Cognitive Systems Performance Laboratory

Technical Report

Effects of Workload, Mood, and Perceived Stress on Error Detection In A Simulated Pharmacy Verification Task

Prepared by:

Sean Reilley, MA, Anthony F. Grasha, Ph.D, & John Schafer, Ph.D

Department of Psychology

University of Cincinnati

Technical Report: 4-1299

Cognitive Systems Performance Laboratory

Department of Psychology

429 Dyer Hall [ML 376]

University of Cincinnati

Cincinnati, Ohio 45221

Tele: 513-556-5543/ FAX 513-556-1904

Email: reilles@email.UC.Edu or Tony.Grasha@UC.Edu

Report Notes

The information in this technical report is based upon research recently published in Perceptual and Motor Skills, 2002, 95, 27-46. The paper (s) submitted using information in this report typically will appear in a different from to conform to journal standards for style, scope and length of articles. Also, the reader is reminded that based upon the outcomes of peer review, some aspects of the presentation of information in this report, and the conclusions drawn may be modified, The authors wish to thank Maggie Palazzolo, Thiago Winterstein, Gretchen Heideman, and Tara McNulty for assistance with data collection. Portions of this technical report were presented at the 11th Annual Meeting of the American Psychological Society, Denver, CO, June, 1999.

Effects of Workload, Mood, and Perceived Stress on Error Detection in a Simulated Pharmacy Verification Task

Sean A. Reilley, MA, Anthony F. Grasha, Ph.D, and John Schafer, Ph.D
Cognitive Systems Performance Laboratory
Department of Psychology
University of Cincinnati

Abstract

The relationships among workload, stress, and performance efficiency are topics of applied interest and theoretical importance to researchers in the areas of human performance. Such interest naturally extends to a variety of occupational areas including inpatient, outpatient, and community pharmacies. In that context, these relationships have become a consumer health issue due to concerns that workload contributes to job stress and a significant decline in dispensing accuracy. The present study examined the self-reported stress, mood, and detection accuracy of 102 trained college-aged individuals in a high-fidelity simulated pharmacy task. Participants evaluated the presence or errors while verifying simulated prescription sets under conditions of low and high workload. Overall, cumulative and detection theory indices of errors were compatible with estimates from pharmacy field studies. When rates of sensitivity and specificity for detection were examined, substantial variations in the identification of existing errors (sensitivity) and difficulties with detection of data-entry mistakes were observed in the high workload condition, but only modest effects emerged for the low workload condition. Finally, although increases in objective workload were associated with modest declines in detection accuracy, objective workload did not significantly impact negative mood (Dundee Stress State Questionnaire) or perceived workload (NASA Task Load Index) as expected.

Investigations of the effects of workload on subjective reports of stress, mood, and performance efficiency are common in many industrial (e.g., safety and quality control inspection, air traffic control) and medical type settings (e.g., radiology) in which workers are asked to perform multiple, successive judgements under high pressure conditions (Swets, 1992). Relationships among such variables also have appeared in the theoretical work of researchers in the area of applied human performance (cf., Parasuraman & Mouloua, 1996; Hancock & Meskkati, 1988). Our interest in this study is how such factors impact performance on a high-fidelity simulated pharmacy task and their generalization to actual pharmacy field settings as well as other occupations having sequential perceptual-motor tasks with a final verification or checking component.

Pharmacy workload is quickly becoming a topic of considerable media interest and recent public policy given that errors appeared in an estimated 3-5% of the 2.8 billion prescriptions filled in retail, outpatient, and ambulatory settings in 1998 (Proctor, 1999). Conservative estimates indicate that 0.87 – 1.5% of these errors are potentially injurious to patients (Abood, 1996; Guernsey, Ingrim, Hokanson, Doutre', Bryant, Blair, & Galvan, 1983). Other studies suggest that as many as 24% of the prescriptions filled contain mistakes and that 3-4% of the misfills are clinically significant (Flynn, Barker, Gibson, Pearson, Berger, & Smith, 1999; Allan, Barker, Malloy, & Heller, 1995). While methodological and sampling considerations are thought to account for the fluctuation in estimates of dispensing error, professionals in this area generally agree that existing rates are too high. Since four billion prescriptions are expected to be filled yearly by 2005 (Proctor, 1999), considerable debate has emerged concerning the impact of workload and other environmental factors on subsequent dispensing accuracy and worker morale.

Perceptions that objective levels of workload (i.e., hours worked, number of scripts filled) are a major contributor to the decline in dispensing accuracy and to the elevated stress levels of pharmacists emanate from personal interviews with pharmacy personnel and from pharmacy focus groups (Grasha & O'Neill, 1996). Field surveys also indicate that high prescription workload is primary among the factors preventing pharmacists from directing their efforts toward increased patient counseling and other prescription dispensing practices that are designed to lower rates of error (Raisch, 1993; Barnes, Riedlinger, McCloskey, & Montagne, 1996).

Despite these subjective reports, controlled empirical studies connecting increments in workload with pharmacy dispensing errors are slowly beginning to emerge. A review of the available research to date provides initial support for a relationship between stress and workload, but mixed evidence concerning the workload-error relationship. With regard to the former, several recently completed pharmacy simulation studies (Grasha & Schell, 1999a, 1999b; Schell & Grasha, 1999) report significant elevations in self-reported stress associated with increases in dispensing workpace. Estimates of the workload-error relationship, on the other hand, indicate that this phenomenon is more complex than previously thought (Chi, 1999). For instance, a review of the findings from studies conducted at pharmacy field sites yields large discrepancies about the purported negative effects of prescription workload. In some cases, moderate (r’s ranging from .40 to .56) to strong (r = .78) linear associations between prescription workload and error are reported (Guernsey et al., 1983; Rupp, DeYoung, & Schondelmeyer, 1992; Allan, 1994). In stark contrast, published data from other sources (Kistner, Keith, & Hokanson, 1994; Grasha et al., 1999a, 1999b) yield relatively weak statistical trends (r’s on the order of .10).

Several factors may account for the substantial discrepancies in estimates of the prescription workload-error relationship. First, in a majority of these studies, error is commonly measured across pharmacists using a cumulative or a running index of errors committed during an hour or total errors per work shift. Operationalizing error in this manner ignores “prescription exposure” or the ratio of actual errors relative to total prescription workload in a given period of time. Grasha’s (1998) secondary analysis of data reported by Allan (1994) provides a good illustration of the potential bias associated with cumulative error measures that are not equated for prescription exposure. According to Grasha, the moderate correlation (r = .56) between the average number of prescriptions filled per pharmacist half-hour and dispensing accuracy reported by Allan is virtually eliminated (r = .02) when prescription exposure was statistically controlled.

A similar set of outcomes are reported by Grasha et al. (1999a, 1999b) who employed a laboratory-based pharmacy dispensing task to investigate the psychomotor components associated with filling accuracy. Such findings raise the possibility that current estimates of the workload-error relationship may be positively biased due to over-use of cumulative error measures that do not account for the effects of prescription exposure. Use of more modern error measures, including decision-based indices, which take into account prescription exposure would allow for increased precision in determining the baseline rates of error.

