Learning and Individual Differences 40 (2015) 73–83

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Transforming response time and errors to address tradeoffs in complex measures of processing speed

Linda J. Sorensen [a,*], Dan J. Woltz [b,1]

[a] University of Utah, College of Education, Urban Institute for Teacher Education, 1721 Campus Center Drive, SAEC 2248, Salt Lake City, UT 84112, United States
[b] University of Utah, College of Education, Department of Educational Psychology, 1721 Campus Center Drive, SAEC 3268, Salt Lake City, UT 84112, United States

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form 27 February 2015
Accepted 13 April 2015

Keywords:
Errors
Response time
RR_{Adj}
Speed–accuracy tradeoff

ABSTRACT

An experiment evaluated a transformation of response time (RT) into response rate adjusted for errors (RRAdj) for its ability to accommodate different speed–accuracy tradeoffs (SATs). Participants solved 2-step arithmetic problems under instructional conditions that emphasized speed versus accuracy of responding. RT and error variables were transformed using RRAdj and three additional computations to determine which better equated performance scores under the two tradeoff conditions. Effective adjustment for SAT strategy was evaluated by the equivalence of predictive relationships with other variables, regardless of SAT instructional set. Of the scoring computations compared, RRAdj alone equated performance in the tradeoff conditions, and it was optimized when accuracy was adjusted for guessing. Thus, in certain multistep cognitive tasks, incorporating RT and error data in the RRAdj computation could at least partially adjust for differing SAT strategies and embody additional meaningful variance compared to the commonly used RT for correct responses.

1. Introduction

Given the central role of response time (RT) in cognitive research, one might assume that representing individual differences in processing time would be straightforward. However, Detterman (1987) detailed a number of potential pitfalls which can distort RT data, even in relatively simple tasks such as choice reaction time. These include but are not limited to individual variation in motivation, comprehension of instructions, and tradeoffs related to accuracy versus latency of responses. As researchers employ tasks demanding more complex cognitive processing, the potential pitfalls of RT for accurately representing ability differences are even more serious. Thus, in the present experiment, our concern is with the use of RT measures for investigating individual differences in performance on relatively complex processing-speed tasks (as opposed to issues related to psychometric testing). These tasks demand knowledge-dependent encoding and decision processes and, in many cases, temporary maintenance of the information being processed across multiple processing steps. We evaluated a method of incorporating response accuracy with RT as a general approach to improving the measurement of individual differences in processing speed within complex cognitive tasks, especially when there are differing speed–accuracy tradeoff (SAT) strategies.

1.1. The task complexity spectrum

Defining task complexity is itself a complex task, but the topic has increased in theoretical importance as researchers from divergent fields have investigated both objective task characteristics and cognitive processes required of the performer (Campbell, 1988). We envision task complexity as encompassing a wide spectrum. This ranges from simple RT tasks which involve relatively few cognitive processes and primarily reflect perceptual-motor abilities (Carroll, 1993; Deary, 2000) to tasks such as verbal comprehension and mental arithmetic that are moderately complex due to the amount and variability of information involved and because multiple processing steps may be required (Campbell, 1988; Jensen, 2005).

In moderately complex tasks, the common practice of ignoring errors as a measure of individual differences and analyzing RT only for correct responses may be misleading as to underlying ability for two reasons. First, there can be meaningful variance in performance accuracy, even if error rates are five to 10%. Second, RT alone can distort true individual differences if participants differentially engage in SAT strategies. This latter issue has been noted in the use of simple cognitive tasks (Detterman, 1987; Lohman, 1989), and the unwanted influence of SAT strategy differences could be even greater in more complex measures. Ideally, the measure of individual ability derived from any cognitive task should not depend on the SAT compromise of the participant. In

* Corresponding author. Tel. (801) 587-1743.
E-mail addresses: lsorensen59@gmail.com (L.J. Sorensen), dan.woltz@utah.edu (D.J. Woltz).
1 Tel.: (801) 581-1735.

http://dx.doi.org/10.1016/j.lindif.2015.04.002
1041-6080/

pursuit of this goal, we tested several transformations of RT and accuracy measures from a mental arithmetic task for their ability to provide equivalent indices of performance under different speed–accuracy instructions.

1.2. Approaches to integrating response time and errors

Despite the logical attraction of including error information into analyses of RT in tasks of greater complexity, there is no commonly accepted method for doing so. RT for correct responses and the number of response errors could be included as separate variables in simple or multiple correlation analyses. However, including additional variables has statistical costs, and it is not clear that inclusion of response errors along with RT in a linear prediction equation adequately adjusts for different SAT strategies. Alternatively, there have been a few approaches developed to incorporate error and RT into integrated measures of performance, and these will be reviewed briefly.

A direct approach for measuring performance while accounting for both time and accuracy is estimating an SAT tradeoff function for each individual participant. This generally requires the measurement of performance accuracy when processing time is systematically varied. Wickelgren (1977) derived the most commonly used tradeoff function, $d'(t) = \lambda\,(1 - e^{-\beta\,(t\,-\,\delta)})$, which describes how response accuracy (d') increases non-linearly as a function of experimentally controlled processing time, t (also see Dosher, 1982; Reed, 1976). All three parameters of this function, λ, β and δ, capture aspects of individual performance, but rate parameter β may be the purest measure of performance incorporating both time and accuracy as it provides a rate of accuracy increase as a function of additional processing time.

