Overview of the Gradual Onset Continuous Performance Task (City or Mountain)

Authors: Roger Strong and Amy Price

Table of Contents:

- Background & Scientific Purpose
- Methodology
- Data and Analysis Guidelines
    - Data
    - Suggested Outcomes
    - Quality Control Guidelines
    - Calculating Test Reliability
    - Correlates of Interest
- References

Background & Scientific Purpose

Maintaining focused attention for extended periods of time is difficult, yet the ability to sustain attention is related to important everyday outcomes such as driving ability, educational performance, and occupational performance. Additionally, difficulty sustaining attention is a common symptom of several neurological and psychiatric disorders, highlighting the importance of this cognitive ability in health.

The TestMyBrain Gradual Onset Continuous Performance Test (Fortenbaugh et al., 2015; Hawks et al., 2023; Riley et al., 2016, 2017; Singh et al., 2021; Treviño et al., 2021; Vogel et al., 2020) is a test created for the *TestMyBrain* platform, adapted from a test originally developed by the Boston Attention and Learning Lab (Esterman et al. 2013). This is a test of sustained attention and inhibitory control, core components of executive functioning.

The test has previously been used by researchers in multiple studies, including: to assess age-related changes in sustained attention across the lifespan (Fortenbaugh et al., 2015), time-of-day related changes in sustained attention (Hawks et al., 2023; Riley et al., 2017), the association of gender inequality with sustained attention performance (Riley et al., 2015), and the association of childhood adversity with sustained attention in adulthood (Vogel et al., 2020).

Methodology

In the TestMyBrain Gradual Onset Continuous Performance Test, participants view a sequence of 300 circular, grayscale images of cities and mountains constructed to have equal mean luminance and contrast. Images gradually transition from one to the next every 800 ms. Participants are instructed to make a keyboard or screen press when images of cities are presented (89.3% of trials), and withhold from responding when mountain images are presented (10.7% of trials). The total duration of the image sequence is 4 minutes. Before completing the 300 scored test trials, participants complete 36 unscored practice trials, which are divided into three separate practice rounds. All participants view the same image sequence. See Figure 1 for an overview of the task structure.

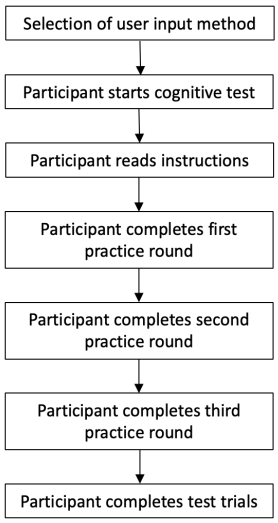

Figure 1: Flow diagram for Gradual Onset Continuous Performance Test

Participants make their response either using a keyboard press or touch input. At the start of the test, participants using devices with touch input must select whether they will use touch input or a keyboard to complete the task. Participants without touch-compatible devices are automatically assigned to keyboard input.

After the response selection method is determined, participants view brief instructions for the test, then complete the first of three rounds of practice trials. This first round of practice trials includes six total images (50% city images, 50% mountain images). Unlike the actual test where images gradually transition from one to the next, this first practice round presents static images with full visibility (100% opacity). At the start of each trial, the screen is blank for 1000 ms before the image appears. The trial's image is then displayed for 2000 ms, or until a response is made. When a participant correctly makes a response (city trials) or correctly withholds a response (mountain trials), a message reading "Correct" is displayed on the screen for 1000 ms. When a participant incorrectly withholds a response (city trials) or incorrectly makes a response

(mountain trials), a feedback screen appears informing the participant that they did not respond properly. This feedback screen remains on the screen until the participant presses the spacebar (keyboard input) or clicks a button (touch input) to continue. Practice trials for which the correct response was not made are repeated until the correct response is made. Note that no data is logged for the first round of practice trials.

Participants next view brief instructions before completing a second round of practice trials that includes ten trials (80% city images, 20% mountain images). Like the test trials that will be completed later on, the images gradually transition from one to the next. The image sequence begins with a noise image at full visibility (see Figure 2 for a visual depiction of the design), overlaid with the first practice trial's image at zero visibility. The scrambled image gradually decreases to zero visibility, while simultaneously the first practice trial image gradually increases to full visibility. When the scrambled image reaches zero visibility (and the first trial image reaches full visibility), the scrambled image is replaced by the second practice trial image, which gradually increases from zero visibility to full visibility (as the first practice trial image simultaneously decreases from full visibility to zero visibility). This process is repeated for each practice trial's image, with each new image starting at zero visibility, gradually increasing to complete visibility, and then gradually decreasing to zero visibility again, at which point it is replaced by a new image. At any given moment, there are always two images on the screen overlaying one another, whose visibility sums to 100% (Figure 2). Following the last trial image, the sequence ends with a full-visibility noise image.

