Predicting Dual-Task Performance With the Multiple Resources Questionnaire (MRQ)

David B. Boles, University of Alabama, Tuscaloosa, Alabama, **Jonathan H. Bursk,** Sikorsky Aircraft, Stratford, Connecticut, and **Jeffrey B. Phillips** and **Jason R. Perdelwitz,** University of Alabama, Tuscaloosa, Alabama

Objective: The objective was to assess the validity of the Multiple Resources Questionnaire (MRQ) in predicting dual-task interference. **Background:** Subjective workload measures such as the Subjective Workload Assessment Technique (SWAT) and NASA Task Load Index are sensitive to single-task parameters and dual-task loads but have not attempted to measure workload in particular mental processes. An alternative is the MRQ. **Method:** In Experiment 1, participants completed simple laboratory tasks and the MRQ after each. Interference between tasks was then correlated to three different task similarity metrics: profile similarity, based on r^2 between ratings; overlap similarity, based on summed minima; and overall demand, based on summed ratings. Experiment 2 used similar methods but more complex computer-based games. **Results:** In Experiment 1 the MRQ moderately predicted interference ($r = +.37$), with no significant difference between metrics. In Experiment 2 the metric effect was significant, with overlap similarity excelling in predicting interference ($r = +.83$). Mean ratings showed high diagnosticity in identifying specific mental processing bottlenecks. **Conclusion:** The MRQ shows considerable promise as a cognitive-process-sensitive workload measure. **Application:** Potential applications of the MRQ include the identification of dual-processing bottlenecks as well as process overloads in single tasks, preparatory to redesign in areas such as air traffic management, advanced flight displays, and medical imaging.

INTRODUCTION

Subjective workload measures attempt to quantify the effort exerted in work activity, using numerical ratings or other indicators that do not themselves directly measure either task performance or physiological responses to work. Two prominent examples are the Subjective Workload Assessment Technique (SWAT) and NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988; Reid & Nygren, 1988). Both emphasize psychological dimensions of workload, having in common items that measure subjective levels of mental effort, time load, and stress or frustration. NASA-TLX in addition assesses subjective levels of physical demand and performance.

Generally speaking, the manipulation of single-task parameters that affect performance also affect subjective workload measures (Adams & Biers, 2000; Astin & Nussbaum, 2002; Biers & Anthony, 2000; Hitchcock, Dember, Warm, Moroney, & See, 1999; Schaab & Dressel, 2001). However, a major test of the validity of subjective measures is whether they successfully predict dual-task performance. The logic is that when the load of one task is low, it should be possible to perform a second low-load task simultaneously with little interference because the total resources available are not exceeded. On the other hand, when the resource load of both tasks is high, combining the tasks should result in substantial interference attributable to resource limitations. If a subjective workload measure is valid, it should be sensitive to resource demand and predict the extent to which two tasks will interfere. In general, subjective workload ratings do change with changes in dual-task

Address correspondence to David B. Boles, Department of Psychology, P.O. Box 870348, University of Alabama, Tuscaloosa, AL 35487; dboles@bama.ua.edu. *HUMAN FACTORS,* Vol. 49, No. 1, February 2007, pp. 32–45.

load (e.g., Kilmer et al., 1988), although mixed outcomes have also been observed (Coyne & Baldwin, 2003; Vidulich & Tsang, 1985).

As useful as the SWAT and NASA-TLX measures have proved to be, they both emphasize global psychological dimensions such as effort and stress and do not attempt to assess workload within specific cognitive resources that may be available to performance, such as those associated with particular perceptual or response processes. Yet theoretical bases exist for a more specific approach. One such basis is Wickens's (1984) multiple resource theory, which states that specific resources are devoted to codes (verbal and spatial, vocal and manual), modalities (visual and auditory), and stages (encoding/central processes and response processes). An extension of this theoretical basis was provided by Boles and Law (1998), who built on Wickens's (1984) seminal approach by arguing that every mental process possesses resources, a conclusion they supported inductively with dual-task data.

Thus an alternative approach to subjective workload assessment would be a measurement instrument that assesses multiple mental resources independently. The ideal instrument of this type would measure every independent resource. However, if Boles and Law (1998) were correct that every mental process possesses resources, the total number of resources is likely to be very large.

The sheer number of potential specific resources is therefore an impediment to the development of a resource-based alternative. Nevertheless, any reasonably diverse selection of resources based on process independence is likely to produce benefits in terms of the sensitivity and diagnosticity of subjective workload measures.

Our approach is based on an extensive research program undertaken to identify independent processes from factor analysis of performance asymmetries (Boles, 1991, 1992, 1996, 1998, 2002). The underlying assumption is that resources in general, but especially perceptual processing resources, typically are asymmetrically distributed between the cerebral hemispheres. They can be detected and measured in many cases by presenting appropriate stimuli to the left or right of midline under appropriate conditions and by assessing the speed or accuracy of recognition. For example, words briefly presented to the right of midline, whether visually or auditorily, usually are better or more quickly recognized than those presented to the left of midline, an outcome reflecting left hemisphere processing of language materials in most individuals. Yet individual differences exist, with some people showing a larger right-sided advantage than others, and with some even showing a left-sided advantage. Such variation forms the basis of an approach in which sets of tasks are presented to participants, with the resultant performance asymmetries constituting variables that are then factor analyzed to uncover mental processes common to two or more tasks, as opposed to those that appear to be unique to a task.

Table 1 lists the processes identified by this approach (Boles, 1991, 1992, 1996, 1998, 2002). Although the set of processes should not be considered complete, it is certainly diverse in terms of both modality and type of mental operation. It also has one particularly strong feature with respect

TABLE 1: Processes Identified Through Factor Analysis of Performance Asymmetries

Process	Typical Tasks
Auditory emotional	Recognizing vocal emotion
Auditory linguistic	Recognizing auditory words, digits, or syllables
Facial figural	Judging face similarity or expression
Facial motive	Performing a facial gesture such as winking
Planar categorical	Judging whether one position is above or below another
Spatial attentive	Focusing attention on a position in space
Spatial concentrative	Judging the spacing of numerous visual objects
Spatial emergent	"Picking out" a visual object from a cluttered background
Spatial positional	Recognizing a visual location in space
Spatial quantitative	Judging numerical quantity represented by a bar graph or small cluster of objects
Tactile figural	Recognizing shapes using the sense of touch
Visual lexical	Recognizing visual words, letters, or multiple digits
Visual phonetic	Matching visual letters by rhymed endings
Visual temporal	Judging brief time intervals between visual objects

to human factors-related needs in that it emphasizes perceptual processes in general and spatial processes in particular, at least partially capturing a resource domain in which workload is typically high and in which dual tasks often interfere.

