CRITIQUE

Methodological and Theoretical Concerns in Multitask Performance: A Critique of Boles, Bursk, Phillips, and Perdelwitz

Michael A. Vidulich, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, and **Pamela S. Tsang,** Wright State University, Dayton, Ohio

Boles, Bursk, Phillips, and Perdelwitz (2007, this issue) purported to assess the validity of the Multiple Resources Questionnaire (MRQ) in predicting dual-task performance. As a relatively easily administered rating scale, if validated, the MRQ would constitute a valuable addition to the system designers' tool set for assessing the cognitive suitability of human-machine interfaces in complex environments. Unfortunately, several methodological issues rendered the validation problematic. Also, theoretical derivation of the MRQ resource structure appeared to be problematic, potentially limiting the applicability of the tool and its contribution to advancing the understanding of the underlying mechanisms of dual-task performance.

RESOURCES AND MULTIPLE-TASK PERFORMANCE

The MRQ is an attempt to produce a workload measure to predict the impact of overlapping resource demands as defined by Boles and Law's (1998) expanded multiple resource model on multiple-task performance (see also Boles, 1991, 1992). Boles et al. (2007, p. 34) stated that "Specifically, we wanted to determine whether questionnaire ratings could be used to predict the amount of interference between simultaneously performed dual tasks."

A central tenet of resource theories since Kahneman's (1973) seminal work has been that information processing resources are competed for and shared among simultaneously performed tasks. An additional tenet of multiple resource theories is that processing resources are subdivided in multiple types and that the degree of mutual interference between tasks is modulated by the degree of similarity and the overall demand of resources required by the time-shared tasks. Based on an analysis of existing dual-task data, Wickens's (1980, 1984) multiple resource model originally was defined by three resource dimensions and recently was updated to a four-dimension model (Wickens, 2002).

Using a battery of tasks and a factor analytic technique to examine the patterns of hemispheric lateralization effects, Boles and Law (1998) elaborated on Wickens's (1984) model and dramatically extended the number of resource dimensions. The tasks used in this work did not appear to represent a comprehensive multiple resource model of information processing, causing 14 of the 17 items on the MRQ derived from the earlier work to be perceptual processes. Recognizing the lack of representativeness, Boles et al. (2007; see also Boles & Adair, 2001) added three additional memory- and response-related items dealing with short-term memory, manual response, and vocal response. These additions are certainly reasonable, but reasonableness is not a scientific principle.

Boles et al. (2007) further pointed out that their list of resources (Table 1, p. 33) should not be considered complete. But they asserted, "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" (p. 43). It was nevertheless problematic because there was no mention of how their list of resources might be completed, what the inclusion criteria might be, and how one might validate additional resources. The danger of Boles et al.'s (2007) unconstrained approach is of course the well-recognized potential for resource proliferation to the point of uselessness.

Alhough Boles and Law (1998) argued that the

Address correspondence to Michael A. Vidulich, AFRL/HEC, 2255 H St., Wright-Patterson Air Force Base, OH 45433-7022; michael.vidulich@wpafb.af.mil. *HUMAN FACTORS,* Vol. 49, No. 1, February 2007, pp. 46–49.

Wickens (1984) model is overly restrictive, and it most certainly has limitations, they offered little information on how their expanded model would enhance predictability of time-sharing performance or how their MRQ tool might improve upon other multiple resource-based subjective instruments such as the Workload Index (W/INDEX; Riley, Lyall, & Wiener, 1994) and Workload Profile (Tsang & Velazquez, 1996) techniques. For example, Phillips and Boles (2004) compared the MRQ with the Workload Profile and found no significant difference between them, even though the former exhibited less variability.

In much of Wickens's (e.g., 1980) early multiple resource work, difficulty insensitivity, performance trade-offs between tasks, and a number of other checks were often used to support definitions of resource dimensions. Boles and Law (1998) acknowledged the potential value of difficulty insensitivity as an indicator of separate resources and incorporated it into their analyses. But their results were mixed, supportive of their presumed resource structure only some of the time. It seems advisable that attempts to validate the MRQ should meet the same criteria that generally have been used to define and develop resource architectures in the literature.

VALIDATION OF THE MRQ TOOL

There were two major methodological concerns in Boles et al.'s (2007) MRQ validation process. One concern was whether simultaneous performance occurred to the extent that would call for a resource explanation. A second concern was whether the experimental tasks used would reveal the workings of resources.