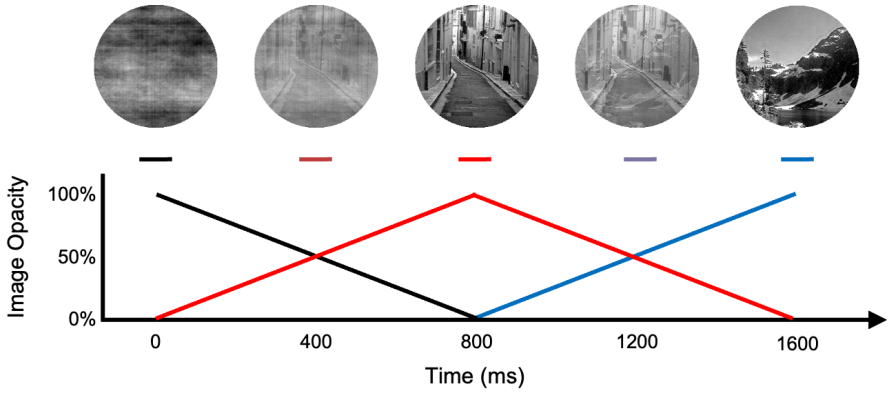

Figure 2: Design overview of image transitions. Image transitions are 800 ms for test trials and the third round of practice trials, and 2400 ms for the second round of practice trials.

The duration of each image's transition from zero visibility to full visibility is 2400 ms for the second round of practice. When a participant correctly makes a response (city trials) or correctly withholds a response (mountain trials), a message reading "Correct" is displayed below the image for 500 ms. When a participant incorrectly makes a response (mountain trials), a message reading "Incorrect" is displayed below the image for 500 ms. Incorrect trials are not repeated from this point in the test onwards.

Participants then receive brief instructions before completing the third and final round of practice trials. This last round of practice trials includes 20 total images (90% city images, 10% mountain images). No feedback is given during the third round of practice trials or for the subsequent test trials. Images gradually transition from one to the next as they did for the second round of practice, except that the duration of each image's transition from zero visibility to complete visibility is 800 ms for the third round of practice, matching the timing of the upcoming test trials.

After completing all three rounds of practice, participants view brief instructions before starting the test trials. The test image sequence includes 300 total images (89.3% city images, 10.7% mountain images), gradually transitioning from one to the next with the same timing as the final round of practice (800 ms to transition from zero visibility to full visibility - see Figure 2).

Data & Analysis Guidelines

Data

As described in Introduction to Cognitive Testing Data in the All of Us Research Program, there are three main categories of data available for cognitive tests: (1) trial-level data, (2) summary scores, and (3) metadata. Please see the Exploring the Mind Data Dictionary for a description of the trial-level data (trial_data), summary score (outcomes), and metadata (metadata) variables for this test (GradCPT).

Suggested Outcomes

The test's suggested primary outcome is *dprime*: a measure of discrimination ability that represents the participant's ability to withhold responses to mountains while making responses to the more prevalent city images. Higher values indicate better performance. Primary outcome *dprime* and secondary outcome *crit* (see below) are based on a signal detection framework - see Stanislaw & Todorov (1999) and the supplement of Fortenbaugh et al. (2015) for instructions on how to calculate *dprime* and *crit*.

As secondary outcomes, researchers should consider analyzing:

- *crit*: response bias - the extent to which a participant is biased to make a response (positive values) or withhold their response (negative values)
- *medianRTc*: median reaction time (ms) of correct responses to city images, a measure of processing speed. Higher values indicate slower reaction times
- *cvRTc*: coefficient of variation of reaction time of correct responses to city images, a measure of processing speed consistency. Lower values indicate more consistent processing speed.