A second explanation for the conflicting estimates of the workload-error relationship may be the number of uncontrolled environmental factors operating in the pharmacy field sites in which a majority of these studies were conducted. For example, factors such as the illumination level of the pharmacy (Buchanan, Barker, Gibson, Jiang, & Pearson, 1991), interruptions from customers and co-workers, and environmental distractions (Allan, 1994; Flynn et. al., 1999) have been shown to negatively affect pharmacy error rates and to interact with prescription workload effects. Given the lack of control over such variables, it remains unclear whether prescription workload makes a unique and substantial contribution to the variance associated with dispensing errors or whether its contribution is vastly overestimated due to confounding variables and/or other covariates. For example psychosocial factors such as field-independence, social and significant-other stress, and perceptions of workload were much stronger predictors of error than were measures of objective workload on a simulated pharmacy task (Grasha et al., 1999a, 1999b; Schell & Grasha, 1999).

It is also important to note that the effects of prescription workload on filling and checking tasks are rarely parceled from one another. Instead, a composite error measure is typically employed, which as Facchinetti, Campbell, & Jones (1999) point out, may obscure relatively high rates of undetected pharmacy errors. In addition, prescription workload may also differentially affect filling and checking processes as well as the specific subtypes of errors (e.g., data entry, selection, and counting mistakes) associated with both tasks. In fact, Grasha et al. (1999a, 1999b) report that in combination with other psychosocial variables, data-entry errors were associated with working slowly, while working quickly increased product selection errors. No study to date, however, has attempted to address this latter issue by examining different types of errors during the final verification or checking task as a function of different levels of workload. Nor has existing work documented the levels of stress and changes in mood associated with verification.

In the current study, the effects of controlled manipulations in prescription workload were examined on subsequent performance efficiency and reports of mood and stress using a high-fidelity pharmacy simulation task. The latter involves common psychomotor components of the dispensing task (data-entry, product selection, count & pour) and the types of perceptual judgments (i.e., discriminating correct product, verifying the accuracy of data entered and products counted, and final judgments concerning overall accuracy of the order) required of actual pharmacy personnel. Only one part of the simulated pharmacy task was employed for the present inquiry—the final order verification phase.

Given that a majority of existing research has confounded workload effects by assessing accuracy across both dispensing and verification (checking) tasks, and few have adequately controlled extraneous sources of environmental error, the present study isolates and controls such factors in order to examine performance on a verification type task. The major goals of the current investigation were to: (1) to collect initial baseline estimates of the workload-error relationship under controlled conditions using cumulative as well as decision-based, measures of error, and (2) to provide an empirical test of the purported effects of prescription workload on self-reported stress. More specifically, the following hypotheses and research questions were addressed:

> [1a] Do increases in simulated prescription load provide a direct, negative contribution to detection accuracy for a pharmacy-like verification task when prescription exposure is experimentally controlled?
>
> [1b] Do increases in simulated prescription load affect the detection accuracy for specific classes of errors under similar conditions?
>
> [2a] Based on anecdotal evidence and empirical support provided by Grasha et al. (1999a, 1999b), differential elevations in negative mood concomitant with increases in simulated prescription workload are expected between groups of individuals subjected to high and low prescription workload. Reports of negative mood are expected to be significantly higher among participants subjected to high prescription load in contrast to their low workload counterparts.
>
> [2b] Consistent with mood effects, task-related stress should also be manifested by differential reports of perceived workload between prescription workload groups. Again, greater elevations in perceived workload were expected for high workload individuals.

Method

Participants

A convenience sample of one hundred and two (37 men and 65 women; ***Mdn*** age = 22) upper level psychology undergraduates (***M*** college completed = 3.36 years, SD = 0.83) at the University of Cincinnati were trained to work as pharmacy personnel for the present study. Participation was restricted to individuals with normal or corrected normal eyesight. They were asked not to consume any caffeinated, alcoholic, or tobacco-based products one hour prior to participation[1]. In order to increase participants' motivation for maintaining performance standards similar to those of pharmacists and technicians, they were informed that the top ten most accurate on the task would receive a gift certificate for $25 at the end of the study.

Simulated Pharmacy Environment & Verification Task

A medium-sized (4.45m X 3.60m), off-white simulated pharmacy equipped with a two-way mirror and a pharmacy-like workstation served as the testing room. Although a detailed description is available elsewhere (Grasha et al., 1999a), in essence, the pharmacy simulator shown in Figure 1 was used to increase participant's awareness of a pharmacy-like environment and to minimize environmental factors (illumination, noise and temperature levels) while they performed the verification task.

In brief, the bulk of participants' work for the present study consisted of performing final evaluations for a series of pre-assembled orders. Specifically, participants were asked to visually judge the accuracy of the completed orders against their handwritten order cards while seated at the simulated workstation. At the conclusion of the verification process for each order, participants entered simple Yes/No decisions into a computerized database using an Apple SE-30 microcomputer. Three decisions were made for each order: (a) the presence or absence of errors with respect to information found in the order label (described below), and whether or not the (b) "class" and "strength" as well as the (c) quantity of simulated drugs were accurate.

Simulated Prescription Orders

The orders used in the verification task each consisted of a brown (13 cm x 7.9 cm x 27 cm) bag with a laser-printed pharmacy-like label. Each label contained the following patient information (customer name, employer, product and product quantity, and special directions for use), along with an order number, dispensing date, and the name of a fictitious pharmacist who filled the order. Inside each prescription-like bag was a clear Ziploc® (17.8 cm x 20.3 cm) bag containing a 7.5 cm x 12.5 cm handwritten, white order-card and a given quantity of "simulated drugs." To mimic the diversity of actual written orders encountered in a pharmacy, information on the prescription label was either transcribed by hand or written in script using both blue and black ball-point ink. Care was taken to insure that the handwriting and printing was

Figure 1: Panoramic view of the pharmacy simulation work space

legible and that both the order card and the computer-printed label contained the same information.

As in a pharmacy, participants encountered simulated drugs of varied colors, sizes and shapes in the pre-assembled orders they were asked to evaluate. In line with prior work (Grasha et al., 1999a), various arts and crafts materials (BEADS), hardware material (NUTS and WASHERS), paper (CLIPS), and pieces of cereal (TRIX) were chosen to mimic the diversity of actual drugs found within a pharmacy. An attempt was also made to simulate drugs that had similar spellings and sounds. As such, each type of BEADS, CLIPS, NUTS, WASHERS, AND TRIX had a corresponding misspelled set (BEEDS, CLEPS, NOTS, WOSHERS, AND TREX) that was identical to the original forty-eight materials except for a tiny hand painted colored mark (a dot or a line).

For reference purposes, extraneous amounts of each of these products were available from one of two types of opaque, off-white pharmacy bottles (1400ml and 100ml) or from one of two styles of Tupperware® containers (small [100ml] and medium [185ml]). These bottles were sorted and stored on regulated pharmacy shelving as depicted in the pharmacy workstation in Figure 1. With practice, participants were able to easily identify each "simulated drug" by inspecting the pre-printed bottle labels (e.g., BEADS-GL 1.0) that contained nomenclature referencing products by class (e.g., beads, nuts), color (e.g., gold, red, green), and size (e.g., 1.0 vs. 10.0).