Largely based on the processes emerging from this work, a new subjective workload instrument called the Multiple Resources Questionnaire (MRQ) has been devised. Originally published in 2001 (Boles & Adair, 2001), we reproduce it now in the Appendix.

Of the 17 items on the MRQ, 14 are directly derived from the factor analytic work using performance asymmetries. Nearly all of the 14 represent perceptual processes, however, so it was decided to include 3 additional memory- and response-related items dealing with short-term memory, manual response, and vocal response. The justification is that dual-task performance is impacted by all three (e.g., Adams & Biers, 2000; Fracker & Wickens, 1989; Wickens, Sandry, & Vidulich, 1983).

The questionnaire is intended to be user-oriented in the sense that ratings are made by the users of technology rather than workload experts. Initial investigations using the questionnaire revealed moderate to substantial interrater reliability, ranging from $r = +.57$ to $+.83$ over a number of laboratory tasks and computer-based games. Further gains in reliability came from aggregating results over sets of raters, a common practice in workload assessment. It was found that when results were aggregated over 8 or more raters, reliability increased to approximately +.90 as assessed by correlating values between aggregates (Boles & Adair, 2001).

Besides being based in known mental processes, the MRQ has the advantage of being fast and easy to administer. A typical administration requires about 5 min following the performance of a task, and ratings from the MRQ are also intended to be used "as is" without further scale derivation.

The Present Study

The present study was designed to examine the validity of the MRQ as a workload measure. Specifically, we wanted to determine whether questionnaire ratings could be used to predict the amount of interference between simultaneously performed dual tasks. In principle this should be simple to do: Administer tasks singly, collect MRQ ratings after each, then assess the rated similarity between tasks and determine whether the similarity values correlate with dual-task interference.

In practice, however, there are some complexities inherent in this approach. First, it is not clear which of the three similarity metrics computationally demonstrated in Table 2 should be used. One possibility is to assess the squared correlation between the ratings for the two tasks, across the 17 items of the MRQ, as a measure of variance shared. Because this approach quantifies task similarity in the "peaks and valleys" of demand across resources, we call it *profile similarity.* A second possibility is to determine the minimum of task demands on each resource as a measure of competition for it, followed by summation across resources. We call this *overlap similarity.* Finally, the total demand of both tasks could be assessed by summing the ratings over both tasks and all the resources, a measure we call *overall demand.*

Further complexity arises in deriving a measure of dual-task interference. One can either examine interference on each task separately or combine

TABLE 2: Computation of Three Similarity Metrics for One Participant Across Four Resources

	Rating			
Resource	Task 1	Task 2	Minimum	Sum
Auditory linguistic	0	1	0	1
Spatial attentive	4	3	3	7
Spatial emergent	3	3	3	6
Visual lexical	2	0	0	2
Similarity metric:	$r^2 = +.47$ Profile similarity		Sum = 6 Overlap similarity	Sum = 16 Overall demand

Note. In actual computations all 17 resources are used.

interference across tasks as a joint measure. As we developed our experimental design, we realized that although our dual tasks would be performed simultaneously, the responses would be asynchronous. We also realized that we had no basis for asking participants to emphasize one task over another, and because task emphasis was not manipulated, a performance operating characteristic approach (Navon, 1984) could not be used. To us, these considerations argued for a single joint measure of interference as a means of smoothing out performance variations due to accidental couplings and decouplings in response demands and idiosyncratic choices among participants as to which task to emphasize. We therefore elected to use a measure of interference that averages over both tasks in a dual-task pairing, which we call *ensemble interference.*

In Experiment 1, we investigated the viability of the MRQ as a predictor of interference between relatively simple laboratory tasks. Experiment 2 examined more complex computer-based games. In both experiments, we made the assumption that the degree to which two tasks demand the same resources should predict the amount of interference between them when they are performed together.

EXPERIMENT 1: METHOD

Participants

A total of 24 participants completed the study, all of them students enrolled in introductory psychology courses. They signed an informed consent form and received a debriefing as well as course credit following their completion of the tasks. The participants were randomly assigned to four groups (6 participants each) that determined which subset of three tasks they received in both the single-task and dual-task sessions, of the total set of four tasks. Thus Group 1 received bar graphs, crosslines, and letters tasks; Group 2 received bar graphs, crosslines, and words tasks; Group 3 received crosslines, letters, and words tasks; and Group 4 received bar graphs, letters, and words tasks. This omission of one task per group was an intentional aspect of the design, because any correlation found between task similarity and task interference over all four groups could not then be attributed to a fortuitous combination of tasks.

Apparatus

The tasks had been used in previous factor analytic work (Boles, 1991, 1992, 1996, 1998, 2002) and were programmed for Apple II computers running the Apple-Psych system of experimental control (Barnes & Burke, 1988). Two such computers were arranged side by side, displaying stimuli on 12-inch (30.5-cm) amber monitors that were pushed together until touching. The physical separation between the centers of the two monitors was 33 cm. Responses were made on external keyboards, each oriented with one key located closer to and one key farther from the participant. In single tasks, participants used the leftmost computer and monitor and their choice of response hand. In dual tasks, two keyboards were arranged for use by the left and right hands, each hand responding using the computer and monitor on the same side.

Stimuli and Procedure

The experiment was run in two sessions, the first for single tasks and the second for dual-task pairings. In most cases the sessions were 1 day apart.

Single-task trials. In all tasks, a trial began with a central fixation cross presented for 750 ms, followed by a 100-ms blank period, and then a display of stimuli until the participant either gave a response or a deadline of 6 s passed. Brief feedback was shown giving the reaction time (RT) if the response was correct; the word "ERROR" appeared if it was incorrect or if the deadline passed. This was followed by an intertrial interval of 500 ms (bar graphs, crosslines, and words tasks) or 800 ms (letters task). Displays consisted of two randomly selected stimuli, one to the left of the screen and one to the right, with an arrowhead ("<" or ">") between them. The stimulus to be responded to was indicated by the arrowhead. Participants were instructed to respond both as quickly and as accurately as possible.

