Trait and state predictors of error detection accuracy in a simulated quality control task

Kraig L. Schell *, Abbie Woodruff, G. Brandon Corbin, Ellen C. Melton

Department of Psychology, Angelo State University, P.O. Box 10907, ASU Station, San Angelo, TX 76909, United States

Received 14 April 2004; received in revised form 11 October 2004; accepted 7 December 2004
Available online 26 January 2005

Abstract

This study examined whether personality traits or mood and workload states are more predictive of detection accuracy in a simulated quality control task mimicking prescription checking in a pharmacy. Ninety-one undergraduate volunteers checked 80 simulated scripts for accuracy in a basic task environment with error rates ranging from 26% to 38%. Four dimensions of perceived workload and three task-related mood factors were assessed multiple times during the task, and several theoretically relevant personality traits were measured before the simulation. Performance was measured using hit rate and false alarm ratios. Results suggested that trait factors were more predictive of hit ratios in this study than were measures of task-related mood and perceived workload. Implications and future research directions are discussed.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Pharmacy; Quality control; Traits; Performance; Error detection; States

1. Introduction

The state versus trait issue is an enduring one in psychology, paralleling in many ways basic arguments such as the nature/nurture debate. In the context of quality control, the relative influence of traits and states on human performance could lead to better interventions designed to
improve it (Grasha & O’Neill, 1996, 1998). This could be useful in the pharmacy, where a misfilled prescription could cause severe injury to a patient. However, pharmacy has been slow to recognize the importance of the pharmacist and/or technician in the error control process, viewing them implicitly as automatons and focusing attention instead on technological and workflow issues (Cohen, 1999). In response to this, we have studied quality control tasks in both experimental and field settings, outlining some potential factors that are relevant to detecting errors in a real-world detection task (Grasha, Reilley, Schell, Tranum, & Filburn, 2000; Schell, Reilley, Grasha, & Tranum, 2003; Schell, Woodruff, Corbin, Bilsing, & Melton, 2002). It is an explicit goal of our research to understand how individual differences affect pharmacy error control so that improvement methods can be designed with those differences in mind.

In this research, we examined the relative contribution of task-related state factors and personality trait measures to error detection in a simulated prescription checking task. Our simulation task has consistently demonstrated error rates and work speeds commensurate with those observed in pharmacies, despite using pharmacy-naïve individuals in our studies (Grasha & Schell, 1999; Grasha et al., 2000). In the simulation, participants verify for correctness multiple sets of “scripts” and input their findings into a customized computer database. Since the scripts represent common items such as beads and paper clips and no specialized medical knowledge is required, these participants can learn the task quickly and perform effectively in most cases.

State and trait factors were included in this study based on previous literature suggesting a relationship to performance on these kinds of tasks. These factors are listed in Table 1. First, measures of anxiety such as trait anxiety, Neuroticism, social anxiety and inhibition motivation have been related to performance in the simulated pharmacy environment (Grasha & Schell, 1999, 2001; Schell & Grasha, 2000; Schell et al., 2002), justifying further study. Second, research shows that Type A tendencies were positively related to error detection accuracy (Grasha & Schell, 1999; Schell et al., 2002), even though the structure of the task is not the stereotypical one that such a person might choose. Third, extraversion was included because previous research

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Variables included in the proposed trait and state models of quality control performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dependent measure: overall hit ratio</td>
</tr>
<tr>
<td></td>
<td>Model T—Trait model</td>
</tr>
<tr>
<td>Predictors</td>
<td>Predictors (workload)</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>Mental demand</td>
</tr>
<tr>
<td>Inhibition motivation</td>
<td>Temporal demand</td>
</tr>
<tr>
<td>Social anxiety</td>
<td>Effort</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>Frustration</td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
</tr>
<tr>
<td>Type A Trait</td>
<td>Predictors (mood)</td>
</tr>
<tr>
<td>Locus of control</td>
<td>Engagement</td>
</tr>
<tr>
<td></td>
<td>Predictors (mood)</td>
</tr>
<tr>
<td></td>
<td>Tension</td>
</tr>
<tr>
<td></td>
<td>Hedonic tone</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
</tr>
<tr>
<td></td>
<td>Task-related motivation</td>
</tr>
</tbody>
</table>

