***American Journal of Epidemiology* Submitted Manuscript**

Special Collection: N/A

Title: Evidence Supports the Validity and Reliability of Response Times from a Brief Survey as a Digital Biomarker for Processing Speed in a Large Panel Study

Authors: Raymond Hernandez[1], Arthur A. Stone[1,2], Elizabeth Zelinski[3], Erik Meijer[1], Titus Galama[1,4,5], Jessica Faul[6], Arie Kapteyn[1], Doerte U. Junghaenel[1,2,3], Haomiao Jin[1,7], Margaret Gatz[1,2,3], Pey-Jiuan Lee[1], Daniel Maupin[7], Hongxin Gao[7], Bart Orriens[1], Stefan Schneider[1,2,3]

ORCiD IDs: Raymond Hernandez (0000-0002-6761-9276), Arthur A. Stone (), Elizabeth Zelinski (0000-0003-3994-399X), Erik Meijer (0000-0001-7959-4670), Titus Galama (0000-0002-1036-396X), Jessica Faul (0000-0002-6969-6478), Arie Kapteyn(0000-0002-1855-5528), Doerte U. Junghaenel (0000-0002-7733-571X), Haomiao Jin (0000-0001-8908-1959), Margaret Gatz (0000-0002-1071-9970), Pey-Jiuan Lee (0000-0001-6163-2325), Daniel Maupin (0000-0002-0510-6627), Hongxin Gao (0000-0002-6030-1118), Bart Orriens (0000-0002-1304-7735), Stefan Schneider (0000-0002-4562-0524)

Correspondence Address: University of Southern California, Dornsife Center for Self-Report Science 635 Downey Way, Los Angeles, CA, 90089-3332, United States

Joint Authorship: N/A

Affiliations:

[1]Center for Economic & Social Research, University of Southern California, Los Angeles, CA, US

[2]Department of Psychology, University of Southern California, Los Angeles, CA, US

[3]Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, US

[4]School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[5]Tinbergen Institute, Amsterdam, Netherlands

[6]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, US

[7]School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

Key words: processing speed; panel studies; paradata; cognitive health

Acknowledgments[1]: N/A

Funding: This work was supported by the National Institute on Aging at the NIH [grant numbers NIH/NIA R01AG068190 and R01AG079554]. It was also supported by the National Institute of Occupational Safety and Health [grant number 1K01OH012739]. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Conflict of Interest: AAS is a Senior Scientist with the Gallup Organization and a consultant for Lore Contagious Health. The other authors have no conflicts of interest to report.

[1] Study investigators, conference presentations, preprint publication information, thanks.

Disclaimer: N/A

Data Availability Statement: Data from the UAS Study is available via their website (https://uasdata.usc.edu/) after completion of a data use agreement.

Abstract

Survey response times (RTs) have hitherto untapped potential to allow researchers to gain more detailed insights into the cognitive performance of participants in online panel studies. We examined if RTs recorded from a brief online survey could serve as a digital biomarker for processing speed. Data from 9,893 adults enrolled in the nationally representative Understanding America Study were used in the analyses. Hypotheses included that people's average survey RTs would have a large correlation with an established processing speed test, small to moderate correlations with other cognitive tests, and associations with functional impairment. We also hypothesized that survey RTs would have sensitivity to various participant characteristics comparable to the established processing speed test's sensitivity (e.g., similar standardized means by gender). Overall, results support the validity and reliability of people's average RTs to survey items as a digital biomarker for processing speed. The correlation between survey RTs (reverse scored) and the formal processing speed test was 0.61 ($p<0.001$), and small to moderate associations with most other cognitive and functional status measures were observed. Sensitivity of survey RTs to various participant characteristics was nearly identical to the formal processing speed test. Survey RTs may be useful as proxies for processing speed.

Introduction

Digital biomarkers have been defined by the U.S. Food and Drug Administration as measurements collected from digital health technology capturing normal biologic processes, pathologic processes, or biological responses to intervention.[1] Examples of actively collected digital biomarkers include blood pressure readings from a digital sphygmomanometer and cognitive performance scores from computerized cognitive testing.[2] Passively collected digital biomarkers include physical activity from accelerometers,[3] smartphone-based audio recordings of coughs that can give information about respiratory status,[2] and phone behaviors (e.g., swipes, taps, and keystroke events) to infer various aspects of cognitive performance (e.g., working memory and executive function).[4] Digital biomarkers have several potential uses including monitoring the status of a disease or medical condition, assessing people's response to a medical product, and identifying the likelihood that a person experiences a clinically significant event.[2]