It should be pointed out that the experiment was not intended to look at the effects of cerebral lateralization per se, and so there was no attempt to enforce eye or head position. Arrowheads pointing to one stimulus were included only because they were used in the original lateralization research employing the tasks (Boles, 1991, 1992, 1996, 1998, 2002), and we did not want to change that aspect of the displays when it was not necessary to do so.

Other than the nonenforcement of eye or head position, the main change from the previous factor analytic work using the tasks was that the stimuli remained present until the response or deadline,

the earlier work having used brief presentations. The change was a practical necessity of the dual-task design, so that participants could not easily miss a stimulus because of momentary attention to the other display. The response deadline was also lengthier for the same reason.

For each task, either 144 trials (bar graphs, crosslines, words) or 128 trials (letters) were administered. The intertrial interval in each task was selected to allow all four tasks to end in approximately 15 min. After each task was completed, the participant completed the MRQ to assess the resources used by the task.

The bar graphs task involved recognizing the whole-number content of a vertical bar graph and deciding whether it represented an odd or even number. A bar graph (Figure 1) consisted of a vertical rectangle plotted against unlabeled reference lines at the 0, 4, and 8 levels and took on a value in the range from 1 through 8. A bar graph was 2.7° × 7.9° in horizontal by vertical extent, with an eccentricity of 2.6° as measured from the fixation point to the near edge. Participants decided whether the bar graph was odd or even and pressed a corresponding key.

A crosslines stimulus (Figure 1) consisted of a short horizontal line segment, 1.1° in horizontal extent, presented 1.6° above or below an imaginary horizontal line through the central fixation cross, at an eccentricity of 2.1° as measured from the near edge to the vertical meridian. The bilateral displays therefore consisted of two "up" line segments, two "down" segments, or one of each, with an arrowhead indicating the segment whose position was to be judged. Participants pressed either the key located away from them for an "up" judgment or the one toward them for a "down" judgment.

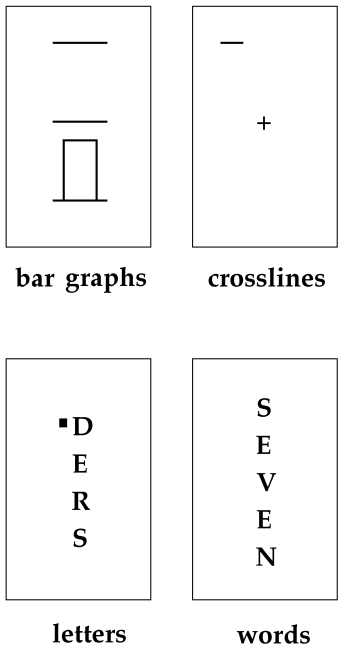


Figure 1. Stimuli used in Experiment 1 (not drawn to scale).

A letters stimulus (Figure 1) was composed of a four-letter string, one letter of which was marked for vowel-consonant classification. Sixteen strings were used, half words and half nonwords, and vowels and consonants were equiprobable at all four positions. A string was arrayed vertically, with 1.1° × 6.6° horizontal by vertical extent, at 3.9° eccentricity as measured from the fixation point to the near edge. The marker was a small dot adjacent to a single target letter, on the side nearer the screen's center. Participants decided whether the target letter was a vowel or consonant and pressed a corresponding key.

In the words task (Figure 1), sometimes called *word numbers* or *visual word numbers* in previous publications (Boles, 1991, 2002), a stimulus consisted of the word name of a number (e.g., "ONE") from one through eight. It was presented vertically with 0.6° horizontal extent, and depending on the length of the word, from 2.7° to 4.7° vertical extent. Eccentricity was 2.1° as measured from fixation to the near edge. Participants decided whether the word represented an odd or even number and pressed a corresponding key.

Dual-task trials. A different subset of three of the four tasks was assigned to each of the participant groups. Among the three, all possible pairings were made, and the pair members were run simultaneously as dual tasks. The ordering of the dual-task pairings was completely balanced over participants, with balancing of the pair members to the left or right computer system nested within that balancing. The number of trials and provision of feedback were as in the single-task trials.

Although all four tasks were of approximately 15 min duration, the programs nevertheless ran independently of one another and, when paired in dual tasks, did not generally end at exactly the same time. Accordingly our design called for analyzing only the data from the first half of the trials from each task, a cut point sufficient to ensure that all

analyzed data were collected under dual-task conditions. Participants were instructed to emphasize both tasks equally.

Data Reduction

Data from the first (single-task) session were initially reduced by calculating median correct RTs and percentage errors for each task. Those from the second (dual-task) session were handled similarly, except only the first half of the trials from each task pairing were used. For each task of a given dual-task pairing, an RT interference measure was calculated by subtracting single-task RT from dual-task RT. Ensemble interference was then calculated by summing these interference measures across both tasks of the dual-task pair. The result for each participant was a measure of ensemble interference in RT, generated by each of the three dual-task pairings. A similar procedure was followed for the percentage errors, likewise resulting in a measure of ensemble interference for each of the dual-task pairings.

For each participant and each dual-task pairing, a *profile similarity* metric was calculated, which is the amount of variance shared by the two tasks. This was calculated as r^2 across the item ratings from the MRQ. An *overlap similarity* metric was also calculated, focusing on the minimum rating for each item on the MRQ. For example, if ratings of 1 and 4 were given for two tasks on an item, the overlap on that item was 1, reflecting a small degree of competition for that resource. (Note that the maximum value cannot be used because it does not reflect competition for resources; for example, a task rated 0 for *no usage* does not compete even if the other task is rated 4 for *extreme usage*). Summing these minimum values across the 17 items then yielded the overlap similarity metric. Finally, an *overall demand* metric was calculated simply as the grand sum of all 17 ratings for the two tasks.

