Measuring attentional blink magnitude: Reliability and validity of a novel single-target rapid serial visual presentation task index in a psychiatric sample

Chloe C. Hudson[1]

Jenna Traynor[1]

Courtney Beard[1]

Thröstur Björgvinsson[1]

Marie Forgeard[2]

Kean J. Hsu[3]

[1]McLean Hospital

[2]William James College

[3]Georgetown University

Author note: Correspondence should be addressed to Chloe C. Hudson. Address: McLean Hospital, 115 Mill Street, Belmont MA 02478. E-mail address: chudson@mclean.harvard.edu.

Abstract

Rapid serial visual presentation (RSVP) tasks have been frequently used to assess attentional control in psychiatric samples; however, it is unclear whether RSVP tasks exhibits the psychometric properties necessary to assess these individual differences. In the current study, we examined the reliability and validity of single-target computerized RSVP task outcomes in a sample of 63 participants with moderate to severe psychiatric illness. At the group level, we observed the classical attentional blink phenomenon. At the individual level, conventional indices of attentional blink magnitude exhibited poor internal consistency. We empirically evaluated a novel index for assessing attentional blink magnitude using a single-target RSVP task that involves collapsing across experimental trials in which the attentional blink phenomenon occurs and disregarding performance on control trials, which suffer from ceiling effects. We found that this new index resulted in much improved reliability estimates. Both novel and conventional indices provided evidence of convergent validity. Consequently, this novel index may be worth examining and adopting for researchers interested in assessing individual differences in attentional blink magnitude.

Keywords: Rapid serial visual presentation task, attentional blink, reliability, validity, psychopathology, mental illness

Deficits in attentional control have been proposed as transdiagnostic mechanisms that underpin phenotypic expressions of psychiatric disorders (e.g., Goschke, 2014; Michelini et al., 2021). In clinical samples, more severe attentional impairments are associated with poorer functional outcomes, including a more severe course of psychiatric illness and difficulties living independently (e.g., Green, 2006; Keller et al., 2019). As such, understanding the nature and range of attentional impairments in psychiatric samples can help to identify salient therapeutic targets that may improve the course of mental illness and functional outcomes. Yet, researchers' confidence in the role of attention in psychopathology has been hindered by the recent discovery of the poor psychometric properties of many attentional process measures. Indeed, studies examining common attention tasks have revealed poor internal consistency (e.g., Hedge et al., 2018; Rodebaugh et al., 2016). Research relying on psychometrically weak outcomes can undermines our understanding of how individual differences in attention are associated with psychopathology. Thus, further research that examines the psychometric properties of commonly used attentional processing tasks in psychiatric populations is urgently needed (Parsons et al., 2019).

Over the past two decades, the Rapid Serial Visual Presentation (RSVP) task has been used to assess the "attentional blink" phenomenon in both psychiatric samples and samples unselected for mental illness (Martens & Wyble, 2010). In this task, participants are instructed to identify a target stimulus embedded within a stream of distractor stimuli, usually presented at a rate of 10 stimuli per second. To attend to the target, participants must disengage from previously attended distractor stimuli. When all distractor stimuli are inconspicuous (e.g., all of the same color), participants generally perform very accurately on this task; however, when the target is closely proceeded in time by a salient distractor (e.g., typically, when a distractor of a different color precedes the target by 500 milliseconds [ms] or less), participants often fail to detect the target due to attentional interference. This phenomenon is referred to as an "attentional blink" (Martens & Wyble, 2010; Raymond et al., 1992). Because the RSVP task requires participants to selectively orient their attention with minimal demands on memory or sustained attention, attentional blink is thought to reflect attentional control (Bredemeier et al., 2011; Martens et al., 2006; Peers & Lawrence, 2009; Rokke et al., 2002). Individual differences in attentional blink are correlated with several cognitive processes implicated in attentional control, including working memory and inhibition (Willems & Martens, 2016), therefore providing evidence of convergent validity. The attentional blink phenomenon has been reproduced over the past 25 years using many different iterations of the RSVP task that vary in terms of type of stimuli, valence of stimuli, number of response options (e.g., identifying one vs. two targets), and type of response options (e.g., identifying the presence vs. the nature of a target).