Despite the attractiveness of estimating individual SAT functions, this approach has limited applicability in many types of research. Obtaining the requisite number of observations at each stimulus exposure requires experimenters to administer more trials than is typical of most experiments (Reed, 1973; Wickelgren, 1977). Moreover, it is likely to be appropriate only for simple processing tasks (Dennis & Evans, 1996; Lohman, 1989). In complex tasks in which some processes are necessarily sequenced, response accuracy at early response deadlines might not adequately capture initial information accrual. Furthermore, the assumption of a negatively accelerated relationship between processing time and response accuracy might be untenable across multiple processing steps if one assumes that this holds within each individual step.

The Ratcliff (1978) diffusion model also accounts for the relationship between accuracy and latency of responses. This method utilizes frequency and RT of both correct and incorrect responses, estimating parameters for each individual from these RT distributions that correspond to psychological processes. Wagenmakers (2009) identified seven parameters associated with the diffusion model, two of which are under subjective control: a, boundary separation, and z, mean starting point. Boundary separation is the pertinent parameter here because it quantifies response caution and, ultimately, modulates the SAT. The diffusion model has been successfully used in a wide range of choice RT experiments (Van Zandt, 2000), but it too has the disadvantage that a relatively large number of trials are required to estimate an RT distribution for infrequent error responses (Wagenmakers). Even with Ratcliff and Tuerlinckx's (2002) minimum number of five error RTs per trial condition, it would take 100 trials in each condition to derive a satisfactory model fit with accuracy levels of 95%. In more complex tasks, where error rates are greater, fewer trials could suffice. Still, number of trials can be an issue when there are multiple within-subject conditions using stimulus domains that are limited in size or when the researcher has multiple RT tasks in the study.

Yet another approach to the SAT problem has been to derive a single measure from a transformation that combines errors and RT. Dennis and Evans (1996) evaluated several such computations using Monte Carlo simulations to reflect variations in accuracy and latency. Analysis of their simulated data using these different scoring methods suggested that the *LogA* index performed better than alternative methods of combining RT and errors. The *LogA* index essentially computed an estimate of the rate parameter β in Wickelgren's (1977) SAT function under simplifying assumptions that circumvented the need to obtain a range of accuracy measures associated with different processing times. This approach, however, could have limited generalizability across levels of task complexity in that the other parameters of the SAT function need to be fixed. Further empirical evidence would be required to accurately determine these values for different tasks outside the realm of simulation. Moreover, fixing these values assumes no individual differences in asymptotic performance and processing time required for accuracy to rise above chance. This could be a reasonable assumption in simple tasks, but it may not be for tasks of greater complexity.

Others have used a simpler adjustment to RT based on the frequency of errors, including Singley and Anderson (1989) in a computerized line-editor learning task. They adjusted total time on each task segment by something roughly equivalent to dividing by the proportion of correct responses, resulting in a metric reflecting the approximate number of seconds subjects spent per correct operation. Similarly, Townsend and Ashby (1983) investigated self-terminating versus exhaustive memory search strategies by calculating $\overline{\mathrm{RT}}/P(\mathrm{corr})$, a ratio of mean RT and proportion of correct responses. Both these computations yield a score that remains interpretable as response latency, but the calculation has the effect of increasing individual RT estimates as accuracy decreases.

Another approach to dealing with RT data in general is to derive a simple rate of responding, or an index of response speed rather than time. Relevant here, an earlier work by Townsend and Ashby (1978) computed $P(\mathrm{corr})/\overline{\mathrm{RT}}$, the reciprocal of their 1983 $\overline{\mathrm{RT}}/P(\mathrm{corr})$ ratio, as an index of effective capacity. With $P(\mathrm{corr})/\overline{\mathrm{RT}}$ proportion correct is divided by mean response time and yields an index of response speed rather than time. Resultant quantities are interpreted as rate of accurate responding, with higher scores indicating greater task proficiency (i.e., more correct responses per unit of time).

None of the methods for combining global task error and time variables described thus far is linked to a process theory as is Ratcliff's (1978) diffusion model, nor do they explicitly deal with SAT function as does Wickelgren's (1977) complete model. However, they may have utility in a wider range of cognitive tasks because they yield psychologically interpretable values and do so with relatively few observations per task condition. Because these performance measures incorporate accuracy in defining RT or response rate, they have the potential to make an adjustment for SAT differences. In addition, they may have further advantages over RT for correct responses only ($\mathrm{RT_c}$) under some circumstances. First, they could increase the systematic variance reflecting the construct of interest that might be distributed across both outcome measures. Second, the response rate transformation, which uses the reciprocal of RT, has the potential advantage of improving the symmetry of $\mathrm{RT_c}$ distributions, which tend to be positively skewed (Fazio, 1990).

The $P(\mathrm{corr})/\overline{\mathrm{RT}}$ index, hereinafter referred to as *response-rate-adjusted*, or $\mathrm{RR_{Adj}}$, was the primary transformation evaluated in this study. It takes the equation form $\mathrm{RR_{Adj}} = (PC/\overline{\mathrm{RT}})60$, with RR representing response rate, PC the proportion correct, and $\overline{\mathrm{RT}}$ the mean response time in seconds for all responses, correct and incorrect. The quotient of $PC/\overline{\mathrm{RT}}$ is multiplied by 60 to produce the number of correct responses per minute, so as to be easily interpreted for a range of tasks. Note that $\mathrm{RR_{Adj}}$ increases as RT decreases, for any given level of accuracy.

The present experiment was designed to empirically evaluate the utility of the $\mathrm{RR_{Adj}}$ transformation for correcting SAT compromises people make in a processing-speed task of moderate cognitive complexity. $\mathrm{RR_{Adj}}$ was used previously to evaluate semantic priming effects that were evident in both RT and accuracy measures in tasks that had moderately complex semantic processing demands (Was, 2010; Was &