For each participant, ensemble RT interference was correlated to each similarity measure using Pearson's *r,* which was then converted to a Z_r score using the *r*-to-*Z* transformation (Glass & Stanley, 1970). For example, if a hypothetical participant showed ensemble RT interference of 182, 662, and 469 ms across the three task pairings and corresponding overlap similarity of 18, 19, and 21, the correlation would be $r = +.43$, and using the *r*-to-*Z* transformation, $Z_r = +0.46$. The same was followed for percentage error interference.

The final result of the data reduction procedures was a set of six Z_r scores for each participant, each reflecting the size of the relationship, across the three dual-task pairings experienced by that participant, between a given similarity metric (profile, overlap, or overall demand) and the amount of ensemble interference in each measure (RT or percentage errors).

EXPERIMENT 1: RESULTS

Table 3 shows the mean ratings of each task across the 17 resources of the MRQ, with the three highest ratings for each emphasized in italic type.

Data from 3 participants (1 each in Groups 1, 3, and 4) proved unusable because a metric produced the same value across all three task pairings, resulting in an undefined correlation with ensemble interference. The mean Z_r scores for the remaining 21 participants appear in Table 4, along with their equivalent *r* values as determined using the *Z*-to-*r* backtransformation. The Z_r values were analyzed using a within-subject ANOVA with the factors of metric (three levels) and measure (two levels). A significant source of variation was the constant elevation of values above zero, $F(1, 20) = 5.56$, $p < .05$ (i.e., the T1 source of variance; Dienes, 2000). Equivalently, the mean Z_r value across all six values (3 levels × 2 measures) can be calculated for each participant and subjected to a one-sample *t* test against zero, with the result $t(20) = 2.36$, $p < .05$. These tests indicate that collectively speaking, the Z_r values shown in Table 4 are significantly positive. However, the ANOVA shows that there was no effect of metric, measure, or their interaction, all $p > .05$, results indicating that the single negative *Z* value in Table 4 is of no particular significance.

Finally, to gain some idea of the degree to which task interference varied as a function of task similarity, we calculated a single regression between overlap similarity and average RT interference across the six pairings of tasks (i.e., collapsing the data over all participants). Overlap was selected as the similarity metric for this purpose because of its slightly (though nonsignificantly) better fits than profile similarity, as evident in Table 4. The regression equation was RT interference = (91.8 × overlap) – 1222. The equation indicates that an increase of a single unit of overlap (i.e., a single rating point) increased summed RT interference by 92 ms. A regression equation based on errors

TABLE 3: Mean Resource Ratings for the Four Tasks in Experiment 1

Resource	Bar Graphs	Crosslines	Letters	Words
Auditory emotional	0.11	0.28	0.17	0.17
Auditory linguistic	0.17	0.06	0.67	0.67
Facial figural	0.00	0.17	0.06	0.17
Facial motive	0.39	0.11	0.44	0.61
Manual	*3.61*	*3.72*	*3.50*	*3.72*
Short-term memory	2.17	2.17	1.83	1.56
Spatial attentive	*3.61*	*3.78*	*3.89*	*3.72*
Spatial categorical	2.78	*3.44*	2.33	2.78
Spatial concentrative	2.00	1.44	0.78	1.06
Spatial emergent	1.11	1.06	2.22	0.72
Spatial positional	2.39	2.61	2.61	1.50
Spatial quantitative	*3.89*	0.89	0.94	1.28
Tactile figural	0.11	0.56	0.33	0.33
Visual lexical	0.72	0.67	*3.72*	*3.78*
Visual phonetic	0.11	0.33	0.78	1.00
Visual temporal	1.67	2.00	1.39	2.06
Vocal	0.17	0.06	0.06	0.33

Note. For each, the three highest ratings are shown in italic type.

indicates that an increase of a single unit increased summed percentage errors by 2.4%.

EXPERIMENT 1: DISCUSSION

The results indicate that the MRQ significantly predicted dual-task interference, with any differences between the metrics and measures either nonexistent or below the limits of statistical detectability. The mean Z_r value across metrics and measures in Table 4 is +0.39, corresponding to an *r* value of +.37. This can be taken as a representative measure of relationship between the predictions of the MRQ and the observed interference between the dual tasks used in Experiment 1.

Table 3 is encouraging in showing, in descriptive terms, that the MRQ appears to capture differences between tasks in specific resource demands. Unfortunately the statistical significance of the differences cannot easily be assessed because no participant received all four tasks. Different subsets of participants received different combinations of tasks, a design aspect that makes the data set neither fully dependent nor fully independent. Nevertheless, based strictly on an inspection of the means, the results are encouraging in suggesting that whereas all four tasks heavily demand manual and spatial attentive resources, there is a divergence between the verbal and spatial tasks, with letters and words requiring the visual lexical resource but the bar graphs and crosslines tasks requiring varying spatial resources. Specifically, the bar graphs task appears to make high demands on the spatial quantitative resource, whereas the crosslines task seems to require spatial categorical resources to a high degree.

On one hand, the rating results are not surprising in that the four tasks were used in the

TABLE 4: Mean Z_r Scores and Their Equivalent *r* Values From Experiment 1

	Measure			
	RT		Percentage Errors	
Metric	Z_r	*r*	Z_r	*r*
Profile similarity	+0.30	+.29	–0.13	–.13
Overlap similarity	+0.43	+.41	+0.62	+.55
Overall demand	+0.66	+.58	+0.44	+.41

original factor analytic studies that led to the identification of the MRQ resources. On the other hand, the fact that naive raters were apparently able, in effect, to reproduce factor structure based only on questionnaire descriptions of each resource reflects positively on the validity of the MRQ.

From a human factors perspective, the results of Experiment 1 suggest that the MRQ may prove to be both a sensitive and diagnostic tool for assessing workload. Sensitivity is indicated by the observation that increasing resource similarity, as assessed by the MRQ, results in increasing dual-task decrements. Diagnosticity is indicated by the observation that MRQ items appear to identify specific bottlenecks in performance, in this case in the manual, spatial, and visual lexical resources.

One limitation of Experiment 1, however, was that the tasks were relatively simple and had themselves previously served as the source of the factor analytic data that gave rise to the MRQ. Extension of the findings to other, more complex tasks was necessary to establish the usefulness of the questionnaire. A second limitation was that the statistical significance of differences between tasks in the use of specific resources could not be assessed because different subgroups of participants received different sets of tasks. In Experiment 2 we investigated the questionnaire's ability to predict interference between computer-based games, and all of the participants received all of the games.