Although the attentional blink phenomenon is consistently observed at the group level (Martens & Wyble, 2010), individuals vary in the extent to which they exhibit an attentional blink (referred to as 'attentional blink magnitude'; e.g., Dale & Arnell, 2010; Willems et al., 2013). Thus, researchers have begun to investigate the correlates of attentional blink magnitude (see Willems & Martens, 2016 for a review). Yet, there is a dearth of research examining whether the RSVP task provides reliable measures of these individual differences. Such knowledge is particularly germane because many tasks that detect reliable group-level effects fail to reliably capture individual differences in the same phenomena (Hedge et al., 2018). To date, only four studies have examined the internal consistency of RSVP tasks (Dale & Arnell, 2013; Dale et al., 2013; Martens & Johnson, 2009; Martens & Valchev, 2009).[1] Each of these studies

[1] A summary of the design characteristics and internal consistency of attentional blink magnitude for each of these tasks is presented in Supplementary Table S1.

recruited samples unselected for mental illness and have focused on iterations of the dual-target RSVP task, in which participants are instructed to identify two targets. The average estimate of internal consistency across these studies was .68, but estimates varied from .48 (Dale & Arnell, 2013) to .90 (Martens & Johnson, 2009). The large variability in reliability estimates from dual-target RSVP tasks suggests that variations in task design, sample, or how authors calculate attentional blink magnitude may influence whether RSVP task outcomes reliably capture individual differences in attentional blink magnitude.

In the current study, we investigated the reliability and validity of outcomes from a single-target RSVP task in a psychiatric sample of adults with a wide range of mental health diagnoses (e.g., mood, anxiety, personality, and psychotic-spectrum disorders). This study advances the existing literature in two important ways. First, prior reliability estimates of the RSVP task focused exclusively on dual-target RSVP tasks. Single-target RSVP tasks differ from dual-target tasks because participants are not explicitly asked to monitor the distractor stimuli (Spalek et al., 2006). Although dual-target tasks have the advantage of ensuring that the initial target stimuli capture participants' attention, single-target RSVP tasks allow for inferences about spontaneous attentional engagement to salient distractors, and thus may represent attentional control processes in a more ecologically valid way. Second, prior work has focused exclusively on the reliability of the RSVP task using samples unselected for mental illness. Given that attentional processes are of particular interest in psychiatric populations due to their potential contributions to psychopathology (e.g., Goschke, 2014; Michelini et al., 2021), and that the RSVP task has been used with numerous psychiatric populations that are characterized by impairments in attentional control (e.g., MacLean et al., 2010; Mathis et al., 2011), it is crucial to evaluate the psychometric properties of the RSVP task outcomes in psychiatric samples.

Method

Participants

Participants were recruited from <*name blinded for review*> in <*blinded for review*>. In brief, patients who attend this treatment program experience a range of moderate to severe mental health concerns (e.g., mood, anxiety, personality, and psychotic-spectrum disorders; see <*citation blinded for review*> for more details on this treatment program). All patients attending the treatment program were eligible for the current study, and those interested in participating provided informed consent to take part in an IRB-approved research study on cognitive flexibility between February 2016 and June 2017. Data included in the current analysis were derived from the full study sample of 89 participants. Of these participants, 25 did not complete the RSVP task due to scheduling limitations, technical difficulties, or clinical concerns requiring premature withdrawal from the research study. Further, one participant performed more than 3.5 standard deviations from the mean on control trials (i.e., trials in which no distractor was present) and was therefore excluded from analyses. As such, the final sample included 63 participants. These participants did not differ from the 26 participants who were excluded in terms of age, gender, ethnicity, or symptom severity, all ps > .10.

Participants ranged in age from 18 to 68 years (M = 29.24, SD = 11.07). Roughly equal proportions of the sample identified as men (54%) and women (46%). Most participants were Non-Hispanic White (69.8%), followed by Asian (12.7%), Hispanic/Latinx (6.3%), Multiracial (7.9%), or Black/African American (3.2%). In terms of education, the largest proportion of our sample had some college education (38.1%), followed by a 4-year college degree (25.4%), post-college education (23.8%), high school diploma/GED (11.1%), and either grade eight or less

(1.6%). Approximately half of participants in the sample (50.8%) were hospitalized for psychiatric problems in the 6 months preceding study participation.

Measures

Rapid Serial Visual Presentation Task (RSVP). Participants completed a single-target RSVP task designed to measure deficits in attentional disengagement from neutral stimuli (Bredemeier et al., 2011; Spalek et al., 2006). In this task, participants were instructed to look at a computer screen where a stream of stimuli were briefly presented (66 ms) in the center of the screen. The stimuli were digits (0-9) and letters (all uppercase letters of the English alphabet, except I, O, Q and Z). A single target (red letter) was embedded in the stream of black digits in all conditions and participants were instructed to search for the red letter. At the end of each trial, participants identified the target letter by key press. In the experimental trials, a salient distractor (a green letter) was embedded in the stream of stimuli 100 ms, 200 ms, 300 ms, or 700 ms (lags) before the target. In the control trials, no distractor was presented.