Sequential Shifts in Multitasking

A major methodological concern was that the MRQ tool was not tested in an environment amenable to simultaneous task processing and thereby to tests of resource structures. Two experimental tasks were presented on two side-by-side monitors 33 cm apart. Further, the two tasks were run by independent unsynchronized programs. The resulting visual scanning demands and unconstrained timing between tasks could not ensure that the participants were truly time-sharing the tasks in the same sense that resource researchers would use the term *time-sharing* (i.e., simultaneous cognitive processing of two or more tasks). It was just as possible that the participants simply shifted attention from one task to another sequentially. If so, the observed effects might have nothing to do with any multiple resource structure. Boles et al. (2007) acknowledged that the task arrangements might have fostered attention shifting rather than attention sharing. They raised four reasons for dismissing this methodological problem. We revisit those reasons here and point out the importance of both theoretical and methodological considerations in the development and validation of practical tools.

First, Boles et al. (2007) reported observing a similar pattern of dual-task interference as predicted by their multiple resource model when they used a single display of two time-locked stimuli in their earlier research (Boles & Law, 1998). But the fact that dual-task interference was found in both the earlier and current studies does not necessarily mean that the interference in the two studies was caused by the same mechanism. This was especially the case when the nature of the task demands was quite different in the two sets of studies. Dual-task interference could result from a number of factors, only one of which was resource competition. Without other corroborating information, it is not possible to ascertain the true underlying mechanism of the interference observed in Boles et al. (2007).

Second, Boles et al. (2007) asserted that sequential shifting is a universal aspect of multitasking. They specifically invoked Wickens's multiple resource model to support their claim, "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" (p. 42). This portrayal of Wickens's (1984) response-processing resources (or any processing resources) sorely missed the theoretical distinction between a bottleneck and a resource limitation of multiple-task performance. A response bottleneck prohibits more than one response selection (assuming that the bottleneck is at the response selection stage; e.g., Pashler, 1984) at any given time, and thus it mandates sequential processing at the bottleneck stage. In contrast, Wickens's (1984) multiple resource model does permit simultaneous processing of multiple responses, albeit at reduced efficiency if the combined demands exceed the total response processing resources available. Note that Wickens's (1984) model also permits sequential processing. Sequential processing is necessary or strategically advantageous if the demand of one task is extremely high and it is important to devote all of one's resources to maintain a certain level of performance of that task. Sequential processing is obligatory only with the bottleneck mechanism. The issue is not whether sequential processing is possible but whether the underlying cause of interference can be attributed to a bottleneck, resource limitation, or some other peripheral interference.

Third, Boles et al. (2007) stated that the likelihood of sequential processing in their experiment is acceptable because it is representative of many real-world tasks. The issue is not whether sequential processing occurs in the real world. Of course, it often does. The issue is the applicability of the multiple resource model when sequential processing is necessitated not by inherent mental limits of the human operator but by physical constraints (e.g., the need to glean information from two spatially disparate locations). Would this distinction have any practical implications such as that in mental workload assessment? It certainly would if mental workload is to be distinguished from physical load and resource limitation is to be distinguished from peripheral interference. These distinctions are made for more than esoteric theoretical reasons. Effective design solutions might hinge on knowing the true causes of the difficulties and inefficiencies inherent in the human-machine interface.

Finally, Boles et al. (2007) argued that the concern about sequential versus simultaneous processing is unimportant because their results have demonstrated that the MRQ works. They emphasized that the imperative is that the MRQ be disseminated and applied such that the extent of its value and generalizability can be assessed. Encouraging further research is certainly a laudable goal, but engineering psychologists also have the responsibility of disseminating an accurate portrayal of the theoretical model upon which the application is based. Otherwise, it is difficult to see how accumulation of knowledge is likely to proceed.

Experimental Tasks Used

The second methodological concern was that the transient nature of the experimental tasks made it difficult to ascertain whether the participants were truly time-sharing. In Experiment 1, four tasks were used. Not only were the tasks presented on independent screens with no control of between-task timing, none of the tasks appeared to exert any continuing demand, such as a continuing memory load. It appeared quite possible that one could simply glean the information from one screen and complete the processing of that task before switching to processing the other task. The dual-task interference here would be caused by the shifting of eye fixations, and there would not be any need to invoke a resource competition account.