Outcome Type	Outcome Name	Description
Primary	dprime	a measure of response accuracy that represents the participant's ability to withhold responses to mountain images while making responses to city images
Secondary	crit	the extent to which a participant is biased to make a response (positive values) or withhold their response (negative values)
	medianRTc	median reaction time (ms) of correct responses to city images
	cvRTc	coefficient of variation of reaction time of correct responses to city images

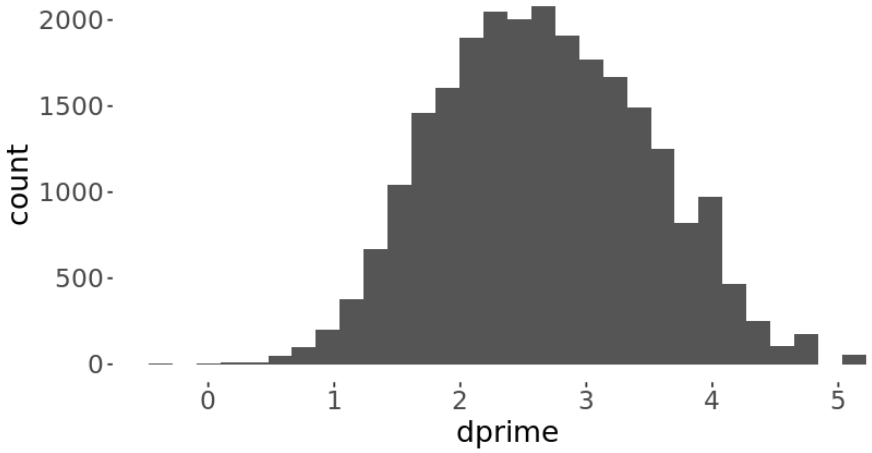

Figure 3: Histogram of primary outcome metric (dprime) for all participants in the CDR v8 off-cycle release

Quality Control Guidelines

The following guidelines are provided for the purpose of flagging extreme deviations in performance from what is typically seen in participants performing the task in a valid manner. Researchers must use their own judgment when determining whether flagged participants should be excluded from analyses. Researchers may also consider implementing their own quality control criteria separately from these recommendations. For more details about quality

control criteria, please see Introduction to Cognitive Testing Data in the All of Us Research Program.

Quality control variables are provided both in trial-level data and full-test outcomes data. The table below summarizes the quality control variables available for this test.

Flag Type	Variable Name	Description
Trial-level	flagged	Indicates whether the duration of the image on the screen was greater than 1600 ms (1 if *trialLength* > 1600, 0 otherwise, null for practice trials). 1600 ms is double the expected trial length for test trials, and suggests the participant switched to a new browser tab or application in the middle of the test before returning to complete the test.
Full-test	flag_nonResponse	Has a value of 1 if the participant, at any point during the test, withheld their response for 75 consecutive test trials (60 seconds), and a value of 0 otherwise. The test requires participants to make frequent responses, and such a long delay between responses indicates that the participant was not properly engaged with the test, or experienced a technical error.
	flag_omissionErrorRate	Has a value of 1 when a participant withholds their response on more than 50% of go test trials (city images), and a value of 0 otherwise. Because 89.3% of trials overall require a response, responding to less than 50% of city images suggests that the participant was not properly engaged with the test, or experienced a technical error.
	flag_trialFlags	Has a value of 1 when the cumulative trial length (*trialLength*) of all flagged test trials (*flagged* = 1) exceeds 60 seconds, and a value of 0 otherwise. This suggests that the participant was taking breaks by switching to a new tab or application for an extended period of time, invalidating their performance on the test.

Gradual Continuous Processing Task (N = 24,858)[1]		
	Yes	**No**
Non-Response Flags	<1%[2]	>99%
Omission Error Rate	<1%	>99%
Trial Flags	<1%	>99%
Any Flag	<1%	>99%

Table 1: Percentage of participants with quality control flags in the Exploring the Mind CDR v8 off-cycle release.

Calculating Test Reliability

To calculate the reliability of performance differences between participants in a given sample, we recommend calculating *split-half reliability* (Pronk et al., 2022) using the following steps:

1. For each participant, separate go-trials (city images) and nogo-trials (mountain images).
2. Mark whether the sequential order of each trial was “even” or “odd.” For example, the first city trial would be marked “odd” and the second city trial would be marked “even.” Correspondingly, the first mountain trial would be marked “odd” and the second mountain trial would be marked “even.”
3. Compute *dprime* separately for odd trials and even trials (see supplement of Fortenbaugh et al., 2015, for instructions for calculating dprime).
4. Compute the Pearson correlation (r) between (1) *dprime* on odd trials and (2) *dprime* on even trials.
5. Use the Spearman-Brown prediction formula to compute full-test reliability: reliability = (2*r) / (1+r)

[1]This count is defined as the total number of unique participants who completed the task.
[2]Due to the data dissemination policy, counts of less than 20 participants cannot be shared publicly. Users can view exact counts in the corresponding featured workspace after logging into their Researcher Workbench account.