*Note: Participants’ false alarm ratios will be used in both models as a predictor.*
has linked it to work performance in other domains (Matthews & Amelang, 1993; Matthews, Davies, & Holley, 1990). Finally, locus of control was included since control beliefs are important in moderating anxiety and increasing motivation (e.g., Shapiro, Schwartz, & Astin, 1996) and anxiety and motivation have been implicated as predictors of performance in our task environment before (Schell & Grasha, 2000; Schell et al., 2002).

Task-related mood and perceived workload are important state factors because of their extensive use in the performance literature (Braarud, 2001; Hancock & Meshkati, 1988; Matthews, Davies, Westerman, & Stammers, 2000; Nygren, 1991). To capture changes in these states over time, measurements were taken on multiple occasions as participants worked. Measures of energy, tense arousal and hedonic tone were measured as task-related mood constructs. These factors have been related to performance in a number of task situations, including driving (Matthews & Desmond, 2001), mail sorting (Matthews, Jones, & Chamberlain, 1992), and our simulation (Tranum, Matthews, & Grasha, 2001). Perceived workload has also been connected to human performance in many ways (Becker, Warm, Dember, & Hancock, 1995; Bunce & Sisa, 2002; Carayon-Sainfort, 1992; Prinzel & Freeman, 1997; Smith & Bourke, 1992). Specific dimensions of perceived workload also have been connected to accurate error detection using our simulation (Grasha & Schell, 2001; Reilley, Grasha, & Schafer, 2002; Schell & Grasha, 2000), supporting their inclusion.

The measurement of performance in the pharmacy simulation task draws from signal detection theory (e.g., Davies & Parasuraman, 1982), since the basic skill is the detection and overt identification of errors (targets) amongst correct orders (noise). We used hit ratios as our measure of accuracy, and false alarm ratios as our measure of response bias. Further, we used false alarm ratios as predictors of hit ratios in this study instead of calculating d-prime and beta statistics for three reasons. First, the nature of the problem in this study (medication errors) demanded that hit ratios were more important than false alarm ratios, due to differences in their potential real-world consequences. Second, we were interested more in group-level analysis of performance ability on this task rather than any comparison of individuals. That is, our intent was not to determine between two people who was a better performer of the task, but instead to understand how certain factors were generally related to human performance in this domain. Third, assumptions of d-prime include normally distributed signal and noise distributions as well as consistent signal strengths for targets and non-targets (Davies & Parasuraman, 1982; Wickens, 1992). Because the stimuli in this task were so complex, signal strength was impossible to hold constant, and so the signal and noise distributions were probably not normal. So, d-prime and beta, although technically calculable in this study, were not used.

Our predictions were based on previous work employing the quality control simulation and theoretical arguments concerning the connection between personality traits, states and performance in work contexts (Barrick & Mount, 1991; Johnson, Schneider, & Oswald, 1997; Mount & Barrick, 1998). Our work with this particular task suggests that personality factors such as trait anxiety, extraversion and conscientiousness are related to performance when objective task demands (time pressure, etc.) are relatively low (Schell & Reilley, 2004; Schell et al., 2002). But, meta-analyses have found more consistent relationships between certain traits (i.e., conscientiousness) and rated work performance (Barrick & Mount, 1991; Conte & Jacobs, 2003; Tett, Jackson, & Rothstein, 1991). It is not clear why these relationships are not stronger when performance is measured objectively rather than subjectively. It has been suggested that accounting for the characteristics of
the task being studied may help answer this question (Hockey, 1984), and so we feel that personality must be studied in specific task contexts to understand how it relates to objective performance measures. Furthermore, state variables should affect performance since, for example, stress and anxiety can lead to cognitive ruminations and other distractions which can then affect performance either positively or negatively (Sarason, Sarason, & Pierce, 1990; Zeidner, 1998). So, it is not clear whether state or trait factors should be more influential on performance in a quality control task.

