS = 0 or 1) or in the mild-to-moderate deficit range (NDS = 2). Those individuals performing within a standard deviation of the low anxiety observed group mean on this test, however, would be considered to be performing either in the mild-to-moderate (NDS = 2) or severe (NDS = 3) deficit range. Thus, although the univariate test for the observation condition by trait anxiety interaction on the TPT-localization subtest was not statistically significant, examination of appropriate norms and scoring conventions suggests a clinically significant interaction such that individuals low in trait anxiety may be more susceptible to the possible adverse impact of observer presence on performance of certain tests such as those assessing recall for spatial location.

Examination of performance patterns among subjects in each observation condition is warranted in order to elucidate potential mechanisms surrounding social facilitation and anxiety effects on tests of executive functioning. As hypothesized, performance on a verbal fluency test and on a subtest of the TPT were associated with significant third party observer or observer by anxiety effects. Additionally, in accordance with findings of \(\text{Kehrer et al. (2000)}\) and \(\text{Lynch (2005)}\), performance on the TMT did not differ between observation groups. Scores on the TMT were also similar among high and low anxiety subjects. With respect to the nature of the associations among anxiety, observation, and performance, these may have varied in part due to the different nature of the tests used in the present study. For example, high trait anxiety was associated with a greater negative impact of observer presence on performance of the
fluency-animals test, as compared with low trait anxiety. However, this relationship was not found for the localization subtest of the TPT. Given the oral nature of responding on the fluency-animals test, it is possible that evaluation apprehension and overall arousal rose to an impairing level among individuals already prone to anxiety when an observer was present. Additionally, spontaneous cognitive flexibility may be particularly vulnerable to the potentially anxiety-inducing and impairing effects of a third party observer. The timed nature of the fluency-animals test likely serves as an additional source of anxiety or arousal; as discussed previously, in comparison with untimed tasks, timed tests have been associated with worse performance among anxious individuals (Siegman, 1956). On the TPT-localization subtest, however, responses cannot be readily observed by an individual sitting 1 m behind the examinee. Furthermore, this test is not timed.

While the present study provides further support for the association between observer presence and poorer neuropsychological test performance, in addition to shedding light on interactions of observation with examinee anxiety, certain limitations of the study should be noted. First, self-rating scales were used to determine subjects’ levels of anxiety, and given that self-report scales are often prone to social desirability response sets, confounds may have been present in examining interactions between anxiety and social facilitation effects. An additional limitation of this study is that all subjects were college undergraduates; this homogeneity among participants limits the generalizability of the findings of this study. Future research on the effects of observer presence and anxiety on neuropsychological test performance in other populations is warranted.

Despite these limitations, the present study contributes to the existing literature suggesting that the validity of neuropsychological test results obtained while a third party observer is present is significantly compromised. Furthermore, these findings have added to the previous research by suggesting that impairments in performance that result from the presence of a third party observer occur not only in tests of effort, attention, concentration, learning, and memory, but in tests of executive functioning as well. The statistically significant finding that performance on tests of verbal fluency may be negatively impacted by the presence of a third party observer replicated that of Kehrer et al. (2000). Additionally, this study is the first known to demonstrate an effect of observer presence on performance of a subtest of the TPT. The medium and large effect sizes found suggest that these associations between test performance and observer presence are clinically as well as statistically significant. Findings from the present study also provide evidence that examinee trait anxiety interacts with third party observation to yield different performance patterns among individuals high and low in anxiety.

Acknowledgements

The authors would like to thank Suzanne Milnes, M.A. for her assistance in data collection as well as Julie K. Lynch, Ph.D. and Christine L. Yantz, Ph.D. for their careful reading and editing.

References