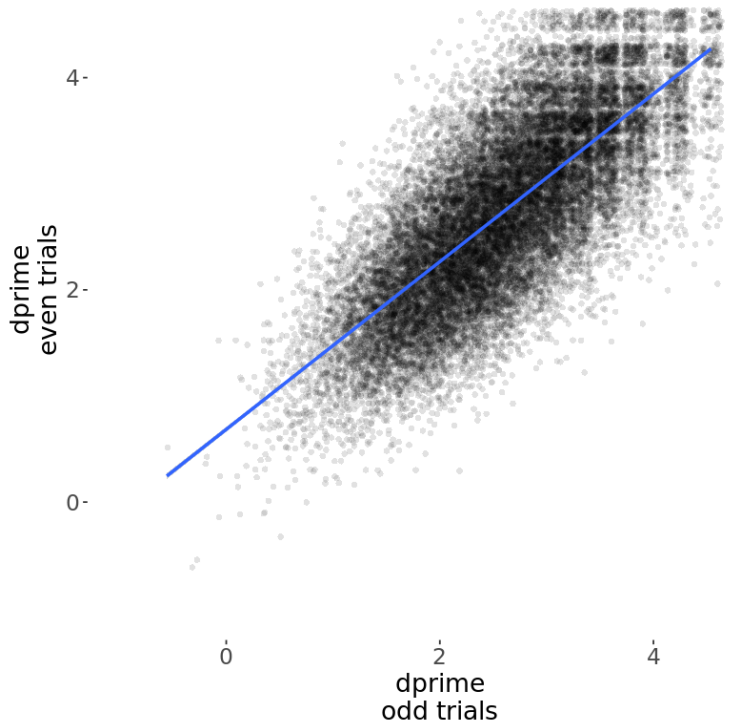

Figure 4: Correlation of participants' dprime on even and odd trials (Spearman-Brown split-half reliability = 0.86)

Correlates of Interest

Prior data collection has found associations between the following demographic variables and TestMyBrain Gradual Onset Continuous Performance Test outcomes. Therefore, researchers may consider including the following variables as covariates in analyses.

1. *age*: Fortenbaugh et al. (2015) reported the following associations with age:
    - *dprime*: sharp improvement with increasing age from age 10-16, followed by gradual improvement from age 17-43, followed by gradual declining performance at older ages
    - *crit*: increased age is associated with less impulsive responding (lower values), starting around age 15
    - *medianRTc*: increased age is associated with higher medianRTc (slower processing speed) starting around age 15
    - *cvRTc*: sharp decrease (indicating more consistent processing speed) from age 10-16, followed by gradual decrease from age 17-43, followed by gradual increase at older ages.
2. *gender*: Riley et al. (2015) reported the following associations with gender identity:
    - *dprime*: small advantage for male participants relative to female participants, on average

- *crit*: on average female participants had lower response bias (less impulsive responding) than male participants
- *medianRTc*: on average male participants had faster reaction times than female participants
- *cvRTc*: on average males had more consistent reaction times than female participants.

3. r*esponse input format*:
    - On average *dprime* and *crit* are higher for participants using keyboard input than for participants using touch input[3]