We expected that trait factors would account for more variance in hit ratios than would state factors for two reasons. In previous studies using the simulation task, traits have shown stronger correlations with performance than states (e.g., Grasha et al., 2000; Grasha & Schell, 1999; Schell & Grasha, 2000). However, much of this research involved a prescription dispensing task instead of a checking task. Since the dispensing task clearly requires different skills and behaviors than the checking task used in this study, we needed to determine if traits were more predictive of pharmacy performance in general or if their predictive utility was restricted to dispensing only. Second, the stimuli in a prescription checking task can be considered a variably-mapped set (Schneider & Shiffrin, 1977), which forces a controlled processing mode for a longer period of time (Ackerman, 1987, 1988). Research has shown that abilities and personality traits are more predictive of performance during controlled processing, especially in early task performance (Ackerman & Kanfer, 1993; Ackerman, Kanfer, & Goff, 1995). Thus, traits were expected to show stronger relationships to performance in our task.

To summarize, we expected that trait factors would be more predictive of performance on the task than were state measures. In other words, it was expected that trait-based models would account for significantly more variance in hit ratios than would similarly constructed state-based models. Because the state variables (i.e., mood and workload) were measured both at mid-task and at post-task points in time, separate models were constructed and tested for each measurement point.

2. Method

2.1. Participants

Ninety-one students enrolled in psychology courses at Angelo State University volunteered for the research in exchange for course credit. In the sample, there were 27 men and 64 women with a mean age of 20 (SD = 4).

2.2. Instruments

Avoidant motivational tendencies were measured using a subset of a self-report survey (Carver & White, 1994) composed of seven items scaled from 1 (very true for me) to 4 (very false for me). Trait anxiety was assessed using the trait section of the State–Trait Anxiety Instrument (Spielberger, 1983), consisting of 20 items scaled from 1 (almost never) to 4 (almost always). Extraversion and Neuroticism were assessed using the Eysenck Personality Inventory (Eysenck, 1981), consisting of 57 questions to which respondents answered “yes” or “no”. Finally, Type A tendencies and social anxiety were assessed using the Grasha Holistic Stress Test (Grasha, 1990). The Type A and
the social anxiety scales consisted of 22 questions and 10 questions, respectively, both scored on a 5-point scale from 1 (disagree) to 5 (agree).

The Dundee State Stress Questionnaire (DSSQ: Matthews, Jones, & Chamberlain, 1990, 1992) was used to measure task-related mood, motivation and cognitions associated with stress, arousal and fatigue. The mood section of the DSSQ consisted of 29 items scored from 1 (definitely) to 4 (definitely not). Respondents were presented with adjectives and asked whether they experienced those feelings as they worked on the task. Energization accounted for 8 of the 29 items, the tension mood scale of the DSSQ accounted for 8 of the 29 items, and the hedonic tone scale accounted for 8 of the 29 items. All of these scales had minimum and maximum ranges of 8–32. Finally, the anger scale of the DSSQ accounted for the last five items, with minimum and maximum scores of 5–20. A separate section of the DSSQ measured task-related motivation on eight items scored from 0 to 9. Because of reverse scoring procedures, each question contained different verbal anchors. This scale carried a potential range of 0–72, with low scores indicating lower motivation. The DSSQ has been used to study a variety of behaviors and has been shown to be valid and reliable across work domains (Matthews et al., 1990; Matthews & Desmond, 1998; Matthews et al., 1992).