Experimental Design

A two-between, one-within mixed factorial design with task period serving as the within factor was used. The two between factors, prescription workload (high, low) and the class of errors introduced into a subset of the orders (counting, selection, and data-entry mistakes) were factorially combined to yield a total of six experimental conditions. Across conditions, prescription workload was operationally defined as the quantity of orders to be judged within a 60-minute period. Based on personal communications with pharmacy experts (Shapiro, February, 26, 1998), participants in the "high" workload (HW) condition examined 120 orders, distributed evenly over two 60 minute task periods, while participants in the "low" workload (LW) condition examined 72 orders (40% less than the HW condition) in a similar time frame.

A defective order (i.e., one containing a mistake) was the critical signal for detection, while accurate orders served as background events for the present study. Both accurate and defective prescriptions were selected from those used in prior work (Grasha et al., 1999a) as well as those supplemented by the principal investigator. Signal probability, or the likelihood an order contained an error, was held constant at a rate of 24%[1]. The number of actual defective orders was deliberate and reflects an attempt to provide ecological validity with regard to reports of more progressive pharmacy error rates.

The selection of the class of errors (data-entry, selection, and counting) also reflected those most frequently encountered in outpatient pharmacy settings. In the present study, a data-entry error was defined as a mismatch of information (e.g., patient name, drug name) between the pre-printed computer label that appeared on the brown bag and the handwritten order card contained in the Ziploc® bag. Discrepancies between a requested item or an item quantity referenced on the order card and the actual items in the Ziploc® bag served as the vehicle for presenting selection and counting errors. Specifically, orders containing both the wrong "simulated drug" (i.e., CLIPS 1.0 vs. CLEPS 1.0) and a "simulated drug" of the wrong strength (i.e., CLIPS 1.0 vs. CLIPS 10.0) were used to equally represent selection errors. In a similar fashion, counting errors reflected roughly equal amounts of excesses and deficiencies in the quantity of the "simulated drugs" in the Ziploc" bag to that requested on the order-card.

Procedures and Instruments

Upon arrival for the 3-hour experiment, participants completed informed consent procedures and then evaluated their mood using the University of Wales Institute of Science and Technology Mood Adjective Checklist (UMACL; Matthews, Jones, & Chamberlin, 1990). The UMACL is composed of twenty-nine adjectives that are rated on a four-point scale. From these items, three independent scales (energetic arousal, tension, and hedonic tone [pleasant mood]) are formed each of which demonstrates satisfactory internal consistency [Cronbach's (1951) alpha coefficients exceeded .74 for each scale in the present study] and moderate test-retest reliability (Matthews, Campbell, Joyner, Huggins, Falconer, & Gilliland, 1997). Next, participants were oriented to the simulated pharmacy environment and completed a 15-minute training session consisting of nine prescriptions (2 defective + 7 accurate). All participants were informed that both accuracy and speed were important for this task, but not to sacrifice accuracy for speed.

At the completion of training and each subsequent 60-minute task period, participants completed a modified version of the Dundee Stress State Questionnaire (DSSQ; Mattp$ws et al., 1997) that included the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988)[2]. Participants responding to the NASA-TLX are asked to rate six dimensions of perceived workload using separate 100-point scales. From these ratings, a global estimate of perceived workload and the unique contribution imposed by task demands (mental, physical, and temporal demands) and those associated with the interaction of the participant with the task (perceived effort, frustration, and performance concern) can readily be identified. The NASA-TLX has been used extensively across a variety of experimental and industrial type tasks and has acceptable test-retest reliability and internal consistency (Cronbach's alpha exceeded .65 for the present study). Although data of an exploratory nature were also collected using the supplementary scales of the modified DSSQ, these data are not discussed here as they did not pertain directly to the specific, a priori hypotheses and scope of the present investigation.

Following completion of the training session and post-test measures, any questions pertaining to the experimental task were answered. Participants were then given a delivery receptacle containing 1/6 of the pre-assembled prescriptions for their respective workload (high vs. low prescription load) and error conditions (data-entry, product selection, and counting mistakes). They were instructed to inspect and make judgements regarding the accuracy of each prescription in a fashion similar to those made during the training period. Participants were also informed that equivalent amounts of simulated prescriptions would be sent via the experimenter "at a standard rate, designed to be appropriate for college students to complete the task (actually every 10-minutes)." The experimenter placed a target time next to a small digital clock which participants could use to monitor progress. In addition, participants were reminded that the pharmacy counting tray used during training could be used to help verify the accuracy of the number of "simulated drugs." Finally, a five-minute verbal warning was given prior to the end of the 60-minute task period. After such time, participants filled out the modified DSSQ and the NASA-TLX and the second 60-minute task period commenced. Procedures for the second 60-minute task period were identical to those already discussed.

Results

Most participants (94%) demonstrated sufficient understanding of the verification task after assessing nine practice orders (M training score = 8.27/9.00, SD = 0.95). The scores from thirteen individuals however did not exceed the pre-set inclusion criterion of 5/9 judged correctly. These individuals were given a second opportunity to re-read the instructions and to re-assess the training set. Participants who failed to meet the inclusion criterion after a second training session (n = 8) were thanked for their time and given a short debriefing. Five of the thirteen individuals, however, did meet the inclusion criterion after their second attempt and were included in the final sample of ninety-four participants (36 males and 58 females).

Statistical Analyses

Due to the number of statistical tests to be conducted, a more conservative, omnibus alpha level of 0.01 was chosen for all analyses in an attempt to correct inflation of the Type I error rate. The major assumptions underlying all parametric statistical tests to be performed were assessed and in several cases arc-sine transformations were applied in an attempt to stabilize the variance associated with accuracy measures. The efficacy of these variance stabilization procedures were evaluated using Kolmogorov-Smirnov tests (Smirnov, 1948). For major analyses involving tests of group differences, a series of Repeated Measures Analysis of Variance procedures using three between factors (workload group, error class, and gender) and one within factor (task period) were performed. Significant F-tests were followed by pre-planned contrasts to test specific directional hypotheses unless otherwise noted.

Effects of Increases in Simulated Prescription Workload on Overall Detection Accuracy and Detection of Specific Classes of Error

A comparison of the workload-accuracy relationship afforded by both cumulative and decision-based detection indices was of interest. First, rates of verification accuracy were calculated from cumulative errors similar to those commonly found in the pharmacy literature (cf., Allan, 1994; Kistner et al., 1994). Second, two decision-based detection indexes, commonly referred to as sensitivity and specificity were calculated. Sensitivity, within a pharmacy context, is the ability to detect flawed prescriptions (a false negative problem). By contrast, specificity references a false positive problem that could pose potential delays to the verification process in a work environment because one would try to correct an error when none is indeed present. More important, given our criticism of cumulative error measures, sensitivity and specificity reflect rates of erroneous decision making (false negatives and false positives, respectively) while taking into account all possible outcomes of such decisions[3].