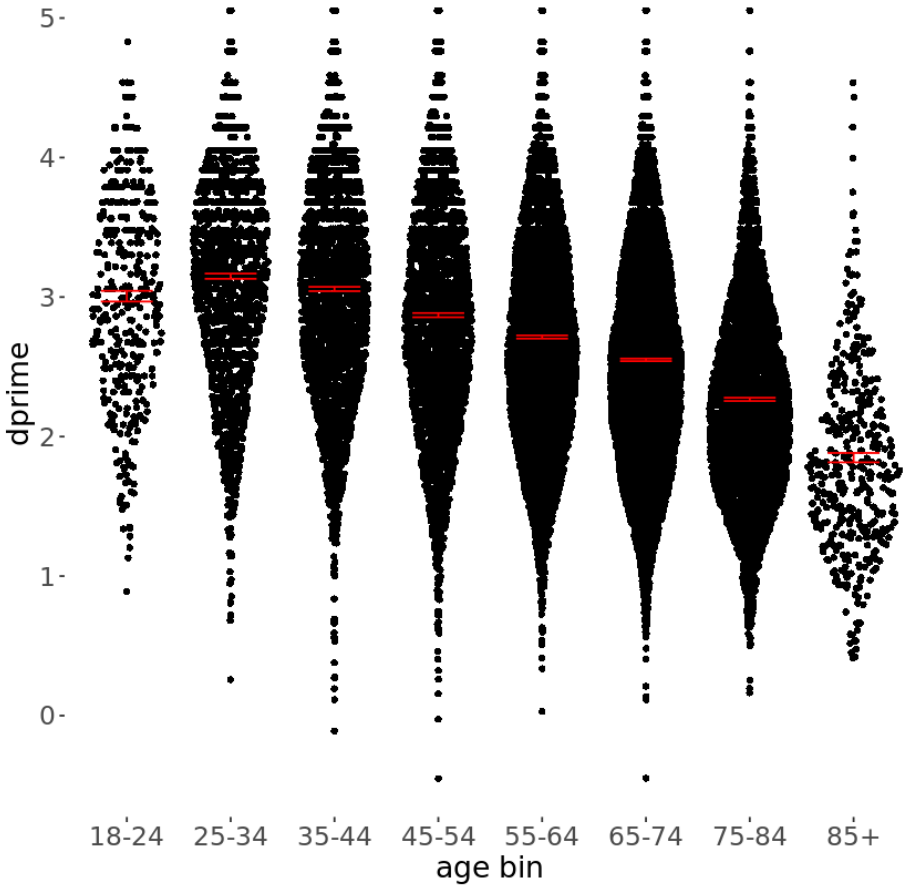

Figure 5: dprime by age bucket. Red lines represent mean dprime for each bucket. Width of distributions (black dots) represent the relative density of participants at each value of dprime.

[3] based on validation data collected by TestMyBrain.org and initial *All of Us* Exploring the Mind data collection

References

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712-2723. https://doi.org/10.1093/cercor/bhs261

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science*, *26*(9), 1497–1510. https://doi.org/10.1177%2F0956797615594896

Hawks, Z. W., Strong, R., Jung, L., Beck, E. D., Passell, E. J., Grinspoon, E., ... & Germine, L. T. (2023). Accurate prediction of momentary cognition from intensive longitudinal data. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(8), 841-851. https://doi.org/10.1016/j.bpsc.2022.12.002

Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. *Psychonomic Bulletin & Review, 29*(1), 44-54. https://doi.org/10.3758/s13423-021-01948-3

Riley, E., Okabe, H., Germine, L., Wilmer, J., Esterman, M., & DeGutis, J. (2016). Gender differences in sustained attentional control relate to gender inequality across countries. *PloS One*, *11*(11), e0165100. https://doi.org/10.1371/journal.pone.0165100

Riley, E., Esterman, M., Fortenbaugh, F. C., & DeGutis, J. (2017). Time-of-day variation in sustained attentional control. *Chronobiology International*, *34*(7), 993-1001. https://doi.org/10.1080/07420528.2017.1308951

Rosenberg, M., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. *Attention, Perception, & Psychophysics, 75*(3), 426-439. https://doi.org/10.3758/s13414-012-0413-x

Singh, S., Strong, R. W., Jung, L., Li, F. H., Grinspoon, L., Scheuer, L. S., Passell, E. J., Martini, P., Chaytor, N., Soble, J. R., & Germine, L. (2021). The TestMyBrain Digital Neuropsychology Toolkit: Development and psychometric characteristics. *Journal of Clinical and Experimental Neuropsychology*, *43*(8), 786-795. https://doi.org/10.1080/13803395.2021.2002269

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. Behavior *Research Methods, Instruments, & Computers, 31*(1), 137-149. https://psycnet.apa.org/doi/10.3758/BF03207704

Treviño, M., Zhu, X., Lu, Y. Y., Scheuer, L. S., Passell, E., Huang, G. C., ... & Horowitz, T. S. (2021). How do we measure attention? Using factor analysis to establish construct validity of neuropsychological tests. *Cognitive Research: Principles and Implications*, *6*(1), 1-26. https://doi.org/10.1186/s41235-021-00313-1

Vogel, S. C., Esterman, M., DeGutis, J., Wilmer, J. B., Ressler, K. J., & Germine, L. T. (2020). Childhood adversity and dimensional variations in adult sustained attention. *Frontiers in Psychology*, 11, 691. https://doi.org/10.3389%2Ffpsyg.2020.00691