A modified version of the NASA Task Load Index (TLX: Hart & Staveland, 1988) was used to measure perceived workload during the task. Six dimensions of workload are measured on the NASA Task Load Index: mental demand, physical demand, temporal demand, perception of performance, effort, and frustration, of which four will be used in this study. The TLX is perhaps the most respected measure of perceived workload currently in use (Nygren, 1991), but because of psychometric problems with the recommended weighting procedure used to obtain a global workload score (Dickinson, Byblow, & Ryan, 1993), we treated the six dimensions of workload as separate variables.

2.3. Simulation environment

Two computer workstations were located in a medium sized room. Each workstation was composed of a table, chair and computer with a keyboard and mouse. A small digital clock was positioned so participants would be aware of the time of day. Lighting, temperature and noise levels were held constant within comfort regions. The workstations faced away from one another to avoid unintentional eye contact by participants.

Products consisted of a prescription bottle containing an amount of simulated drug materials and an order card, individually packaged in plastic bags. Simulated drugs were made out of beads, nuts, washer, and paper clips of several different colors, sizes and/or shapes, half of which were marked to simulate sound-alike/look-alike drug errors. Each item class was divided into two sets, half of which were marked with a black permanent marker and their names slightly altered to indicate their designation. For example, “BEADS-D-GL” indicated a gold colored, donut-shaped bead, while “BEEDS-D-GL” was the same bead with a black mark. Additional distinctions were made using simulated dosages such as “1.0” and “10.0” for items where there was more than one size, such as the nuts and washers. More information on the items has been included in earlier reports (see Grasha & Schell, 2001; Schell et al., 2003).

All order cards contained the following information: customer name, address, product name, product quantity, special instructions (a random phrase that was included to enhance realism,
such as “Re-order every 90 days”, etc.), telephone number and an order number. The order cards were handwritten and made of standard card stock paper. The prescription labels used for comparison were typed in 12-point Times Roman font and contained the same basic information. They were pasted on a separate sheet of paper for each order and placed in sequence in a small desk calendar apparatus. The orders were grouped in eight bundles of 5, totaling 40 orders in each of two sets. Thus, participants could verify a maximum of 80 orders during the simulation.

There were three possible types of errors: errors in information (spelling, omission or inclusion), product mismatches, or item counts. It was possible to have more than one error on a particular order. Two different sets of orders were used, each with their own inserted errors with error ratios ranging from 26% to 38%. However, a constant ratio of 2:1 was maintained between errors of incorrect information and product mismatch and counting errors in all sets, since the former type of error is more common in actual pharmacies (Allan, Barker, Malloy, & Heller, 1995; Flynn & Barker, 1999).

2.4. Procedure

Training consisted of an introduction to task materials, an overview of the task, and 10 supervised “practice trials”. A “trial” involved the following steps: an order was selected and three comparisons were performed. The first comparison involved the order card and the printed label, the second comparison involved the order card and the item inside the bottle, and the last comparison involved the order card and the amount of the item in the bottle. On their computer workstation, participants answered three questions about each order, indicating whether (a) the item in the bottle was the correct item, (b) the amount of the item was correct, and (c) the information on the order card matched the information on the printed label. Next, the trial was completed by entering the order number into the computer. Each set of orders contained two blocks of 40 orders each, which the participants were given 80 min to complete (40 min per block). This time limit was chosen because it has been shown to be a comfortable amount of time for most participants to complete the set (Reilley et al., 2002). The DSSQ was administered at three points: before the simulation, after the first set of 40 orders, and after the second set of 40 orders. The TLX was administered following each 40-order set. After all orders had been checked or time had expired, participants were debriefed and released.

3. Results

The data were first analyzed descriptively in order to examine summary statistics over time for the trait and state predictors. Means and standard deviations for these factors and for overall hit ratio are shown in Table 2. Standard deviations for hit ratios were much higher than for false alarm ratios. Also, the mean hit ratio of 65% is not surprising given the complexity of the stimuli being judged, i.e., the potential presence of more than one kind of error. Mean scores on the DSSQ and TLX scales were within expected ranges for a task of this type based on norms available in Matthews et al. (1990). Since our intent was not to induce stress in these participants, there was no need for a manipulation check or a comparison with means observed in other domains. Also, previous research has shown the pharmacy task to be moderate in workload as compared
to a simple card sorting task (Grasha & Schell, 1999), and these data are in line with that observation also.