In the current study, and as shown in Figures 2 and 3, cumulative accuracy ranged from 65% to 98% with a 65% to 93% range for sensitivity and 86% to 98% for specificity. This is similar to field work where rates of cumulative error (82% to 96%) were observed (Cambell & Facchinetti, 1998). Also similar to industry estimates, error detectability, as shown in Figures 2 and 3 varies with the type of error-class introduced into the simulated prescription set. The most striking feature, across accuracy measures, is the consistent decline in the ability of HW individuals to detect prescriptions having flawed labels.

When the magnitude of the relationship between prescription workload and detection accuracy is assessed using point biserial correlations some striking differences emerge. Curiously, the cumulative detection index yielded a moderate effect for prescription workload (***r*** = .42) that, although slightly lower in magnitude, is consistent with reports from pharmacy field studies mentioned earlier. By contrast, sensitivity and specificity, which unlike cumulative estimates, control for prescription exposure produce significantly lower estimates of this relationship (***r's*** ranging from .05 to .14). Further examination of the workload effect for specific types of errors using decision-based measures yielded additional information that is not readily available from the cumulative index. Namely, the reduced detection accuracy of HW individuals appears to reflect a lapse in their sensitivity to label errors, despite relatively high rates of specificity.

Consistent with observations found in Figure 3, the results of the ANOVA on arc-sine transformations of the sensitivity measure yielded significant main effects for Workload (***F*** (1, 82) = 6.26, p < .01), Error condition (***F*** (2, 82) =12.75, p < .01), as well as a significant Workload x Error condition interaction (***F*** (2, 82)= 8.67, p < .01). No statistically significant main effects or interactions emerged for specificity. Post- hoc Scheffé analy-

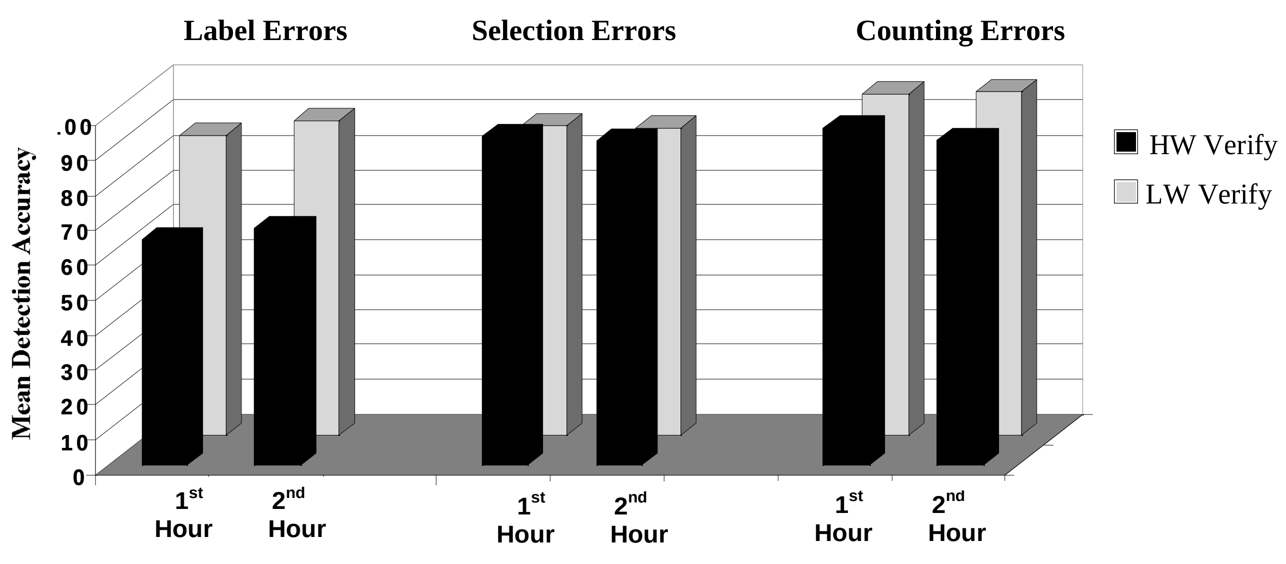

Figure 2: Simple Accuracy as a Function of Workload Group and Type of Verification Error

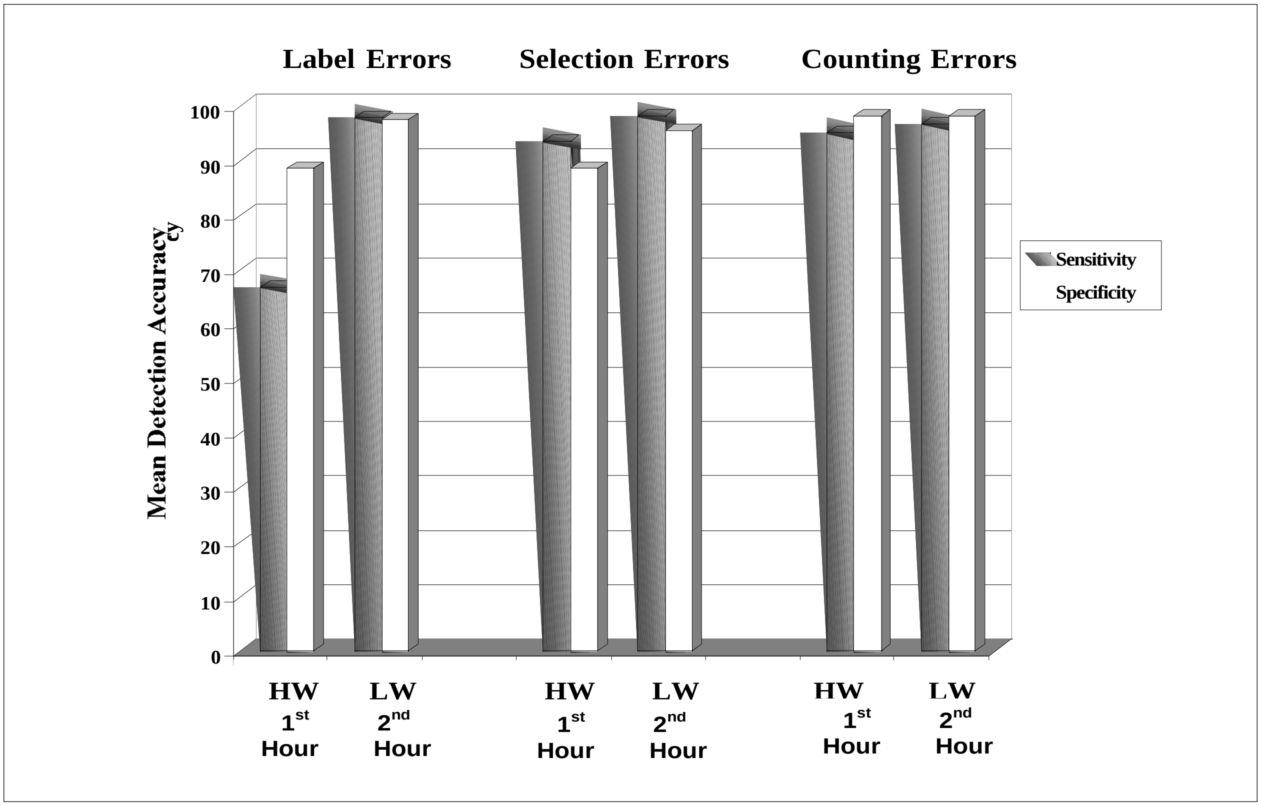

Figure 3: Sensitivity & Specificity as a Function of Workload Group and Type of Verification Error

ses for the sensitivity variable confirmed that HW individuals were least sensitive in their judgements when evaluating simulated prescriptions containing label errors (Period 1, $t\,(1, 32) = -3.05$, $p < .01$; Period 2, $t\,(1, 32) = -3.78$, $p < .01$), compared to errors of other types. Because a violation of normality associated with the decision-based measures could leave the ANOVA procedure with low power to detect Type 2 errors (Wilcox, 1998), we also fit a series of generalized linear models including Poisson and Over-dispersed Poisson models (c.f., McCullagh & Nelder, 1989; Gardner, Mulvey, & Shaw, 1991) to the data.