EXPERIMENT 2: METHOD

Participants

Thirty-one undergraduates in psychology courses passed criteria for single-task performance, described later, and participated in the dual-task conditions, constituting the final sample. An additional 23 undergraduates did not advance to the dual-task conditions because they did not meet the criteria. All received course credit for their participation.

Apparatus

Apple Macintosh computers were used in both experiments, with keyboards and mice. In dual-task trials, side-by-side 14-inch (35.6-cm) color monitors were used, with their sides touching (centers 36 cm apart).

Stimuli and Procedure

The experiment was run as a single-task session followed by a separate dual-task session. In most cases the sessions were 1 day apart.

Single-task trials. Three computer-based games were used – Greebles, Super Maze Wars, and Word Tracer – all of which were available as shareware or freeware. The games were selected because all could be used on Macintosh computers and could be played with a specified 3-min time limit.

In Greebles, a two-dimensional game, a tank-like character is navigated around a maze using up, down, left, and right arrow keys on the standard keyboard while avoiding attacks from bug-like "greebles." Scoring is achieved by pushing against blocks that then move, crushing greebles between other blocks, and by collecting bonuses scattered throughout the maze. Destroying all the greebles results in advance to another round, which consists of a new maze with different greebles.

In Super Maze Wars, the player looks out of a hovercraft's window at the walls and floor of a maze and navigates it using the arrow keys while avoiding attacks from an enemy craft. The player attempts to run over and thus collect golden pyramids that oscillate up and down off the floor. The enemy craft can also be attacked by "firing" with the space key, and its destruction as well as the collection of pyramids scores points.

In Word Tracer, a random 6 × 6 matrix of letters is shown, and the player attempts to form as many unique words as possible by linking adjoining letters. This involves navigating the mouse to a letter, clicking, then navigating to the next letter, and so forth, until the word is completed. The player then clicks an on-screen button to enter the word. Letters adjoin in up, down, left, right, and diagonal directions. Points are awarded based on the length of the word.

The order of games was as evenly counterbalanced as possible across participants. Participants engaged in repeated 3-min bouts of the game until a scoring criterion was reached (i.e., a preset minimum number of points achieved as a running average over three successive games – specifically 1000 points for Greebles, 7 points for Super Maze Wars, and 10 points for Word Tracer), after which they progressed to the next game. The game scoring criteria were set during pilot testing to be high enough that dual-task decrements could subsequently be observed, as it was found that absent criteria, some performances continued to improve during dual-task game play, voiding the logic of the study. Such criteria were not needed in Experiment

1 because the simpler tasks used in that study became sufficiently well practiced within the single-task session. When participants successfully achieved the criterion for a particular game, they completed the 17-item MRQ.

Dual-task trials. All possible pairings of the three games were made, and the pair members were run simultaneously as dual tasks. The assignment of pair members to the left or right computer system and the ordering of the dual-task pairings were balanced as evenly as possible, given the odd number of participants. Participants were instructed to emphasize both games equally.

All games were of the same 3-min duration, so the scores were recorded on completion of both games in a dual-task pair.

Data Reduction

Because the games operated using very different numerical scales (i.e., scores typically in the low 1000s for Greebles, below 10 for Super Maze Wars, and between 10 and 20 for Word Tracer), score intervals had different meanings across games and thus the subtractive procedure used for the RT tasks of Experiment 1 was not appropriate. Following precedent from a similar situation, it was decided instead to use a percentage decrement score (Noy, 1990). This was simply the decrement between single- and dual-task scores, expressed as a percentage of the single task score.

Otherwise the data reduction proceeded in a manner similar to that of Experiment 1. Ensemble interference was calculated by summing the percentage decrements across both tasks of a dual-task pair, and values for profile similarity, overlap similarity, and overall demand were derived from the MRQ data.

EXPERIMENT 2: RESULTS

Table 5 shows the mean ratings of each task in Experiment 2 across the 17 resources of the MRQ, with the 3 highest ratings for each emphasized in italic type.

The mean Z_r scores were +0.72 for the profile similarity metric, +1.17 for the overlap similarity metric, and +0.14 for the overall demand metric. These values correspond to correlations (r) of +.62, +.83, and +.14, respectively, representing the relationship between each MRQ-based similarity metric and interference between tasks.

The Z_r values were analyzed using ANOVA with the single within-subject factor of metric (three levels). A violation of sphericity was detected using the Mauchly sphericity test, $W = 0.54$, $p < .001$, so the degrees of freedom of the ANOVA were adjusted using the Box correction with the Greenhouse-Geisser epsilon value. The effect of metric nevertheless proved to be significant, $F(1.37, 41.1) = 3.87$, $p < .05$. The effect was further examined through planned pairwise comparisons using t tests. It was found that the overlap similarity metric produced significantly higher correlations than did either the profile similarity metric, $t(30) = 2.09$,

TABLE 5: Mean Resource Ratings for the Three Tasks in Experiment 2

Resource	Greebles	Super Maze Wars	Word Tracer
Auditory emotional	0.68	0.45	0.23
Auditory linguistic	0.10	0.16	1.42
Facial figural	0.52	0.45	0.26
Facial motive	0.35	0.32	0.45
Manual	*3.23*	*3.26*	2.19
Short-term memory	1.32	1.74	2.19
Spatial attentive	*3.23*	*3.42*	*3.10*
Spatial categorical	*3.23*	*3.32*	2.55
Spatial concentrative	2.77	2.35	1.35
Spatial emergent	2.26	2.32	*2.94*
Spatial positional	2.39	3.00	2.00
Spatial quantitative	1.00	0.97	0.90
Tactile figural	0.65	0.61	0.10
Visual lexical	0.52	0.29	*3.94*
Visual phonetic	0.19	0.29	1.42
Visual temporal	1.87	1.90	1.29
Vocal	0.06	0.10	0.06

Note. For each, the three highest ratings are shown in italic type.

$p < .05$, or the overall demand metric, $t(30) = 2.45$, $p < .05$. The profile similarity and overall demand metrics did not differ significantly from each other, $t(30) = 1.34$, $p > .05$.