At the beginning of the session, participants read instructions presented on the screen and were invited to ask questions about the procedure. Each trial began with the presentation of a fixation cross in the center of the screen that indicated the location where the stream of stimuli was later presented. Participants initiated each trial by pressing the space bar. A stream of 20 digits and letters was then presented, each for 66 ms with an inter-stimuli interval of 26 ms, yielding a presentation rate of approximately 10 stimuli per second. After 8 practice trials, participants completed 120 trials (60 control trials and 15 trials for each lag the distractor is presented in the experimental trials). As has been done in prior studies, control trials were assigned a label of 100 ms, 200 ms, 300 ms, or 700 ms lag for analysis purposes, despite there being no distractor to lag the target in these trials. Performance on the RSVP task was operationalized as percent accuracy. This task took approximately 6 to 8 minutes to complete.

Stroop Task. Participants completed the classic Stroop task as a measure of inhibition, which has been characterized as a form of attentional control (e.g., Eysenck & Derakshan, 2011). This task had three conditions. In the first condition ("color naming"), participants viewed a series of color patches on a sheet of card and were asked to say the color of the patch out loud. In the second condition ("word reading"), participants viewed a series of color words (e.g., "BLUE") displayed in black ink on a sheet of card and were asked to read the word out loud. These first two conditions provide a baseline measure of lower-level functions that are fundamental to the task. In the third and final condition ("inhibition"), participants were presented with color words displayed in an incongruent colored ink (e.g., the word "BLUE" printed in red colored ink) on a sheet of card. Participants were instructed to say the color of the ink the color word is printed in, which requires them to inhibit the prepotent response of reading the word. For each condition, time required for completion (ms) and number of uncorrected and self-corrected errors were recorded.

Procedure

Participants were approached in the milieu during downtime in their partial hospital treatment about opportunities to participate in research. Those that expressed interest were invited to take part in two 25-to-30-minute testing sessions on the second and/or third day of their treatment program. All participants provided written informed consent. The RSVP task was completed during the second testing session on a laptop computer using OpenSesame (Mathôt et al., 2012). At the first testing session, participants completed a Stroop task, self-report questionnaires assessing mental health symptoms and cognitive processes, and other tasks. At the second testing session, participants completed the RSVP task followed by additional self-report questionnaires. This study was approved by the <*blinded for review*> Institutional Review Board. This study was not preregistered. De-identified data and syntax are available at our OSF website for this project: <*blinded for review*>.

Data Analysis

To assess the group-level attentional blink phenomenon, we conducted a repeated-measures ANOVA with condition (control, experimental) and lag (100 ms, 200 ms, 300 ms, 700 ms) as within-subject factors in SPSS statistical software version 28.0. A significant interaction was followed up by examining pairwise comparisons using Bonferroni correction. To assess individual differences in attentional blink, we calculated difference scores by subtracting percent accuracy in control trials from percent accuracy in experimental trials for each lag, as has been done in prior studies (e.g., Bredemeier et al., 2011; Kellie & Shapiro, 2004; Spalek et al., 2006). The internal-consistency reliability of the RSVP task outcomes was computed using: (1) Cronbach's alpha[2] of all trials within each trial type using Hayes and Coutts (2020) *omega* macro for SPSS and (2) permutation-based estimates of split-half reliability using the *splithalf* package for R (Parsons, 2021).[3] We used 5,000 random splits to calculate Spearman-Brown corrected reliability estimates and 95% confidence intervals. Finally, convergent validity was assessed by

---

[2] We report the results of Cronbach's alpha using point-biserial correlations, which is the default input correlation for Cronbach's alpha when item responses are dichotomous. For those interested, we have also reported a tetrachoric alpha, in which tetrachoric correlations were used as input values, in Supplementary Table S2. However, we focus on Cronbach's alpha results in the manuscript for two reasons, which are discussed in more detail by Chalmers (2018). First, Cronbach's alpha does not require continuous item response data, and as such, alternatively forms of alpha are not necessary when data are dichotomous. Second, and perhaps more importantly, alpha coefficients with alternative correlations as input are not equivalent to Cronbach's alpha and may result in unacceptably liberal sampling error estimates. Therefore, tetrachoric alpha coefficients cannot be interpreted as though it is a standard estimate of reliability.

[3] We attempted to compute McDonald's omega for each index; however, it could not be computed because several variables had a variance of zero and many of the remaining variables were very weakly or negatively correlated.

computing the Pearson R correlation between indices of attentional blink magnitude and completion time on the inhibition condition of the Stroop task.[4]