Correlations among trait and state predictors are shown in Tables 3 and 4, and correlations between hit ratios and these predictors are shown in Table 5. There were some significant relationships among the predictors, which was expected given the relatedness of the constructs. It was noted that Frustration was the workload dimension most consistently correlated with DSSQ subscales. Also, correlations among DSSQ and TLX factors decreased after the task when compared to mid-task statistics. Finally, though the tables do not show trait–state correlations, we can report that in almost all cases, traits and states were not significantly correlated. The one exception was post-task temporal demand, which was correlated significantly with the anxiety trait measures (trait and social anxiety, inhibition motivation and neuroticism).

Finally, changes in DSSQ and TLX scores over time were assessed using one-factor repeated-measures ANOVAs. These analyses revealed that participants reported less motivation following the task than they reported in the middle of the task (F(1,81) = 28.37, p < .001), but no other DSSQ score changed significantly during the task. For TLX scores, however, mental demand, effort and frustration all increased significantly comparing mid-task to post-task scores (all p’s < .05).

To examine the hypothesis in this study, data were checked for regression assumptions of normality using one-sample Kolmogorov–Smirnov tests. These tests were not significant in all cases, even though false alarm ratios were somewhat negatively skewed. Thus, regression equations were constructed using the following steps. First, bivariate correlations were calculated between all predictors and the overall hit ratio criterion; significant correlations between proposed predictors and

Table 2
Means and standard deviations for overall hit rate and false alarm rate

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall hit rate</td>
<td>65.17</td>
<td>22.41</td>
</tr>
<tr>
<td>Overall false alarm rate</td>
<td>92.48</td>
<td>6.82</td>
</tr>
</tbody>
</table>

*Note:* False alarm rates are calculated as the percentage of correct orders judged as correct. \( r \) (hit rate, false alarm rate) = \(-.284\), \( p < .02 \).

Finally, changes in DSSQ and TLX scores over time were assessed using one-factor repeated-measures ANOVAs. These analyses revealed that participants reported less motivation following the task than they reported in the middle of the task (\( F(1,81) = 28.37 \), \( p < .001 \)), but no other DSSQ score changed significantly during the task. For TLX scores, however, mental demand, effort and frustration all increased significantly comparing mid-task to post-task scores (all \( p \)'s < .05).

To examine the hypothesis in this study, data were checked for regression assumptions of normality using one-sample Kolmogorov–Smirnov tests. These tests were not significant in all cases, even though false alarm ratios were somewhat negatively skewed. Thus, regression equations were constructed using the following steps. First, bivariate correlations were calculated between all predictors and the overall hit ratio criterion; significant correlations between proposed predictors and

Table 3
Correlations among trait predictors of hit rate

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trait anxiety</td>
<td></td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Inhibition motive</td>
<td>52</td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Social stress</td>
<td>36</td>
<td></td>
<td></td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Type A behavior</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td>07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Locus of control</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>6. Extraversion</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Neuroticism</td>
<td></td>
<td>55</td>
<td></td>
<td>49</td>
<td></td>
<td></td>
<td>42</td>
</tr>
</tbody>
</table>

*Note:* Decimals removed for readability. Bolded correlations are significant at \( p = .01 \).
Table 4
Correlations among state predictors of hit rate