Paralleling Experiment 1, the data were aggregated over participants, and a regression was calculated between similarity and average percentage decrement across game pairings. Overlap similarity was selected as the appropriate metric because of its significantly superior prediction of interference. The regression equation was percentage decrement = (2.44 × overlap similarity) + 3.78. The equation indicates that as overlap between tasks increased by one unit, the summed percentage decrement increased by 2.4%.

EXPERIMENT 2: DISCUSSION

As in Experiment 1, in Experiment 2 we tested the predictive validity of the MRQ by first collecting resource ratings in an initial single-task session. From those, three predictors of dual-task performance were derived (profile similarity, overlap similarity, and overall demand). The final determination involved the degree to which each predictor correlated to actual dual-task performance as measured by ensemble interference. The overlap similarity metric produced a correlation of $r = +.83$ to ensemble interference. If anything, this value is larger than that found for the simpler laboratory tasks of Experiment 1, suggesting that the predictive validity of the MRQ does not decline as the task environment becomes more complicated. Thus the results of Experiment 2 indicate that the MRQ successfully predicts dual-task interference when fairly complex computer games are the simultaneously performed tasks.

An additional finding from Experiment 2 was that the overlap similarity metric significantly outperformed both the profile similarity and overall demand metrics. Thus substantial progress has been made over the undifferentiated outcome of Experiment 1, in which overlap similarity did not outperform the other metrics.

The mean ratings in Table 5 appear very descriptive of the games, and the fully within-subject design allows assessment of the significance of differences in resource usage. For each of the three highest rated resources for each game, we conducted a post hoc one-way ANOVA comparing mean ratings across games and, if that was significant, followed up with pairwise *t* tests. For the manual resource, games differed significantly, $F(2, 60) = 18.30$, $p < .001$, or when corrected for violation of sphericity, $F(1.51, 45.2) = 18.30$, $p < .001$, with Word Tracer making less demand than either Greebles, $t(30) = 4.50$, $p < .001$, or Super Maze Wars, $t(30) = 4.90$, $p < .001$, and with the latter two not differing significantly, $t(30) = 0.23$. Continuous maze navigation in Greebles and Super Maze Wars presumably accounts for the higher manual resource ratings of these tasks relative to Word Tracer, which involved periods of manual inactivity as word search progressed.

All three games were highly intensive of spatial attention, requiring close concentration on screen locations, and did not differ significantly from each other in that respect, $F(2, 60) = 1.35$, $p > .05$. However, the games differed significantly in spatial categorical ratings, $F(2, 60) = 6.58$, $p < .01$, or when corrected for violation of sphericity, $F(1.47, 44.1) = 6.58$, $p < .05$, with both Greebles and Super Maze Wars producing higher ratings than Word Tracer, $t(30) = 3.16$, $p < .01$, and $t(30) = 2.65$, $p < .05$, respectively, but not themselves differing significantly, $t(30) = 0.55$. Greebles and Super Maze Wars presumably involved higher use of the resource because of continuous decisions to turn left or right or go forward or backward, whereas such decisions were more intermittent for Word Tracer.

However, Word Tracer involved higher ratings of the spatial emergent resource, $F(2, 60) = 4.33$, $p < .05$, with no violation of sphericity, $W = 0.98$, $p = .71$, than did Greebles, $t(30) = 2.49$, $p < .05$, or Super Maze Wars, $t(30) = 2.56$, $p < .05$, those two games not differing significantly, $t(30) = 0.26$. This was no doubt attributable to the need to "pick out" letters from the letter matrix. Similarly, given the need to form words, Word Tracer involved higher use of the visual lexical resource, $F(2, 60) = 430.56$, $p < .001$, with no violation of sphericity, $W = 0.83$, $p > .05$, than did either Greebles, $t(30) = 21.49$, $p < .001$, or Super Maze Wars, $t(30) = 33.37$, $p < .001$, with the latter two not differing, $t(30) = 1.56$, $p > .05$.

According to these ratings outcomes, the MRQ appears to enjoy considerable face validity.

GENERAL DISCUSSION

The MRQ is an easily administered 17-item measure of subjective workload based on multiple resource concepts. Although previous research

has indicated that the MRQ exhibits acceptable reliability (Boles & Adair, 2001), to be useful the measure should also exhibit validity by predicting changes in task performance following changes in task parameters. Because our emphasis is on dual-task performance, we conducted two experiments to determine whether the MRQ successfully predicts decrements in performance when simultaneously performed tasks are paired in varying combinations.

Together, the results of the experiments support the predictive validity of the MRQ. Correlations of predicted to actual decrements were as high as $r = +.55$ in Experiment 1 and $r = +.83$ in Experiment 2. Significant correlations were observed regardless of whether the tasks were relatively simple laboratory tasks requiring discrete responses or more complex computer-based games requiring continuous performance.

Sequential Shifts in Multitasking

A potential criticism of our experimental paradigm is that the side-by-side nature of the tasks and their uncorrelated timing may have encouraged a sequential shifting of attention between the tasks rather than simultaneous processing. Given the placement of the displays and controls, we have no doubt that some sequential shifting occurred.

Nevertheless, for four reasons we believe the point carries little weight. First, Boles and Law (1998) used single displays requiring simultaneous recognition of two brief (100-ms) time-locked stimuli and found that dual-task interference was predicted by the previously established factor structure on which both their experiments and the current experiments were based. Thus the general approach works regardless of whether single or dual displays are used or whether or not perceptual events are time locked.

Second, we believe that sequential shifting is a universal aspect of multitasking. Psychological refractory period research has shown that the simultaneous selection of two responses is an inescapable bottleneck even if different response modalities are used (Ruthruff, Johnston, & Van Selst, 2001). Wickens (1984) explicitly recognized the response bottleneck problem by incorporating stages-of-processing resources in his model, predicting less interference when encoding and central processing can proceed for one task while responding to another task, compared with the attempt to simultaneously respond to both tasks. The bottleneck created by simultaneous response demands means that to some extent, there is a sequential element to all multitasking that involves separate responses, and our paradigm is not very different in this regard.

