Verbal Time Production as a Secondary Task: Which Metrics and Target Intervals are Most Sensitive to Workload for Fine Motor Laparoscopic Training Tasks?

Russell C. Grant[1,2], C. Melody Carswell[1,2], Cindy H. Lio[2], Brent Seales[2], & Duncan Clarke[2]
Department of Psychology[1], and Center for Visualization and Virtual Environments[2]
University of Kentucky

Introduction

Time estimation is sensitive to mental workload in a wide variety of tasks (e.g., Casali & Wierwille, 1984; Liu & Wickens, 1995; Zakay & Shub, 1998; Liu & Wickens, 1995; Zakay & Shub, 1998). However, several methodological questions remain: 1) How long should the target interval be? and 2) How should performance be summarized?

Target Interval Length: How Long?

Hart (1978) suggested using target intervals (TIs) below 30 s to keep load between estimates constant. However, even intervals below 30 s may differ in sensitivity for two reasons. First, there seems to be a perceptual breakpoint in timing near 2-3 s (see Madison, 2001 for review). Second, research suggests a break point in working memory between 12-15 s (Peterson & Peterson, 1959). Thus additional resources may be required to produce intervals over 12-15 s (e.g., prospective memory)

Summarization Metric: Which One?

No consensus has been reached on how to summarize estimated intervals. The following have been frequently used in the literature and were those tested in the analysis:

Mean: (and Mean-to-Target Interval ratio M/TI)
Median (and Median-to-Target Interval ratio: Mdn/TI)
Standard Deviation (SD)
Absolute Deviation from the Median (ADM)
Absolute Deviation from the Mean (ADMdn)
Coefficient of Variation (CV)

Method

Participants: 16 students (4 female, 12 male) with normal vision
Equipment: Stryker 888 endoscope, Stryker Quantum 300 light source, Stryker 888 zero-degree zero-degree 10 mm camera, 2 Maryland dissectors (Figures 3 and 4).

Task Conditions:

Dish-to-Bucket Task Condition: Participants were required to transfer round plastic beads from a dish to an adjacent bucket (Figure 1).
Dish-to-Peg Task Condition: Participants were required to transfer round plastic beads from a dish and place them onto pegs on an adjacent pegboard. This task was thought to be more difficult because it required greater manual precision and the additional demand of having to rotate the bead to fit onto the peg (Figure 2).

Figure 1. Dish-to-Bucket Task

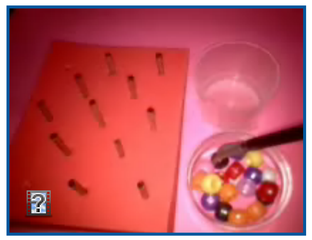


Figure 2. Dish-to-Peg Task

Time Estimation Task:

Participants verbally produced 6, 11, 16 and 21 s intervals.

Procedure:

Participants practiced each of the two conditions xx times before data collection began.

An experimental block was composed of 8 trials – 2 tasks (dish-to-bucket and dish-to-peg) X 4 intervals (6, 11, 16, and 21 sec). Participants completed 2 blocks, with trial order counterbalanced across blocks and participants.

The NASA-TLX (Hart & Staveland, 1988) was administered after each trial.

Figure 3. Equipment

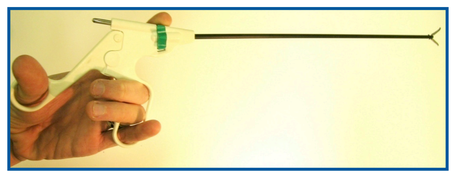


Figure 4. Surgical Graspers

References

Casali, J. G. and Wierwille, W. W. (1984). On the measurement of pilot perceptual workload: a comparison of assessmentt techniques addressing sensitivity and intrusion issues. Ergonomics, 27 (10), 1033-1050.
Hart, S. G. (1978) Time estimation as a secondary task to measure workload: Summary of research, Proceedings of the 14th Annual Conference on Manual Control (NASA CD-2060) (Washington, DC: NASA), 82 - 90.
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A.
Hancock and N. Meshkati (eds), Human Mental Workload (pp. 139-183). Amsterdam: North Holland.
Liu, Y. & Wickens, C. D. (1994). Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. Ergonomics, 37, (11), 1843-1854.
Madison, G. (2001). Variability in isochronous tapping: higher order dependencies as a function of intertap interval. Journal of Experimental Psychology: Perception and Performance, 27 (2), 411-421.
Peterson, L. R. & Peterson, M. J. (1959). Short-term retention of individual verbal items. Journal of Experimental Psychology, 58, 193-198.
Zakay, D. & Shub, J. (1998). Concurrent duration production as a workload measure. Ergonomics, 41, (8),1115-1128.

Results

Statistical Analysis 2 X 2 X 4 within-subjects ANOVAs performed on each of the dependent variables: NASA-TLX, primary task performance, and each of the 6 timing metrics listed previously).

NASA-TLX & Primary Task Performance There were significant main effects for Block and Task Condition for both the NASA-TLX (Hart & Staveland, 1988) and primary task performance highlighting the difference in difficulty of the task conditions and the effect of practice (Block). TLX scores were higher with Dish-to-Peg and lowered with practice. Primary task performance scores were lower with Dish-to-Peg and higher with practice.

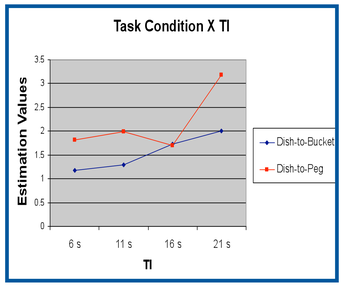
Time Estimation

Interactions:

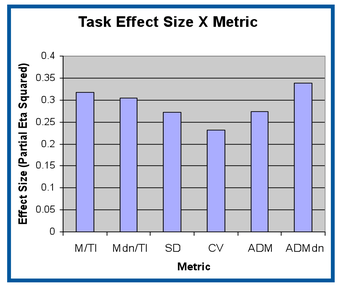
Task Condition X TI: All 8 metrics showed moderate to large effect sizes (.105 < ηp2 < .248) suggesting that sensitivity was dependent upon the target interval used.

Contrasts

Target Interval: For nearly all metrics except Mdn/TI, the 11 s target interval was most sensitive.21 s TI was more sensitive for metrics of dispersion (mean ηp2 = .142) than central tendency (mean ηp2 = .022). The most sensitive TI and metric was 11 s using the ADMdn.



Graph 1. Task X TI



Graph 2. Task X Metric

Conclusion

Task Condition X TI: Indicated that the interval estimated impacted the sensitivity of time estimation.
Short TIs (6, 11 s): The contrasts performed showed that the 11 s interval was most significant, second to the 6 s. This could be because the increase in resources required to produce longer intervals.
Long TIs (16, 21 s): Longer intervals were not as sensitive due to increases in error variance. Literature shows non-linear increases in coefficient of variation (Gibbon, 1997). This may be due either to longer intervals requiring more resources because of the breakpoint in working memory (e.g., prospective memory) or shifting strategies (e.g., participants using task events to produce intervals).
Conclusion:
Care should be taken when generalizing these results given the small sample size. The results may only be applicable to fine-motor tasks. Tasks requiring higher level cognitive resources may see greater task interference with longer intervals.