Results

Group-Level Attentional Blink Phenomenon

We found a main effect of condition, $F(1, 62) = 71.51$, $p < .001$, $\eta_p^2 = .54$ and lag, $F(3, 60) = 16.10$, $p < .001$, $\eta_p^2 = .45$. These effects were qualified by a significant condition by lag interaction, $F(3, 60) = 17.31$, $p < .001$, $\eta_p^2 = .46$. Follow-up analyses found that there was no effect of lag on control trials, $F(3, 60) = 0.39$, $p = .76$, but there was a significant effect of lag on experimental trials, $F(3, 60) = 17.84$, $p < .001$, $\eta_p^2 = .47$. Consistent with the expected attentional blink phenomenon, participant performance was more accurate on experimental trials with a 700 ms lag ($M = 98.52$, $SD = 3.48$) relative to experimental trials with a 100 ms lag ($M = 90.16$, $SD = 11.29$; $t[62] = 5.54$, $p < .001$, $d = 1.00$), 200 ms lag ($M = 89.63$, $SD = 11.60$; $t[62] = 6.32$, $p < .001$, $d = 1.04$), or 300 ms lag ($M = 95.25$, $SD = 7.89$; $t[62] = 3.32$, $p = .002$, $d = 0.54$). Further, participant performance was more accurate on experimental trials with a 300 ms lag relative to experimental trials with a 100 ms lag ($t[62] = 3.10$, $p = .003$, $d = 0.52$) or 200 ms lag ($t[62] = 4.00$, $p < .001$, $d = 0.57$). There was no significant difference in accuracy across experimental trials with a 100 ms vs. 200 ms lag ($t[62] = 0.31$, $p = .76$). See Figure 1 for a visual depiction of this interaction.

Reliability of Attentional Blink Magnitude Indices

Table 1 provides the reliability coefficients of raw accuracy scores across trial conditions as well as the conventional attentional blink magnitude index. Across all variables, reliability did not meet the traditional threshold for acceptability (i.e., .70; Fields et al., 2012; Nunnally &

[4] We also examined this association controlling for color naming and word reading completion time. The pattern of results was the same as those presented below. Results of these analyses can be found on our OSF website.

Bernstein, 1994; Taber, 2017). Reliability coefficients for many of the control conditions were negative, which can occur when the sum of the individual item variances is greater than the scale variance (e.g., Krus & Helmstadter, 1993). We speculate that the poor reliability observed was driven by limited variability in performance accuracy on control trials; on average, participants made fewer than one error across all 60 control trials (M = 0.46, SD = 0.69), and 22 (36.67%) of the control trials had a variance of 0 because all participants answered the question correctly. These results indicate that almost all participants are performing at or close to ceiling on control trials, making discrimination among participants impossible.

Reliability statistics were substantially improved for experimental trials (.21 to .58), but similarly did not meet the threshold for acceptability. Again, we speculated that poor reliability may be due to limited variability in accuracy. On average, participants made 4 errors across all 60 experimental trials (SD = 3.47), with most of these errors occurring in the 100, 200, and 300 ms lag trials. Not surprisingly, the conventional metric used to assess individual differences in attentional blink magnitude (i.e., difference scores) did not surpass the reliability of these raw scores.

To test our proposition that poor reliability was driven primarily by lack of variability across experimental trials, we assessed the reliability of a new attentional blink index that averaged accuracy across all conditions in which the attentional blink phenomenon occurred (experimental trials with a 100 ms, 200 ms, or 300 ms lag). Using this new variable, Cohen's alpha was .71 (95% CIs .59 to .80) and split-half reliability was .71 (95% CIs .57 to .81). As such, when collapsing across these attentional blink trials, reliability statistics improved substantially and met traditional thresholds for acceptable reliability.

Validity of Attentional Blink Magnitude Indices

The correlation between attentional blink magnitude indices and Stroop task inhibition completion time can be found in Table 2. Using conventional indices, larger attentional blink magnitude at 100 ms and 200 ms lag, was significantly associated with slower task completion during the Stroop inhibition condition. Similarly, less accurate performance on our novel index of attentional blink magnitude was associated with slower completion time during the Stroop inhibition condition. Completion time in the Stroop inhibition condition was not associated with conventional indices of attentional blink magnitude at 300 ms or 700 ms lag. In sum, these findings demonstrate convergent validity for conventional indices of attentional blink calculated at 100 ms and 200 ms lag, as well as our novel attentional blink index.

Discussion

Consistent with a large body of existing work, we detected the attentional blink phenomenon at the group level in our sample: participants were significantly more accurate on trials in which the target stimulus was immediately proceeded by a salient distractor relative to trials where all distractors were inconspicuous. The existence of this group-level effect across numerous samples, including those that vary substantially on characteristics that are likely to impact attentional control (e.g., samples unselected for mental illness vs. psychiatric samples), suggests that this effect is highly replicable at the group level.

In direct contrast, our results also demonstrate that conventional indices for assessing individual differences in attentional blink magnitude exhibit poor reliability when data is obtained from a single-target RSVP task with a psychiatric sample. Across both control and experimental trials in each lag condition, reliability statistics did not reach the traditional threshold for acceptability. Control trials are typically included in indices of attentional blink magnitude to account for variation in performance that is unrelated to attentional blink; however, our results suggest that control trials in our task design are not reliably capturing individual differences in any processes (related to attentional blink or not). Indeed, the reliability of control trials was so poor that it suggests that most variability was due to measurement error, or put simply, chance (Lord et al., 1968). Although experimental trials demonstrated better reliability than control trials, they also failed to reach traditional thresholds for acceptability. We speculated that reliability may be undermined due to limited variability in performance.