The third reason we view the sequential criticism as having little impact is that practically speaking, much real-world multitasking is carried out under conditions similar to those of our experiments. The most frequently mentioned multitasking situations in the literature are high-performance flight, air traffic management, and nuclear control. All are characterized by multiple visual displays, presented simultaneously or serially, requiring shifting of attention. It is reasonable to ask under what conditions interference will be minimized in these real-world applications. We believe using dual displays with uncorrelated events is as true a multitasking situation as a pilot having to read instruments while visually scanning terrain, an air traffic controller consulting multiple radar displays, or a nuclear power plant operator examining multiple dials that display control parameters.

The preceding point leads to our fourth reason, which may be the most important. This is that so far, the MRQ approach works. But only if the MRQ is disseminated and applied will it be possible to assess the true extent to which its value is generalizable.

The Optimum Metric

In the course of addressing the predictive validity of the MRQ, considerable progress has been made toward identifying the optimum metric of task similarity to which performance decrements can be related. Experiment 2 showed that overlap similarity, a measure of summed minimum demands across resources, outperforms both the summed demand of dual tasks on resources and profile similarity, a measure of correlation between resource demands.

The fact that the overlap similarity metric outperformed summed demand is theoretically meaningful. The outcome strongly indicates that total resource demand on the system is less important than resource-by-resource correspondences between tasks. Presumably such correspondences indicate bottlenecks in specific mental resources that degrade dual-task performance.

When task resource ratings are explicitly laid out, as in Tables 3 and 5, the nature of the bottlenecks becomes clear. Such tables effectively represent action plans for task redesign through the shifting of mental work onto different resources. For example, the results in Table 5 allow clear predictions that less interference should be observed (a) if spatial attentive demands could be reduced across the board, for example, by using integrated dual-task displays; and (b) if one member of a task pairing could be shifted from manual to vocal control. Of course, in many cases it will be important when altering designs not to purchase improved dual-task performance at the cost of single-task decrements, as, for example, a shift from manual to vocal responding might entail.

Tables of resource usage really provide only suggestions for redesign. Nevertheless, the specificity shown in Tables 3 and 5 provides considerably greater guidance than the cognitively nonspecific dimensions of other subjective workload instruments such as the SWAT and NASA-TLX.

Conclusion

The MRQ appears to hold considerable promise for the subjective measurement of workload. It is sensitive to changes in dual-task pairings, and the workload estimates that are produced are diagnostic of bottlenecks in dual-task performance. In our view, potential applications of the MRQ include the identification of bottlenecks in dual tasks as well as process overloads in single tasks, preparatory to redesign in high-workload areas such as air traffic management (Metzger & Parasuraman, 2005), advanced flight displays (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003), and medical imaging (Klein, Riley, Warm, & Matthews, 2005).

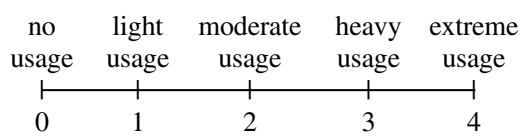
Nevertheless, there also should be some recognition of the limits of the MRQ. Given the assumption that all mental processes have resources (Boles & Law, 1998), it cannot be argued that the MRQ measures resource usage across all possible processes. We encourage flexible use of the instrument, which can include adding items to reflect independent resources not otherwise included, or elimination of items that are obviously irrelevant to a particular task domain. The general approach to workload measurement that we have described should work reasonably well as long as valid multiple resource considerations guide changes to the instrument. Certainly, any derivative of the MRQ should be simple enough to use that its validity can be checked by doing as we have done, correlating a task similarity metric to observed dual-task interference.

A second limitation to the MRQ is that as with other subjective measures, it should be regarded as providing a relative and not absolute measure of workload. Empirically, regressions of task similarity onto task interference show a great deal of variability across individuals, suggesting an idiosyncratic component to individuals' workload estimates (Boles, Phillips, Bursk, & Perdelwitz, 2004). Beyond the question of individual differences, however, there is reason to believe that interference in specific resources accounts for only a portion of dual-task interference. Thus when laboratory tasks are paired, RTs increase beyond what can be accounted for by characteristics of the tasks themselves (Boles & Law, 1998), probably indicating that there are either coordination costs to dual-task performance (Friedman, Polson, Dafoe, & Gaskill, 1982) or some type of generalized resource that produces interference regardless of task characteristics (Wickens, 1984, p. 305). Either alternative is likely to impose a limit on how absolute a measure of workload can be if it is based on specific resources. For these reasons, the regression equations developed as descriptions of the results of Experiments 1 and 2 should be viewed as just that – descriptions – and not as potentially universal relationships that could be expected to hold true across all task domains.

APPENDIX: THE MULTIPLE RESOURCES QUESTIONNAIRE (MRQ)

MULTIPLE RESOURCES QUESTIONNAIRE for task _____

The purpose of this questionnaire is to characterize the nature of the mental processes used in the task with which you have become familiar. Below are the names and descriptions of several mental processes. Please read each carefully so that you understand the nature of the process. Then rate the task on the extent to which it uses each process, using the following scale.



Important:

All parts of a process definition should be satisfied for it to be judged as having been used. For example, recognizing geometric figures presented visually should **not** lead you to judge that the "Tactile figural" process was used, just because figures were involved. For that process to be used, figures would need to be processed tactilely (i.e., using the sense of touch).

Please judge the task as a **whole**, averaged over the time you performed it. If a certain process was used at one point in the task and not at another, your rating should **not** reflect "peak usage" but should instead reflect **average** usage over the entire length of the task.

Auditory emotional process – Required judgments of emotion (e.g., tone of voice or musical mood) presented through the sense of hearing. ____

Auditory linguistic process – Required recognition of words, syllables, or other verbal parts of speech presented through the sense of hearing. ____

Facial figural process – Required recognition of faces, or of the emotions shown on faces, presented through the sense of vision. ____

Facial motive process – Required movement of your own face muscles, unconnected to speech or the expression of emotion. ____

Manual process – Required movement of the arms, hands, and/or fingers. ____

Short-term memory process – Required remembering of information for a period of time ranging from a couple of seconds to half a minute. ____

Spatial attentive process – Required focusing of attention on a location, using the sense of vision. ____

Spatial categorical process – Required judgment of simple left-versus-right or up-versus-down relationships, without consideration of precise location, using the sense of vision. ____

Spatial concentrative process – Required judgment of how tightly spaced are numerous visual objects or forms. ____

Spatial emergent process – Required "picking out" of a form or object from a highly cluttered or confusing background, using the sense of vision. ____

Spatial positional process — Required recognition of a precise location as differing from other locations, using the sense of vision. ____

Spatial quantitative process – Required judgment of numerical quantity based on a nonverbal, nondigital representation (for example, bar graphs or small clusters of items), using the sense of vision. ____

Tactile figural process – Required recognition or judgment of shapes (figures), using the sense of touch. ____

Visual lexical process – Required recognition of words, letters, or digits, using the sense of vision. ____

Visual phonetic process – Required detailed analysis of the sound of words, letters, or digits, presented using the sense of vision. ____

Visual temporal process – Required judgment of time intervals, or of the timing of events, using the sense of vision. ____

Vocal process – Required use of your voice. ____

ACKNOWLEDGMENTS

We would like to thank Jenny Abernathy, Brandon Boyd, Kaz Espy, Beth Tynan, and Christina Weywadt for helping conduct the experiments.