The results of these alternative analyses largely corroborated the pattern of findings for sensitivity and specificity. That is, across error conditions (data-entry, selection, counting), specificity did not emerge as a significant contributor to workload group. Sensitivity, on the other hand, did appear to make a direct contribution to workload status that is specific only to data-entry errors. More precisely, sensitivity offered a unique contribution to high workload status only for the first task period ($F\,(1, 29) = 12.12$, $p < .01$).

Effects of Increases in Simulated Prescription Workload on Negative Mood State

As a whole, the parameter estimates and internal reliability coefficients (Cronbach's α) for the UMACL obtained in the present study are consistent with normative data reported by Matthews et al. (1997). For presentation purposes, mood estimates from each simulation period are depicted graphically in Figure 4 as standardized change scores from the individual's pre-simulation or "baseline" report. A positive departure from zero (baseline mood) indicates an increase in mood state referenced in standard deviation units, while a decrease from baseline reflects a decrease in mood state from the individual's pre-simulation baseline in standard deviation units.

Visual inspection of the plotted standardized change scores reveals modest declines in energy and pleasant mood that parallel elevations in tension. This pattern also appears to be influenced by workload with HW individuals evidencing slightly poorer mood over the course of the simulation when compared to reports from LW individuals. Based on the ANOVA results, decreased energy ($F\,(3, 243) = 19.44$, $p < .01$), pleasant mood ($F\,(3, 243) = 9.64$, $p < .01$), and elevated tension ($F\,(3, 243) = 12.84$, $p < .01$) were evident. Contrary to predictions, the observed increase in negative mood among HW individuals did not significantly differ ($p > .01$) from those of LW individuals across mood measures according to the ANOVA model. Finally, neither gender nor error condition emerged as significant main

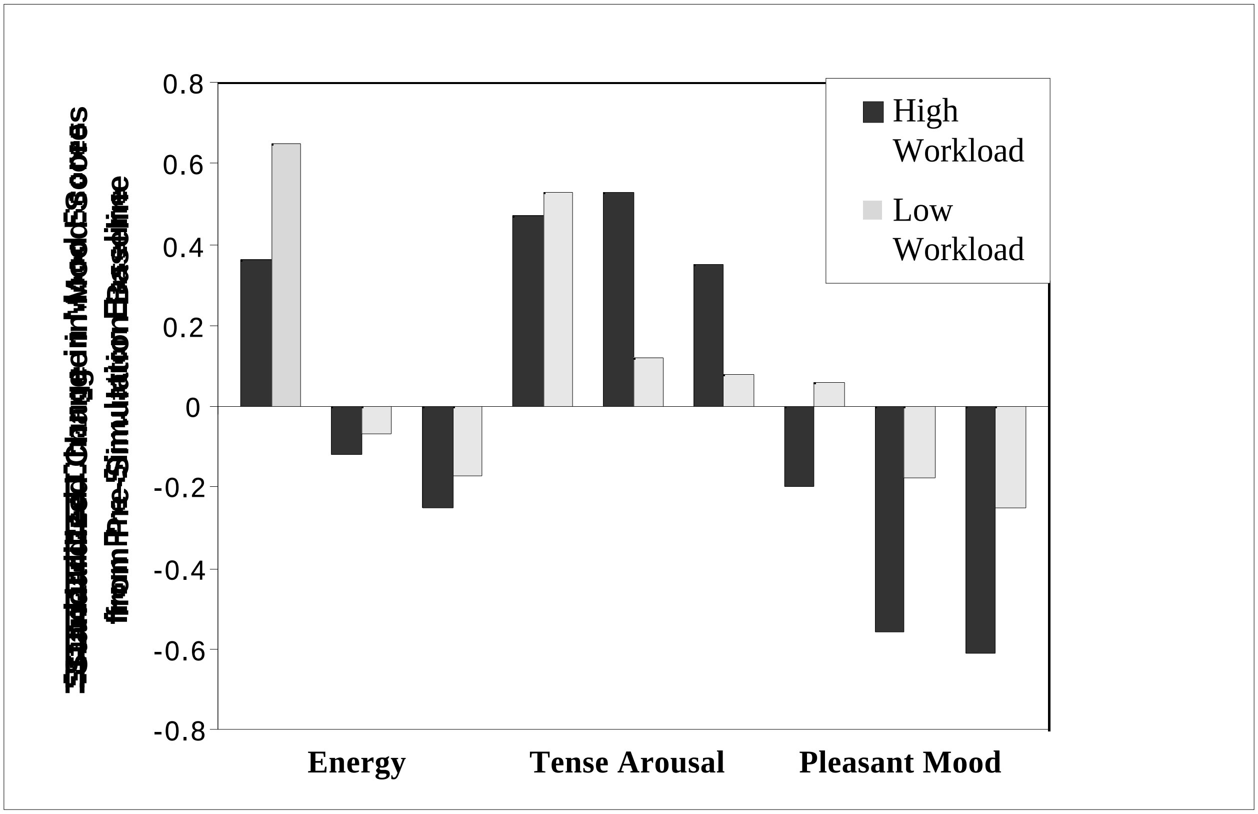

Figure 4: Standardized Change in Prescription Workload Groups'Mood Scores from Baseline Across Verification Task Periods

effects or as significant 1^{st} order interactions with workload. Thus, based on analyses involving reports of negative mood, no statistical support was obtained for the hypothesis that HW individuals' reports of negative mood during verification would be significantly greater than similar reports made by LW individuals.

Effects of Increases in Simulated Prescription Workload on Perceived Workload Scores

Three reports of perceived workload (post-training and once after each experimental period) using the NASA-TLX were collected. Given the current disagreement over weighting procedures (for a review, see either Nygren, 1991; Dickinson, Byblow, & Ryan, 1993), only raw NASA-TLX scores were used to evaluate directional hypotheses involving perceived workload. When NASA-TLX scores from individuals in each of the respective workload conditions across studies are compared, no substantial differences are evident between high (***M*** NASA-TLX = 36.85) and low workload (***M*** NASA-TLX = 37.32) individuals' reports following completion of the training period. However, reports of perceived workload during the first 1-hour verification period remain at a constant, moderate level for HW individuals (***M*** NASA-TLX = 36.57), while slightly lower reports emerge for LW individuals (***M*** NASA-TLX = 31.54). A similar, but less pronounced trend is evident during the second 1-hour verification period in which TLX scores for HW individuals (***M*** NASA-TLX = 34.31) are slightly higher than reports from their LW counterparts (***M*** NASA-TLX = 28.99). The ANOVA model did not readily confirm these observed differences, however. In that analysis, both the main effect for Workload (***F*** (1,82) = 1.37. ***p*** > .01) and the Prescription Workload x Task Period interaction were non-significant (***F*** (2, 162) = 3.73, ***p*** > .01).