Given these limitations, we proposed and evaluated one method for enhancing the reliability of attentional blink magnitude when assessed using a single-target RSVP task: modifying indices. First, we removed control trials from the calculation of attentional blink magnitude because of strong ceiling effects. Control trials help to ensure that task indices are capturing attentional abilities and not other non-target processes (e.g., fatigue, processing speed, motivation, etc.). Indeed, one participant was excluded from the current study because their performance on control trials indicated substantial deviations from average, which likely represents the presence of non-target process that influenced performance. However, our results suggest that more subtle variation in control trial performance does not reliably capture individual differences. As such, we suggest that control trials should not be included in attentional blink magnitude calculations when ceiling effects are detected. Second, because poor reliability can be driven by a lack of variability, we collapsed across experimental trials in which the attentional blink phenomenon occurs and capitalized on this maximized variability. In the current sample and task design, this novel method of calculating attentional blink magnitude resulted in superior reliability relative to conventional metrics of attentional blink magnitude.

Our novel metric for attentional blink magnitude demonstrated convergent validity with Stroop inhibition completion time. Importantly, our index demonstrates equivalent convergent validity estimates to those obtained with the conventional index at 100 ms and 200 ms lags (i.e., medium effect sizes; $z > 1.07$, $p < .14$), and better convergent validity estimates to those obtained with the conventional index at 300 ms or 700 ms lag (i.e., small effect sizes; $z > 1.75$, $p < .04$). These finding suggests that eliminating control trials in our novel index of attentional blink magnitude does not reduce the validity of estimates obtained, and instead, may enhance validity compared to some traditional indices.

Of course, modifying attentional blink indices is only one of many potential ways that researchers may be able to improve the reliability of single-target RSVP task indices. A review of the similarities and differences between RSVP task designs and samples may also help to explain the heterogeneity in reliability estimates obtained in the current and prior studies, and consequently, point towards other means of improving reliability (see Supplementary Table S1). First and foremost, each of the prior studies assessed the reliability of dual-target RSVP tasks, which require participants to attend to two targets, while our single-target RSVP task only required participants to attend to one target. Although dual-target tasks have the advantage of ensuring that the initial target stimuli capture participants' attention, previous research suggests that stimuli do not need to be task-relevant to capture attention and induce an attentional blink effect (e.g., Folk et al., 2002; Maki & Mebane, 2006; Most et al., 2005; Spalek et al., 2006). However, dual-target search is often slower and less accurate than single-target searches (Barrett & Zobay, 2014), suggesting that dual-target RSVP tasks will minimize ceiling effects. Indeed, the accuracy of control trials in studies that used dual-target tasks was lower than the accuracy of control trials in our study (i.e., .77 to .92 vs. .99, respectively). As such, single-target RSVP tasks may need modifications, such as the proposed changes to estimates of attentional blink magnitude, to maximize variability across trials and limit ceiling effects. Nevertheless, single-target tasks may still be preferable to dual-target tasks when research questions warrant assessment of *spontaneous* attentional engagement, because dual-target tasks do not allow for such inferences.

Second, the characteristics defining the target stimuli, distractors, and their similarity varied between studies. As can be seen in Table S1, several different iterations of these characteristics have been evaluated in dual-target tasks. In the current study, our target and salient distractor differed from inconspicuous distractors in terms of two salient features: color (red for target stimuli, green for salient distractor, black for inconspicuous distractors) and category (letters for target stimuli and salient distractor, digits for inconspicuous distractors). These design decisions were made because previous research suggests that featural interference is a major determinant of attentional blink, and greater featural difference between the salient distractor and inconspicuous distractor enhances attentional blink magnitude (e.g., McAuliffe & Knowlton, 2000). Dale and colleagues (2013) "featural RSVP task" (row 4 in Supplemental Table S1) used similar task characteristics to those used in the current study. In this task, targets were red letters, and the remaining inconspicuous distractor stimuli were black digits. Of note, the dual-target version of this RSVP task achieved very similar reliability estimates to the conventional indices of reliability achieved in the current study, even though ceiling effects were not observed in this dual-target version of the task. These findings suggest that the characteristics of the task may influence reliability above and beyond the impact of ceiling effects; however, future research is needed to empirically investigate this possibility.

Finally, the current study is the first to our knowledge that investigates the reliability of RSVP task using a psychiatric sample. Indeed, an important strength of the current study is the use of a clinically representative sample recruited from patients seeking partial hospital-level psychiatric treatment. Patients with mental health difficulties experience a myriad of cognitive impairments associated with functional impairment (e.g., Gupta et al., 2013; Bora et al., 2009). Furthermore, cognition is most impaired among those with the highest level of psychopathology (e.g., Roca et al., 2015). Given that studying these altered attentional processes in psychiatric samples is vital to understanding the ontology and maintenance of symptoms, it is critical that we evaluate the psychometric properties of assessment tools in this population.