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mental demand—Mid-task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Temporal demand—Mid-task</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Effort—Mid-task</td>
<td>47</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Frustration—Mid-task</td>
<td>05</td>
<td>35</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Mental demand—Post-task</td>
<td>82</td>
<td>35</td>
<td>52</td>
<td>08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Temporal demand—Post-task</td>
<td>33</td>
<td>64</td>
<td>23</td>
<td>27</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Effort—Post-task</td>
<td>55</td>
<td>42</td>
<td>47</td>
<td>15</td>
<td>59</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Frustration—Post-task</td>
<td>14</td>
<td>26</td>
<td>19</td>
<td>70</td>
<td>16</td>
<td>29</td>
<td>07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Tension—Mid-task</td>
<td>15</td>
<td>35</td>
<td>14</td>
<td>47</td>
<td>12</td>
<td>25</td>
<td>20</td>
<td>43</td>
<td>-16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Hedonic tone—Mid-task</td>
<td>-01</td>
<td>-13</td>
<td>04</td>
<td>-55</td>
<td>03</td>
<td>-09</td>
<td>-03</td>
<td>-54</td>
<td>21</td>
<td>-40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Anger—Mid-task</td>
<td>09</td>
<td>26</td>
<td>-06</td>
<td>61</td>
<td>03</td>
<td>23</td>
<td>00</td>
<td>63</td>
<td>-50</td>
<td>48</td>
<td>-61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Motivation—Mid-task</td>
<td>21</td>
<td>13</td>
<td>42</td>
<td>-32</td>
<td>34</td>
<td>03</td>
<td>35</td>
<td>-28</td>
<td>41</td>
<td>-13</td>
<td>56</td>
<td>-51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Engagement—Post-task</td>
<td>02</td>
<td>-13</td>
<td>09</td>
<td>-30</td>
<td>03</td>
<td>-17</td>
<td>10</td>
<td>-39</td>
<td>26</td>
<td>-08</td>
<td>21</td>
<td>-33</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Tension—Post-task</td>
<td>08</td>
<td>21</td>
<td>08</td>
<td>07</td>
<td>08</td>
<td>14</td>
<td>19</td>
<td>00</td>
<td>33</td>
<td>-16</td>
<td>19</td>
<td>-04</td>
<td>-45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Hedonic tone—Post-task</td>
<td>-02</td>
<td>-10</td>
<td>-03</td>
<td>-29</td>
<td>-08</td>
<td>-18</td>
<td>01</td>
<td>-41</td>
<td>18</td>
<td>-21</td>
<td>37</td>
<td>-37</td>
<td>17</td>
<td>71</td>
<td>-53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Anger—Post-task</td>
<td>-05</td>
<td>12</td>
<td>00</td>
<td>16</td>
<td>-07</td>
<td>07</td>
<td>-08</td>
<td>22</td>
<td>-09</td>
<td>19</td>
<td>-24</td>
<td>35</td>
<td>-24</td>
<td>-48</td>
<td>56</td>
<td>-73</td>
<td></td>
</tr>
<tr>
<td>18. Motivation—Post-task</td>
<td>05</td>
<td>17</td>
<td>23</td>
<td>-12</td>
<td>19</td>
<td>16</td>
<td>32</td>
<td>-23</td>
<td>26</td>
<td>-08</td>
<td>11</td>
<td>-21</td>
<td>45</td>
<td>46</td>
<td>-13</td>
<td>31</td>
<td>-24</td>
</tr>
</tbody>
</table>

*Note:* Bolded correlations are significant at $p = .01$. Decimals removed for readability.
the dependent measure were identified. Next, exploratory hierarchical models were built using the residual correlation method, where the strongest bivariate predictor was entered into the model and then other predictors were correlated with the residuals of that model. The building process stopped when no other predictors correlated with the residuals of the final model. Relevant model statistics are displayed in Table 6.