Because no statistically significant differences emerged for perceived workload between verification groups and across task periods, we collapsed across these factors and computed mean NASA-TLX scores for comparison purposes with previously collected pharmacy data. Table 1 displays both composite NASA-TLX and individual NASA-TLX-component scores from several sets of comparison data. First, NASA-TLX data reported by Grasha (1998) are presented for fifty pharmacists and fourteen technicians who were asked to give perceived workload estimates based on an average of their last three pharmacy work-shifts. NASA-TLX data are also presented for two groups of college students, a 70-minute (high workpace) and a 90-minute (low workpace) group who completed a pharmacy-like dispensing task

Table 1

Mean composite-TLX scores and individual TLX-component scores for verification data and several relevant sets of comparison data.

Comparison Data	***Total*** TLX	***Mental*** Demand	***Physical*** Demand	***Temporal*** Demand	***Performance*** Concern	***Effort***	***Frustration***
Pharmacists	***70.8***	***88.9***	***31.2***	***74.4***	***87.6***	***85.9***	***57.1***
Technicians	68.4	91.1	47.1	64.1	78.6	71.4	58.2
Simulated Dispensing	57.8	51.9	26.8	48.6	79.5	53.3	27.8
Simulated Verification	42.5	47.6	24.3	35.7	75.7	27.5	24.7
Vigilance	48.0	48.9	14.9	57.5	52.3	51.9	62.4
Card Sorting	20.0						

Note: Sample size for pharmacists and technicians equals 50 and 14, respectively.Total samples of individuals completing separate simulated dispensing and verification tasks equals 71 and 94. Samples sizes for vigilance and card sorting data equal 162 and 18, respectively.

without a verification component (Grasha et a1., 1999a). For these data, mean NASA-TLX scores are expressed collapsed across the 70 and 90-minute groups. Vigilance data collapsed across two short vigils (30-40 minutes in duration) are displayed for one hundred sixty-two college participants based on prior work by Hitchcock, Dember, Warm, Moroney, & See (1999) and See, Simon, Warm, Dember, & Fowler (1995). Finally, composite NASA-TLX data from eighteen college participants who were asked to perform a 104-second card-sorting task (Hitchcock et al.,1999) are also provided as a lower-bound comparison group.

It is clear from examination of the TLX data in Table 1 that pharmacists and technicians report experiencing a higher degree of perceived workload in contrast to other comparison groups. In particular, mental demand, temporal demand, performance concern, and effort make substantial contributions to the workload signature for their job. As one might expect, students completing a filling-only task under distraction-free conditions report significantly lower perceived workload compared to actual pharmacists (***t*** (1, 119) = 4.34, ***p*** < .01). A similar trend for pharmacy technicians occurs, however, due to the small number of available reports significant tests were not conducted. More important, the reports from students in the filling-only condition reflect a similar pattern in the workload signature to those reported by actual pharmacists and technicians.

The data from the present inquiry follow a similar pattern. Given that verification tasks consume approximately 4% of a pharmacist's total workload (Lin, Grasha, and Ivey, 1998), it is not surprising that reports of perceived workload from the verification sample are significantly lower in magnitude (***t*** (1, 163) = 6.48, ***p*** < .01) than are similar scores from participants completing a more cognitively complex dispensing-only condition. This is not to say, however, that a verification task by itself is not mentally demanding. In fact, simple visual inspection of the card sorting data as a lower bound comparison provides evidence suggesting quite the contrary.

More to the point, TLX-data associated with the verification task also reflect a consistent workload signature that has been demonstrated empirically for simulated dispensing tasks and that is evident in reports from actual pharmacy personnel. Such a workload pattern also appears to be different in kind to those of other related non-pharmacy tasks. For instance, the vigilance data referenced in Table 1 displays a workload signature with temporal demand and frustration being the primary contributors to composite-TLX scores. Thus, while the TLX data associated with the verification task are conceivably low, they mirror a unique and consistent pharmacy workload signature that is qualitatively different from other related tasks.

Discussion

Effects of Prescription Workload on Overall Detection Accuracy for the Simulated Pharmacy Verification Task

In general, rates of detection accuracy in the present study appear to be in line with pharmacy field estimates of 82-96% reported by Cambell & Faccinetti (1998). That workload provides a direct effect for detection accuracy when simulated prescription exposure as well as a variety of extraneous environmental factors are experimentally controlled is interesting given the rudimentary nature of the verification task. Perhaps the most striking outcome of the present inquiry is the absence of a moderate to strong linear relationship between workload and detection rates when such factors are controlled.

One explanation for the large discrepancy in estimates of the workload-error relationship concerns the type of error measure used to assess this relationship. No pharmacy study to date has examined verification or dispensing errors using decision-based error measures that appropriately incorporate prescription exposure into the final error estimate. Thus, one implication of these findings is that future work should consider employing error measures that take into account prescription exposure since failing to control for this variable leads to positively biased estimates of the workload-error relationship.

It remains to be seen if the magnitude of the prescription workload effect observed for simulated verification is replicable, and if so, would follow a linear, additive relationship as additional task components (i.e., filling) and extraneous factors (e.g., interruptions) are introduced under controlled conditions. Alternatively, prescription workload may provide only a modest effect which in combination with other factors may best account for the majority of the variance associated with dispensing errors. To begin to address this issue, a follow-up study is in progress that combines both verification and filling tasks over a 6-hour task period.

Effects of Prescription Workload on Detection of Specific Classes of Errors

Consistent with prior literature examining aspects of the dispensing task, prescription workload negatively impacted verification accuracy for specific classes of errors. In particular, detection of label errors was poorest under conditions of high prescription workload. Given the direct experimental manipulation of specific error classes in the present study, one implication of poorer detection ability of HW individuals for label errors may be a more lenient response strategy for checking labels as compared to investigating discrepancies for other potential errors. Under high pressured conditions, it seems intuitive that a pharmacist would be more concerned with making sure the correct medication was given to a patient, rather than checking for misspellings in the patient's name and/or street address. Since the average pharmacist spends less than 10 seconds verifying a prescription (Lin et al., 1998), there would be some advantage to focusing on the more problematic parts of the prescription.

Failing to detect slight differences between two patient names may be one danger in readily adopting this strategy, especially under HW conditions since both individuals might receive the wrong prescription. A signal detection analysis would likely provide relevant data to address such speculation, however, the low rates of false positives in the present study were not appropriate even for a non-parametric signal detection analysis. Unfortunately, this is common outcome for many researchers who investigate a variety of decision-based problems having low rates of error within the health sciences. Nevertheless, future studies that allow assessments over longer portions of the work-shift may yield data that are analyzable within a signal detection paradigm, thus providing sufficient data to address this issue.