It is also possible; however, that psychiatric samples have more volatile attentional control abilities relative to samples unselected for mental illness. Consequently, research assessing reliability of RSVP task outcomes in psychiatric samples may yield lower estimates of reliability compared to studies using the same task with non-psychiatric samples. Future research is needed that directly compares the reliability of attentional control estimates between psychiatric and non-psychiatric samples to empirically test whether our novel index can also increase the reliability of task outcomes in non-psychiatric samples. Nevertheless, we provide evidence that reliable and valid attentional control metrics can be achieved when using a single-target RSVP task with a psychiatric sample, if modifications are made to conventional attentional blink indices. We propose and provide the first empirical evidence for novel index that reliably assesses attentional blink magnitude in treatment-seeking patients with moderate to severe mental illness using the current version of a single-target RSVP task.

The results of the current study should also be interpreted in the context of two notable limitations. First, like much of psychological science, our sample is relatively limited in terms of sociodemographic diversity (Henrich et al., 2010), and as such, it is unclear if the results will generalize to more diverse populations. Further, it is unclear whether the reliability of attentional blink indices vary by demographic characteristics (especially age, given the well-established cognitive decline that occurs later in life). As such, future research is needed to examine whether task reliability is moderated by sociodemographic characteristics. Second, we studied one of many versions of the RSVP task: a single-target task with 120 neutral trials. There are countless iterations of even single-target RSVP tasks that vary in terms of target type, valence of stimuli, and number of trials (see Kelly & Dux, 2011 for a review). It is plausible that altering any of these task parameters may influence reliability, and that it is in fact the design of the current study's task that contributed to low reliability when all trials were included. In other words, modifying the task design may also be an effective strategy for enhancing the reliability of single-target RSVP task outcomes. Of particular relevance to the current study is the number of trials: if lack of variability in responses hindered reliability, it is possible that adding additional trials may help to improve reliability by introducing more variability in performance. Relatedly, varying the speed of stimuli presentation or examining reaction times as an outcome variable may also reduce ceiling effects thereby increasing variability in performance. Unfortunately, given the dearth of reporting of the psychometric properties of task-based measures (Parsons et al., 2019), it is unclear how our findings compare to other iterations of the single-target RSVP task.

It is worth noting that augmenting task design features, especially adding more items, is likely to increase participant burden. Patients with moderate to severe mental illness often experience functional impairments that make participation in research challenging (e.g., Bixo et al., 2021), and as such, design changes that enhance participant burden may systematically prevent individuals from participating, limiting the generalizability of results. As such, our proposed novel index may be one way to increase reliability while effectively balancing concerns inherent in psychiatric research.

Given the replication crisis in psychology broadly, and cognitive science more specifically, it is crucial that we employ experimental tasks that produce replicable effects at both the individual and group level to study individual and group differences in processes of interest. In the current study, we present the first psychometric analysis of individual differences in attentional blink magnitude using a single-target RSVP task and in psychiatric sample. We highlight fundamental problems with conventional means of assessing attentional blink magnitude in a single-target RSVP design and propose a novel method that substantially improves reliability. Specifically, we show that collapsing across experimental trials in which the attentional blink phenomenon occurs, and disregarding performance on control trials which suffer from ceiling effects, greatly improves the reliability of attentional blink magnitude obtained from single-target RSVP tasks while preserving convergent validity. In our study, this effect was observed in a transdiagnostic psychiatric sample and warrants replication in other samples. Nonetheless, by developing this novel index of attentional blink magnitude, our hope is that we have improved the methodological rigor of the single-target RSVP task, which will ultimately help to ensure that researchers are able to generate meaningful and replicable research findings on attentional processes in psychopathology.

Declaration of Interest Statement

The authors report there are no competing interests to declare.

References

Barrett, D. J., & Zobay, O. (2014). Attentional control via parallel target-templates in dual-target search. *PloS one, 9*(1), e86848. https://doi.org/10.1371/journal.pone.0086848

Bixo, L., Cunningham, J. L., Ekselius, L., Öster, C., & Ramklint, M. (2021). 'Sick and tired': Patients reported reasons for not participating in clinical psychiatric research. *Health Expectations, 24*, 20-29. https://doi.org/10.1111/hex.12977

Bora, E., Yücel, M., & Pantelis, C. (2010). Cognitive impairment in affective psychoses: a meta-analysis. *Schizophrenia Bulletin, 36*(1), 112-125. https://doi.org/10.1093/schbul/sbp093

Bredemeier, K., Berenbaum, H., Most, S. B., & Simons, D. J. (2011). Links between neuroticism, emotional distress, and disengaging attention: Evidence from a single-target RSVP task. *Cognition & Emotion, 25*(8), 1510-1519. https://doi.org/10.1080/02699931.2010.549460

Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. *Educational and Psychological Measurement, 78*(6), 1056-1071. https://doi.org/10.1177/0013164417727036

Dale, G., & Arnell, K. M. (2010). Individual differences in dispositional focus of attention predict attentional blink magnitude. *Attention, Perception, & Psychophysics, 72*(3), 602-606. https://doi.org/10.3758/APP.72.3.602

Dale, G., & Arnell, K. M. (2013). How reliable is the attentional blink? Examining the relationships within and between attentional blink tasks over time. *Psychological Research, 77*(2), 99-105.https://doi.org/10.1007/s00426-011-0403-y

Dale, G., Dux, P. E., & Arnell, K. M. (2013). Individual differences within and across attentional blink tasks revisited. *Attention, Perception, & Psychophysics, 75*(3), 456-467. https://doi.org/10.3758/s13414-012-0415-8

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences, 50*(7), 955-960.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.

Folk, C., Leber, A., & Egeth, H. (2002). Made you blink! Contingent attention capture produces a spatial blink. Perception and Psychophysics, 64, 741753.

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research, 23*(S1), 41-57. https://doi.org/10.1002/mpr.1410

Green, M. F. (2006). Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. Journal of Clinical Psychiatry, 67, 3.

Gupta, M., Holshausen, K., Best, M. W., Jokic, R., Milev, R., Bernard, T., ... & Bowie, C. R. (2013). Relationships among neurocognition, symptoms, and functioning in treatment-resistant depression. *Archives of Clinical Neuropsychology, 28*(3), 272-281. https://doi.org/10.1093/arclin/act002

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1-24. https://doi.org/10.1080/19312458.2020.1718629

Hedge, C., Powell, G. & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29-29. https://doi.org/10.1038/466029a

Keller, A. S., Leikauf, J. E., Holt-Gosselin, B., Staveland, B. R., & Williams, L. M. (2019). Paying attention to attention in depression. Translational psychiatry, 9(1), 1-12.

Kellie, F. J., & Shapiro, K. L. (2004). Object file continuity predicts attentional blink magnitude. Perception & Psychophysics, 66(4), 692-712.

Kelly, A. J., & Dux, P. E. (2011). Different attentional blink tasks reflect distinct information processing limitations: an individual differences approach. Journal of experimental psychology: *Human perception and performance, 37*(6), 1867. https://doi.org/10.1037/a0025975

Krus, D. J., & Helmstadter, G. C. (1993). The probabilities of negative reliabilities. *Educational and Psychological Measurement, 53*, 643-650. https://doi.org/10.1177/0013164493053003005

Lord, F.M., Novick, M.R., & Birnbaum, A. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

MacLean, M., Arnell, K., & Busseri, M. (2010). Dispositional affect predicts temporal attention costs in the attentional blink paradigm. *Cognition and Emotion, 24*(8), 1431-1438.

Maki, W. S., & Mebane, M. W. (2006). Attention capture triggers an attentional blink. Psychonomic Bulletin and Review, 13, 125131.

Martens, S., & Johnson, A. (2009). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Experimental Brain Research, 192*(1), 43-52. https://doi.org/10.1007/s00221-008-1551-1

Martens, S., & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews, 34*(6), 947-957. https://doi.org/10.1016/j.neubiorev.2009.12.005

Mathis, K. I., Wynn, J. K., Breitmeyer, B., Nuechterlein, K. H., & Green, M. F. (2011). The attentional blink in schizophrenia: isolating the perception/attention interface. *Journal of Psychiatric Research, 45*(10), 1346-1351. https://doi.org/10.1016/j.jpsychires.2011.04.002

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

Michelini, G., Palumbo, I. M., DeYoung, C. G., Latzman, R. D., & Kotov, R. (2021). Linking RDoC and HiTOP: A new interface for advancing psychiatric nosology and neuroscience. Clinical psychology review, 86, 102025.

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. Psychonomic bulletin & review, 12(4), 654-661.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Olivers, C. N. L., & Nieuwenhuis, S. (2005). The beneficial effect of concurrent task-irrelevant mental activity on temporal attention. *Psychological Science, 16*, 265-269. doi:10.1111/j.0956-7976.2005.01526.x

Olivers, C. N. L., & Nieuwenhuis, S. (2006). The beneficial effects of additional task load, positive affect, and instruction on the attentional blink. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 364-379. doi:10.1037/0096-1523.32.2.364

Parsons, S. (2021). Splithalf: Robust estimates of split half reliability. *Journal of Open Source Software, 6*(60), 3041. https://doi.org/10.21105/joss.03041.pdf

Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science, 2*(4), 378-395. https://doi.org/10.1177/2515245919879695

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink?.*Journal of experimental psychology: Human perception and performance, 18*(3), 849. https://doi.org/10.1037/0096-1523.18.3.849

Roca, M., Vives, M., López-Navarro, E., García-Campayo, J., & Margalida, G. (2015). Cognitive impairments and depression: A critical review. Actas Esp Psiquiatr, 43(5), 187-193.