In Experiment 2, even though the dual tasks were also unsynchronized discrete tasks, the more complex tasks used suggested that some enduring central-processing demands of one task could be interfering with the processing of the other. The Greebles and Super Maze Wars tasks appeared to be continuous games that would require participants to maintain situation awareness of one task in memory while performing the other task in order to enable quick and efficient recognition of any situation changes upon returning attention to the first task. The Word Tracer task did not appear to require maintaining any previously presented information in memory while attending another task. It appeared possible to complete processing this task in its entirety before switching to another task and start afresh whenever attention was returned. The difference between time-sharing two continuously demanding tasks (Greebles and Super Maze Wars) and time-sharing either of them with a discrete task (Word Tracer) might account for the finding that the Greebles and Super Maze Wars tasks interfered with each other more than either interfered with the Word Tracer task. But this account did not have to invoke any multiple resource explanation. A single resource and the temporal demands between tasks could adequately account for the data without resorting to the multiple resource model outlined by Boles et al. (2007, Table 5, p. 40).

In short, the issue with the experimental tasks used by Boles et al. (2007) boils down to whether they were designed to induce resource competition for multiple resources. The key concern is the simultaneity of resource demands – a necessary condition for inferring resource competition. If the tasks were not time locked to occur together and the tasks were not designed to exert a continuous load while the participant performed the tasks, then there does not appear to be any basis for interpreting the results as being caused by resource competition. The methodological concerns do not have anything to do with whether sequential task processing occurs in the laboratory or in the "real world." The concern has to do with the appropriateness of using tasks and methods that promote sequential processing for evaluating tools that are purportedly designed to assess resource competition.

CONCLUSION

Mental workload assessment remains an important tool for human factors evaluation, but the best procedures for assessing mental workload remain debatable (e.g., Tsang & Vidulich, 2006; Vidulich, 2003). New assessment tools that possess demonstrated validity in detecting subtle cognitive demands, such as the competition for specific cognitive resources, would no doubt be well received by the human factors community. However, to support that the MRQ or any other measurement instrument is indeed reacting to competition for multiple resource structures, researchers would be advised to incorporate methodology more sensitive to the workings of multiple resources (e.g., see Gopher, 1994; Gopher & Donchin, 1986; Navon, 1984). Boles and Law (1998) recognized many of the essential experimental controls, such as the simultaneity of demands and checks for difficulty insensitivity. The absence of such sophistication in Boles et al.'s (2007) study does not inspire confidence in the validity of the multiple resource interpretation of their findings or their MRQ tool.

Many human factors researchers and practitioners would agree with the sentiment expressed by Kantowitz (2000, p. 3.456) that "a good theory is the best practical tool." It is the strong connection between theories and human factors tools that would ultimately support confidence in the applications and their extension to new systems and domains. But, to be confident about the applications, it is essential to rigorously conduct and evaluate laboratory validation of those tools within the appropriate theoretical context.

REFERENCES

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: "Unlocalized" functions. *Neuropsychologia, 30,* 963–988.

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., Bursk, J. H., Phillips, J. B., & Perdelwitz, J. R. (2007, this issue). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human Factors, 49,* 32–45.

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.

Gopher, D. (1994). Analysis and measurement of mental load. In G. d'Ydewalle, P. Eelen, & P. Bertelson (Eds.), *International perspectives on psychological science: Vol. 2. The state of the art* (pp. 265–291). Hove, UK: Erlbaum.

Gopher, D., & Donchin, E. (1986). Workload – An examination of the concept. In K. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. II. Cognitive processes and performance* (pp. 41.1–41.29). New York: Wiley.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kantowitz, B. H. (2000). Attention and mental workload. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 3.456–3.459). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 358–377.

Phillips, J. B., & Boles, D. B. (2004). Multiple resources questionnaire and workload profile: Application of competing models to subjective workload measurement. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 1963–1967). Santa Monica, CA: Human Factors and Ergonomics Society.

Riley, V., Lyall, E., & Wiener, E. (1994). Analytic workload models for flight deck design and evaluation. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 81–84). Santa Monica, CA: Human Factors and Ergonomics Society.

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39,* 358–381.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 417–449). Hoboken, NJ: Wiley.

Vidulich, M. A. (2003). Mental workload and situation awareness: Essential concepts for aviation psychology practice. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 115–146). Mahwah, NJ: Erlbaum.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). San Diego, CA: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3,* 159–177.

Michael A. Vidulich is the division technical advisor for the Warfighter Interface Division of the Air Force Research Laboratory's Human Effectiveness Directorate. He received his Ph.D. in experimental/engineering psychology in 1983 from the University of Illinois at Urbana-Champaign.

Pamela S. Tsang is a professor at the Department of Psychology at Wright State University. She received her Ph.D. in experimental/engineering psychology from the University of Illinois at Urbana-Champaign.

Date received: July 17, 2006
Date accepted: July 27, 2006