First, the set of trait predictors was regressed on the overall hit ratio criterion. From the correlation matrix, we selected social stress ($r(72) = .31$) to be entered first, followed by overall false alarm ratios ($r(78) = -.28$) and Type A tendencies ($r(72) = .27$). Further analysis of regression

Table 5
Correlations between hit rate and the trait and state predictors

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Pre-task</th>
<th>Mid-task</th>
<th>Post-task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait anxiety</td>
<td>.12</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Inhibition motivation</td>
<td>.03</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Social anxiety</td>
<td>.31*</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>.11</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Extraversion</td>
<td>- .18</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Type A Trait</td>
<td>.27**</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Locus of control</td>
<td>.15</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Mental demand</td>
<td>n/a</td>
<td>.02</td>
<td>-.04</td>
</tr>
<tr>
<td>Temporal demand</td>
<td>n/a</td>
<td>.03</td>
<td>-.02</td>
</tr>
<tr>
<td>Effort</td>
<td>n/a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frustration</td>
<td>n/a</td>
<td>-.12</td>
<td>.02</td>
</tr>
<tr>
<td>Energy</td>
<td>.12</td>
<td>.13</td>
<td>-.01</td>
</tr>
<tr>
<td>Tension</td>
<td>.03</td>
<td>-.21</td>
<td>-.06</td>
</tr>
<tr>
<td>Hedonic tone</td>
<td>.06</td>
<td>-.17</td>
<td>-.05</td>
</tr>
<tr>
<td>Anger</td>
<td>-.07</td>
<td>-.25**</td>
<td>.09</td>
</tr>
<tr>
<td>Motivation</td>
<td>.30*</td>
<td>.30*</td>
<td>.03</td>
</tr>
</tbody>
</table>

Note: n/a = Variable not measured at this time.
* $p < .01$.
** $p < .05$.

Table 6
Regression equations for overall hit rates as predicted by trait and state factors

<table>
<thead>
<tr>
<th>Predictors</th>
<th>$R^2$</th>
<th>$R^2$Δ</th>
<th>Std. beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model T—Trait Factors ($R = .55$, $R^2 = .31$, $R^2$, adj. = .26 $F(4,64) = 7.07, p &lt; .01$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social stress</td>
<td>.11</td>
<td></td>
<td>.26</td>
</tr>
<tr>
<td>Overall false alarm rate</td>
<td>.08</td>
<td></td>
<td>-.30</td>
</tr>
<tr>
<td>Type A Trait</td>
<td>.06</td>
<td></td>
<td>.25</td>
</tr>
<tr>
<td>Locus of control a</td>
<td>.06</td>
<td></td>
<td>.25</td>
</tr>
<tr>
<td>Model S—Mid-task state factors ($R = .40$, $R^2 = .16$, $R^2$, adj. = .14, $F(2,70) = 6.76, p &lt; .01$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>.09</td>
<td></td>
<td>.29</td>
</tr>
<tr>
<td>Overall false alarm rate</td>
<td>.07</td>
<td></td>
<td>-.27</td>
</tr>
</tbody>
</table>

a High LoC scores indicate internal control beliefs. All predictors univariately significant at least at $p < .05$. 
residuals suggested that locus of control should also be included, even though it was not bivariately related to the overall hit ratio. Each of these four predictors was associated with a significant change in $R^2$ and produced the final model shown in Table 6. This model explains an adequate proportion of the variance in overall hit ratio with an acceptable amount of shrinkage for an exploratory model ($R^2 = .31$, adjusted $R^2 = .26$).

Next, sets of mid-task and post-task state predictors were regressed on the hit ratio criterion. For the mid-task analysis, we selected mid-task motivation ($r(73) = .30$) to be entered first. The only other variable included in the model shown in Table 4 was overall false alarm ratio—anger and effort, even though they were correlated bivariately with hit ratio, did not significantly increase the variance accounted for in the model when they were added. $R^2$ for this model was lower than for the trait model ($R^2 = .16$, adjusted $R^2 = .14$). Also, when post-task state measures were used in the model-building process, a regression model could not be calculated since no post-task state variables were significantly related to overall hit rate.