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., ... & Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology, 125*(6), 840. https://doi.org/10.1037/abn0000184

Spalek, T. M., Falcon, L. J., & Di Lollo, V. (2006). Attentional blink and attentional capture: Endogenous versus exogenous control over paying attention to two important events in close succession. *Perception & Psychophysics, 68*(4), 674-684. https://doi.org/10.3758/BF03208767

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296.https://doi.org/10.1007/s11165-016-9602-2

Willems, C., & Martens, S. (2016). Time to see the bigger picture: Individual differences in the attentional blink. *Psychonomic Bulletin & Review,* 23(5), 1289-1299.https://doi.org/10.3758/s13423-015-0977-2

Willems, C., Wierda, S. M., van Viegen, E., & Martens, S. (2013). Individual differences in the attentional blink: the temporal profile of blinkers and non-blinkers. *PLoS One, 8*(6), e66185.https://doi.org/10.1371/journal.pone.0066185

Figure 1

Percent Accuracy Stratified by Lag and Condition on the Rapid Serial Visual Presentation Task

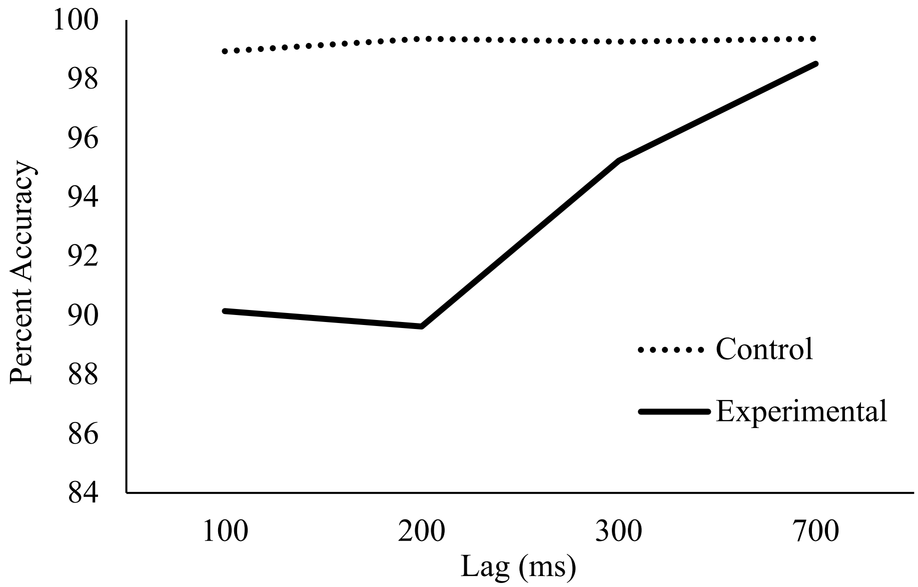

Table 1

Reliability Estimates and 95% Confidence Intervals of Rapid Serial Visual Presentation Task Indices

Raw Scores	Lag (ms)				
	100	200	300	700	Overall
Split-Half Reliability					
Control Trials[5]	.12 (-.14 to .32)	-.09 (-.10 to -.07)	-.10 (-.11 to -.08)	-.09 (-.10 to -.07)	.06 (-.24 to .37)
Experimental Trials	.58 (.38 to .74)	.58 (.38 to .74)	.57 (.38 to .75)	.22 (-.16 to .55)	---
Cronbach's Alpha					
Control Trials	.07 (-.20 to .33)	-.08 (-.15 to .00)	-.10 (-.19 to -.02)	-.09 (-.17 to .00)	.05 (-.34 to .30)
Experimental Trials	.57 (.22 to .72)	.58 (.39 to .69)	.57 (.22 to .70)	.21 (-.13 to .40)	---
Attentional Blink Magnitude					
Split-Half Reliability	.54 (.34 to .71)	.54 (.32 to .71)	.39 (-.02 to .64)	.01 (-.36 to .39)	

[5] Split-half reliability estimates conducted separately for control trials assigned 100 ms, 200 ms, 300 ms, and 700 ms lags were calculated with only 50 permutations of random splits. Additional permutations led to error messages because multiple variables had a variance of zero. Split-half reliability conducted manually with odd and even trials confirmed that reliability estimates are consistently extremely low (< .12) for control trials at each lag.

Table 2.

Correlations Between Attentional Blink Magnitude Indices and Stroop Inhibition Completion Time

	Correlation
Conventional Indices	
100 ms	.31*
200 ms	.46**
300 ms	.18
700 ms	.19
Novel Index	-.48**

Note. *$p < .05$, **$p <. 001$. ms = millisecond.