Effects of Increases in Simulated Prescription Workload on Negative Mood and Perceived Workload

In the present study, general increases in negative mood were evident for both workload groups as they spent more time verifying the accuracy of simulated orders. Statistically significant variations in negative mood across the three task periods were obtained. Contrary to expectations, however, increases in simulated prescription workload did not provide a statistically significant increase in negative mood during verification. The lack of a significant main effect for workload may indicate that simulated prescription workload does not provide a direct and meaningful effect on self-reported mood that is independent of other factors. This conclusion, in our view, is tentative given several factors, most notably: the duration of the verification task, the complexity of the work involved, and the continuity of the prescription workload.

The duration of the verification task may account for a lack of a mood effect in the present study. Participants, in the present inquiry, were asked to perform a simulated verification task only for two hours which may not be representative of the typical work-shift pharmacists and technicians follow prior to taking a work-break. Thus, a mood effect that is directly related to workload may emerge as individuals are asked to spend longer amounts of time on task, thereby better approximating a typical pharmacy work-shift. Data relevant to this issue are also being collected in a follow-up study where participants are asked to spend six hours on task with brief rest breaks incorporated into the design.

Second, a filling task may reflect a more cognitively complex job that, in contrast to verification, is more mentally draining due to the multiple physical and mental operations that are needed. If the relationship between cognitive complexity and negative mood is valid, as the verification task becomes more cognitively complex, for example, by adding a dispensing component, increases in negative mood would be expected to follow in a linear fashion. Support for this notion is bolstered by reports of significant increases in negative mood, as assessed by the UWIST Mood Adjective Checklist, for more complex tasks like driving (Matthews, et al., 1997; Fairclough & Graham, 1999) and sustained attention (Helton, Dember, Warm, & Matthews, 1999). A study is currently in progress to assess the mood effects associated with a more cognitively demanding pharmacy-like task that incorporates both dispensing and verification components.

Finally, the largest increase in negative mood and perceived workload between groups was found after the first simulated task period following training. Such differences, although non-significant, may speak to a more global “stress effect” that is evident for transitions from one prescription workload level to another (i.e., high-low; low-high). That is, in the present study, individuals in the high workload condition evidenced a non-significant trend toward greater elevations in negative mood and perceived workload following an increase of 40% (60 orders / 60 minutes) in their prescription workload from that of training (9 orders/ 15 minutes). Low workload individuals, by contrast, reported trends consistent with smaller increases in negative mood and perceived workload when asked to complete a ratio of orders (36 order / 60 minutes) equivalent to those encountered during training (9 orders / 15 minutes).

Findings by Moroney, Warm, & Dember (1995) regarding work-shift changes and their association with increases in perceived workload may provide a useful parallel to the present data. According to Moroney et al. (1995), the greatest increase in perceived workload scores emerged when participants experienced a shift from a low to high work demands. If applicable to a pharmacy situation, the findings from Moroney et al. (1995) indicate that the most pronounced mood and perceived workload effects should be evident during periods when people experience a substantial shift from having few orders to dispense to an overload in orders to fill during peak hours. Corresponding data on workload transitions in the pharmacy literature unfortunately does not exist despite recent media and empirical interest in the relationship.

Limitations of Present Investigation

Conclusions drawn from laboratory-based studies are subject to a host of sampling and generalizability limitations, of which our study is no exception. Participants in the present study were drawn from a convenience sample of college-students and thus were not trained pharmacy personnel who routinely dispense products that could potentially harm others. A major underlying assumption of this laboratory-based study and others conducted previously using college-aged participants was that the psychomotor components associated with the type of errors pharmacists make are similar in kind to those made by novices performing similar tasks under controlled conditions. Support for this assumption is bolstered by a variety of similar outcomes between the two groups. For instance, performance measures such as time to complete various psychomotor components (e.g.. data entry, product selection, and count-and-pour) as well as the frequency and rates of both dispensing and verification errors obtained are compatible with those found in the pharmacy literature (Grasha et al., 1999a, 1999b; Schell et al., 1999).

While the simulation task approximates what occurs in aspects of filling prescriptions, it does not currently mimic the variety of clinical judgments about legal limits, drug interactions, and appropriateness of prescription information for a patient. Additionally, factors such as interruptions, irregular noise, and inadequate lighting were not present in the current form of the task simulation, although they are common occurrences in outpatient, hospital, and retail pharmacies. While additional studies are needed to evaluate the combined effects of such factors under controlled conditions, the simulated pharmacy task does allow for a degree of experimental control over the situation than cannot be achieved in field-settings. In addition, interventions designed to reduce their impact can be initially and safely tested in the simulation, which is a strength of the laboratory-based approach to this problem. As such, any data obtained from this approach can be compared to existing field-site data in the literature or validated by trying to obtain similar results in field sites.

Finally, significance tests conducted in the present inquiry were moderately affected by low statistical power due to inadequate sample sizes. Although a more stringent alpha level was employed, such procedures in the present study are likely to provide less than a conservative correction to inflation of the Type I error rate due to the number of statistical tests performed. Although experimental economy often preclude such decisions, more research with larger samples employing similar design methodology is needed to better determine the underlying relations between workload, stress, and verification error.

Footnotes

[1] Our rationale for holding signal probability constant across groups was to increase the likelihood that a significant prescription workload effect reflected variance attributable to differing amounts of simulated orders and not exposure differences for the quantities of artificial errors introduced into the prescription set. Thus, within the high workload condition, 30 of the total 120 presented orders contained errors, while 18 of the total 72 orders in the low workload condition were flawed.

[2] Confidential data were also collected for archival purposes concerning participants' use of prescription and over-the-counter medications as well as consumption of illegal substances, alcohol, and tobacco-based products 24-hours prior to the experiment.

[3] That is, both sensitivity and specificity reflect a ratio of errors to total outcomes. For sensitivity, this ratio is given by (Correct detection of flawed orders [CDFO] divided by [CDFO] + False Negatives). Similarly, specificity reflects (Correct detection of accurate orders [CDAO] divided by CDAO + False Positives).

[4] Replacing the perceived workload scale found in the DSSQ with the NASA-TLX was based on several criteria. First, the perceived workload scale in the DSSQ, although quite acceptable, has not been validated as extensively across work domains as has the NASA-TLX. Second, perceived workload estimates, using the NASA-TLX, were available for comparison purposes from a sample of pharmacy technicians and actual pharmacists as well as college-aged participants who performed a simulated dispensing task (Grasha et al., 1999a, 1999b).

References

Abood, R.R. (1996, March). Errors in pharmacy practice. Continuing Education Program # 073-000-96-041-H03. ***U.S. Pharmacist,*** 2-11

Allan, E.L. (1994). ***Relationships among facility design variables and medication errors in a pharmacy.*** Unpublished doctoral dissertation. Auburn University.

Allan, E. L. (1995). Dispensing errors and counseling in community practice. ***American Pharmacy, 12,*** 25-33.