### 4. Discussion

The purpose of this research was to explore whether traits or states are more predictive of error detection in a quality control task. These data suggest that traits are more predictive of hit rates in a quality control task than are state measures such as mood and workload. This is evidenced by larger $R^2$ values for the trait model versus the state model, as well as the larger number of traits that are independently predictive of hit ratios.

Examining the predictors in the trait model (see Table 5), anxiety-related constructs are well-represented. This is consistent with previous research linking anxiety to task performance (Anton & Klisch, 1995; Mahar, Henderson, & Deane, 1997). Social anxiety showed the strongest relationship to hit ratios, a result seen before using our simulation (Grasha et al., 2000). It may seem to some that social anxiety should not be relevant to performance on this task because it was performed alone without performance feedback. Participants may have believed some sort of evaluation of them would be done later, which would connect these results with a large literature on evaluation anxiety (see Sarason et al., 1990, for example). In the pharmacy, social anxiety can come from a number of sources: customers, co-workers, supervisors, etc. The identification of social anxiety as a predictor of error detection suggests that it should be studied further for its impact on pharmacy performance.

The positive partial correlations between social anxiety and performance (as well as between Type A tendencies and performance) could mean that participants thought the task was simple, so that anxiety was facilitative (Zeidner, 1998). However, hit ratios and false alarm ratios were negatively associated, suggesting a liberal criterion shift. Workload signatures on the task are comparable to but slightly lower than those obtained in field studies of pharmacists (Grasha et al., 2000), which would also support the liberal decision criterion explanation. In addition, the regressions in Table 6 show that false alarm ratios explained a significant amount of the variance in hit ratios in both models. This suggests that a criterion shift may explain these data better than an anxiety-facilitation argument. A study that induces anxiety before task performance would add a needed perspective to this question.
With respect to the state model, task-related motivation is an interesting factor. The motivation scale used in this research taps into what Ackerman and Kanfer (1993) call “motivational thoughts”. Questions assessed, for instance, whether the participant thought the task was interesting, how much they wanted to succeed, and how worthwhile they thought the task was. Ackerman and Kanfer (1993) reported that both positive and negative motivational thoughts were related to performance in complex tasks, and the data in this study are consistent with this conclusion as far as positive thoughts go.

It is intriguing that some participants reported the task to be interesting and worthwhile, even though the task seems to be rather boring and trite. It is possible that motivational thoughts were more positive because participants understood the societal importance of the topic under study. Clearly, reducing medication errors is a worthwhile cause, and participants’ motivation may reveal a belief that their time was being used for a “greater good”. It is also possible that motivated individuals “manufactured” interest to make up for the task’s monotony. This would conform to the self-regulation literature, which argues that people will maintain motivation to complete uninteresting tasks by altering them (Ryan & Deci, 2000; Sansone & Smith, 2000). For example, perhaps participants imagined the task as a kind of game, challenging themselves and setting personal performance standards. Another explanation is that cognitive dissonance (Thibodeau & Aronson, 1992; Elliot & Devine, 1994) may have played a role. This was a long task to complete for course credit, so the “motivated” individuals may have reported more motivation to explain why they spent over two hours doing a mundane task. The correlation between motivation and performance was not significant post-task, which could suggest that the charade of motivation had run out of energy by the end of the day.

Future research should address some of this study’s limitations. First, it is possible that other traits not measured in this work are also important; more exploration is needed here. Second, the weak correlations among state factors and performance may be an artifact of the type of performance studied. Future research should enhance the simulation and add other contextual variables. Finally, narrower measures of personality traits may yield clearer conclusions; specific hypotheses in this area should be developed and tested.

Acknowledgments

This research was partially inspired by master’s theses completed by Abbie Woodruff, Brandon Corbin and Ellen Melton and was partially supported by funds from an internal grant supplied by Angelo State University. We acknowledge William Davidson, William Fuller, Rob Mowrer and William Montgomery for their input on the original documents, Luz-Eugenia Cox-Fuenzalida for comments on earlier drafts, and our research team at ASU for their help in many ways.

References