REFERENCES

Adams, C. M., & Biers, D. W. (2000). ModSWAT and TLX: Effect of paired comparison weighting and weighting context. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6.610–6.613). Santa Monica, CA: Human Factors and Ergonomics Society.

Astin, A., & Nussbaum, M. A. (2002). Interactive effects of physical and mental workload on subjective workload assessment. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 1100–1104). Santa Monica, CA: Human Factors and Ergonomics Society.

Barnes, S., & Burke, R. S. (1988). What is Apple-Psych? *Behavior Research Methods, Instruments, and Computers, 20,* 150–154.

Biers, D. W., & Anthony, C. (2000). Context effects in the measurement of subjective workload: Prolonged exposure to the same or different contexts. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6.614–6.616). Santa Monica, CA: Human Factors and Ergonomics Society.

Boles, D. B. (1991). Factor analysis and the cerebral hemispheres: Pilot study and parietal functions. *Neuropsychologia, 29,* 59–91.

Boles, D. B. (1992). Factor analysis and the cerebral hemispheres: Temporal, occipital and frontal functions. *Neuropsychologia, 30,* 963–988.

Boles, D. B. (1996). Factor analysis and the cerebral hemispheres: "Unlocalized" functions. *Neuropsychologia, 34,* 723–736.

Boles, D. B. (1998). Relationships among multiple task asymmetries: II. A large-sample factor analysis. *Brain and Cognition, 36,* 268–289.

Boles, D. B. (2002). Lateralized spatial processes and their lexical implications. *Neuropsychologia, 40,* 2125–2135.

Boles, D. B., & Adair, L. P. (2001). The Multiple Resources Questionnaire (MRQ). In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1790–1794). Santa Monica, CA: Human Factors and Ergonomics Society.

Boles, D. B., & Law, M. B. (1998). A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 204–215.

Boles, D. B., Phillips, J. B., Bursk, J. H., & Perdelwitz, J. R. (2004). Application of the Multiple Resources Questionnaire (MRQ) to a complex gaming environment. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 1968–1972). Santa Monica, CA: Human Factors and Ergonomics Society.

Coyne, J. T., & Baldwin, C. L. (2003). Comparison of the P300 and other workload assessment techniques in a simulated flight task. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 601–605). Santa Monica, CA: Human Factors and Ergonomics Society.

Dienes, Z. (2000). *Using SPSS: Two-way repeated-measures ANOVA.* Retrieved October 17, 2006, from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/SPSS%202-way%20rm.html

Fracker, M. L., & Wickens, C. D. (1989). Resources, confusions, and compatibility in dual-axis tracking: Displays, controls, and dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 80–96.

Friedman, A., Polson, M. C., Dafoe, C. G., & Gaskill, S. J. (1982). Dividing attention within and between hemispheres: Testing a multiple resources approach to limited-capacity information processing. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 625–650.

Glass, G. V., & Stanley, J. C. (1970). *Statistical methods in education and psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Hitchcock, E. M., Dember, W. N., Warm, J. S., Moroney, B. W., & See, J. E. (1999). Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors, 41,* 365–372.

Kilmer, K. J., Knapp, R., Burdsal, C., Borresen, R., Bateman, R., & Malzahn, D. (1988). Techniques of subjective assessment: A comparison of the SWAT and modified Cooper-Harper scales. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 155–159). Santa Monica, CA: Human Factors and Ergonomics Society.

Klein, M. I., Riley, M. A., Warm, J. S., & Matthews, G. (2005). Perceived mental workload in an endoscopic surgery simulator. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 1014–1018). Santa Monica, CA: Human Factors and Ergonomics Society.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors, 47,* 35–49.

Navon, D. (1984). Resources – A theoretical soup stone? *Psychological Review, 91,* 216–234.

Noy, Y. I. (1990). Selective attention with auxilliary automobile displays. In *Proceedings of the Human Factors and Ergonomics Society 34th Annual Meeting* (pp. 1533–1537). Santa Monica, CA: Human Factors and Ergonomics Society.

Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: North-Holland.

Ruthruff, E., Johnston, J. C., & Van Selst, M. (2001). Why practice reduces dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 3–21.

Schaab, B. B., & Dressel, J. D. (2001). Training digital system skills: Increasing transfer and reducing workload. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1820–1822). Santa Monica, CA: Human Factors and Ergonomics Society.

Vidulich, M. A., & Tsang, P. S. (1985). Assessing subjective workload assessment: A comparison of SWAT and the NASA-bipolar methods. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 71–75). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus, OH: Charles E. Merrill.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45,* 360–380.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors, 25,* 227–248.

David B. Boles is a professor in the Department of Psychology at the University of Alabama. He received a Ph.D. in psychology at the University of Oregon in 1979.

Jonathan H. Bursk is the manager of Crew Systems Integration for the CH-53K program at Sikorsky Aircraft. He received an M.S. in psychology from Rensselaer Polytechnic Institute in 2003.

Jeffrey B. Phillips is a researcher at the Naval Aerospace Medical Research Laboratory, Pensacola, FL. He received a Ph.D. in psychology from the University of Alabama in 2006.

Jason R. Perdelwitz resides in Barrigada, Guam. He received an M.S. in the Department of Industrial Engineering at the University of Alabama in 2005.

Date received: December 19, 2004
Date accepted: September 7, 2006