Barnes, J.M., Riedlinger, J.E., McCloskey, W.W., & Montagne, M. (1996). Barriers to compliance with OBRA'90 regulations in community pharmacies. ***The Annals of Pharmacotherapy, 30,*** 1101-1105.

Buchanan, T.L., Barker, K.N., Gibson, T., Jiang, B.C., & Pearson, R.E. (1991). Illumination and errors in dispensing. ***American Journal of Hospital Pharmacy, 48,*** 2137-2145.

Cambell, G.M. & Facchinetti, N.J. (1998). Using process control charts to monitor dispensing and checking errors***. American Journal of Health-Systems Pharmacy, 55,*** 946-952.

Chi, J. (1999, January 18). Workload and errors: It's more complex than you may think. ***Drug Topics, 143,*** 29.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. ***Psychometrika, 16,*** 297-335.

Dickinson, J, Byblow, W.D., & Ryan, L.A. (1993). Order effects and the weighting process in workload assessment. ***Applied Ergonomics, 24,*** 357-361.

Drury, C.G. (1994). The speed-accuracy trade-off in industry. ***Ergonomics, 37***, 747-763.

Faccinetti, N.J., Cambell, G.M., Jones, D.P. (1999). Evaluating dispensing error detection rates in a hospital pharmacy. ***Medical Care, 37,*** 1248-1256.

Fairclough, S.H., & Graham, R. (1999). Monitoring driver impairment due to sleep deprivation or alcohol. In M.W. Scerbo & M. Mouloua (Eds.), ***Automation technology and human performance***. New Jersey: Lawrence Erlbaum Associates.

Flynn, E. A., Barker, K. N., Gibson, T., Pearson, R. E., Berger, B. A., & Smith, L. A. (1999). Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. ***American Journal of Health-System Pharmacy, 56,*** 1319-1323.

Galen, R.S., & Gambino, S.R. (1975). ***Beyond normality: The predictive value and efficacy of medical diagnoses.*** New York: John Wiley & Sons.

Gardner, W., Mulvey, E.P., & Shaw, E.C. (1991). Regression analyses of counts and rates: Poisson, over-dispersed Poisson, and negative binomial models. ***Psychological Bulletin, 118***, 392-404.

Grasha, A.F., & O'Neill, M. (1996, June). Cognitive processes in medication errors. ***U.S. Pharmacist***, 96-109.

Grasha, A.F. (1998). ***Workload and pharmacy dispensing errors.*** Cognitive-Systems Performance Laboratory (Research Update, 1 (2), 1-15). Cincinnati, Ohio: University of Cincinnati, Department of Psychology.

Grasha, A.F., & Schell, K. G. (1999a). Effects of intrapersonal, interpersonal, and task factors on human performance and error in a simulated pharmacy dispensing task. ***(Technical Report 2-0699).*** Cognitive Systems Performance Laboratory, Cincinnati, Ohio: University of Cincinnati, Department of Psychology.

Grasha, A. F., & Schell, K. G. (1999b, August). ***Psychosocial factors and time-accuracy functions in a simulated pharmacy.*** Paper presented at the Annual Meeting of the American Psychological Association, Boston, MA.

Grasha, A. F., Schell, K. G., Reilley, S., & Tranum, D. (1999c). ***Correcting human error in the prescription process.*** Paper presented at the annual meeting of the National Association of Chain Drug Stores, San Diego, CA.

Guernsey, B.G., Ingrim, N.B., Hokanson, J.A., Doutré, W.H., Bryant, S.G., Blair, C.W., & Galvan, E. (1983). Pharmacists' dispensing accuracy in a high-volume outpatient pharmacy service: Focus on risk management. ***Drug Intelligence and Clinical Pharmacy, 17,*** 742-746.

Hancock, N., & Meskkati, N. (1988). ***Human mental workload***. Ansterdam: North-Holland

Hart, S.G., Staveland, L.E. (1988). Development of a multidimensional workload rating scale: Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), ***Human mental workload***. Amsterdam: Elsevier.

Helton, W.S., Dember, W.N. Warm, J.S., & Matthews, G. (in press). Optimism-pessimism and false failure feedback: Effects on vigilance performance***. Current Psychology: Research & Reviews.***

Hitchcock, E.M., Dember, W.N., Warm, J.S., Moroney, B.W., & See, J.E. (1997). Effects of cueing and knowledge of results on workload and boredom in sustained attention. ***Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting,*** 1298-1302.

Kistner, U.A., Keith, M.R., Sergeant, K.A., & Hokanson, J.A. (1994). Accuracy of dispensing in a high-volume, hospital-based outpatient pharmacy. ***American Journal of Hospital Pharmacy, 51,*** 2793-2797.

Lin, A., Grasha, A.F., & Ivey, M. (1998, December). ***Integrating pharmacy automation technology by improving workflow and facility design.*** Spoken session delivered at the annual meeting of the Association of Hospital Pharmacies: Baker APS Continuing Education Program and Symposium, Las Vegas, NV.

Matthews, G., Campbell, S., Joyner, L., Huggins, J., Falconer, S., & Galliard, K. (1997). ***The Dundee Stress State Questionnaire: An initial report***. Unpublished manuscript, University of Dundee.

Matthews, G., Jones, D.M., & Chamberlin, A.G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. ***British Journal of Psychology, 81,*** 17-42.

McCullagh, P., & Nelder, J.A. (1989). ***Generalized Linear Models.*** London: Chapman & Hall.

Moroney, D.W., Warm, J.S., & Dember, J.S. (1995). Effects of demand transitions on vigilance performance and perceived workload. ***Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting,*** 1375-1379.

Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. ***Human Factors, 33,*** 17-33.

Parasuraman, R., & Mouloua, M. (1996). ***Automation and human performance: Theory and applications.*** Mahwah, NJ: Earlbaum.

Proctor, K. (1999). ***Pharmacist demand and pharmacist supply.*** Paper session presented at Pharmacy Practices Issues Seminar sponsored by the National Association of Chain Drug Stores, Chicago, IL.

Raisch, D.W. (1993). Barriers to providing cognitive services. ***American Pharmacy, 33,*** 54-58.

Rupp, M.T., DeYoung, M., Schondelmeyer, S.W. (1992). Prescribing problems and pharmacist interventions in community practice. ***Medical Care, 30,*** 926-941.

Schell, K.G., & Grasha, A. F. (1999). ***Anxiety and speed-accuracy functions in a simulated pharmacy task.*** Manuscript submitted for publication.

See, J.E., Simon, J.O., Warm, J.S., Dember, W.N., and Fowler, J.G. (1995, April). ***The effects of transitions in spatial uncertainty on vigilance performance and subjective workload.*** Paper session presented at the Southern Society for Psychology and Philosophy meeting, Virginia Beach, VA.

Smirnov, N.V. (1948). Table for estimating the goodness of fit of empirical distributions. ***Annals of mathematical statistics, 19,*** 279-281.

Swets, J.A. (1992). The science of choosing the right decision threshold in high stakes diagnostics. ***American Psychologist, 47,*** 522-532.

Wilcox, R.R. (1998). How many discoveries have been lost by ignoring modern statistical methods? ***American Psychologist, 53,*** 300-314.