Survey RTs

Recently, efforts have been made to use passively collected response times (RTs) to internet survey items as a digital biomarker for cognitive performance.[5] Our use of the word "survey" refers to questionnaires used to elicit reports of facts, attitudes, and other subjective states. Prior research has shown that people with cognitive impairment have longer RTs on formal cognitive tests,[6] so the argument was made that RTs on survey items may also be sensitive to cognitive performance.[7] Longer survey RTs may be caused by slower performance of any of the several cognitive steps often involved in responding to survey items including comprehending questions, retrieving the relevant information from memory, making a judgement based on the retrieved information, and reporting an answer within the confines of the response options provided.[8] In our group's prior work, a moderate correlation was found between average survey item RTs and inhibitory control (r = -.41), and small correlations for quantitative (r = -.09) and verbal reasoning (r = -.20) skills.[7] Systematic RT adjustments (i.e., appropriately taking more time on harder questions and less time on easier ones) had moderate associations with both quantitative (r = .43) and verbal reasoning (r = .33), and a small correlation with inhibitory control (r = .05).[7] Both longer RTs on items and less systematic RT adjustments were found to be predictive of future cognitive impairment.[9]

Survey RTs from large online panel studies have hitherto untapped potential to allow researchers to gain insights into the cognitive performance of study participants even in the absence of formal cognitive testing. With advancements in technology and increased frequency of use of devices like computers and smartphones, panel studies are increasingly being administered online.[10] RTs to survey items can potentially be recorded passively each time panel study members complete an online survey. Such information could allow investigation of relationships between cognitive performance and topics of interest that are administered for the panel study. For instance, the Panel Study of Income Dynamics (PSID) asks questions about a variety of constructs including health conditions, health behaviors, type of employment, and annual income,[11] and since 2021, some interviews have been administered online. For participants who completed multiple online surveys in the PSID, it would be possible to examine how trajectories in cognitive functioning (as captured by survey RTs) relate to the aforementioned constructs. Researchers could examine if for example people with certain types of jobs or health conditions experience greater decline in cognitive performance. Note that even in cross sectional surveys, survey RTs could provide information about individuals' cognitive performance, and that their potential utility is not limited to panel studies.

Several online panel studies do administer formal web based cognitive tests,[12,13] but even in these cases survey RTs can have utility as proxies for cognitive performance. In these studies, formal web based cognitive tests are often administered on an annual or biannual basis to limit burden on participants. Surveys however are often administered more frequently than the formal online cognitive tests, so they can potentially provide more frequent (albeit rougher) measurements of cognitive performance when survey RTs are recorded.

The purpose of this paper is to examine the reliability and validity of RTs on survey items as a digital biomarker specifically for processing speed. Processing speed has been theorized to be a basic aspect of cognitive functioning underlying many other cognitive functions.[14–16] Reduction in processing speed has been posited as major factor contributing to age related decline in cognitive functioning.[14] A prior study found evidence supporting the validity of the average of two weeks of RTs to items in Ecological Momentary Assessments (EMA) as a proxy for processing speed, including a large correlation (r = .58) with a formal smartphone-administered processing speed test, the Symbol Search task.[17] We build on this research by further examining the validity of RTs for non-EMA survey items as a digital biomarker for processing speed using rich data available in a large population-representative U.S. internet panel study.

Hypotheses

Several hypotheses were formulated for validity testing. One of our hypotheses was that people's average survey RTs would have a large correlation (i.e., at least 0.5)[18] with an established processing speed test, the Figure Identification task, which involves matching of a target figure to the correct figure in a set of figures as quickly and accurately as possible.[19] We expected a high correlation, but not redundancy, in part because of the performance expectation in the formal processing speed test that is not present when completing surveys. Also, the Figure Identification task tests specifically for perceptual speed,[20] while survey RTs may be affected by two forms of processing speed (i.e., perceptual and decision speed) and by reading fluency. Processing speed has been defined as a factor underlying various forms of speed (e.g., perceptual speed, decision speed, etc.)[20] but has also often been approximated by perceptual speed tests alone.[20,21]

Two additional hypotheses were that survey RTs should be associated with age, and that survey RTs should have sensitivity to various participant characteristics comparable to the Figure Identification score's sensitivity. A large body of prior literature has found that processing speed slows with age.[20,22] We did not have explicit expectations about the direction of associations between various participant characteristics (e.g., education, race, etc.) and processing speed, but we expected the standardized mean levels of Figure Identification scores for particular participant groups in our sample to be similar to the standardized (and reverse scored) mean levels of survey RTs if the two in fact capture the same construct (i.e., processing speed).

We hypothesized that survey RTs would have small to moderate associations with aspects of cognition other than processing speed (e.g., memory), because of prior literature supporting their relationships with other cognitive domains.[14–16] Furthermore, we anticipated that the magnitude of associations between scores on the formal processing speed test and other cognition measures would be similar to associations between survey RTs and those same cognitive measures.

Finally, we hypothesized that survey RTs would have associations with difficulties with activities of daily living (ADLs; e.g., dressing, meal preparation), based on prior literature suggesting a relationship between processing speed and functional status.[23] We anticipated that the magnitude of associations between scores on the formal processing speed test and functional status measures would be similar to associations between survey RTs and those same functioning measures.

Methods

Study Design

We analyzed data collected between late 2019 and early 2024 in the Understanding America Study (UAS), a nationally representative internet-based survey panel of U.S. adults aged 18 and older[13] currently with about 14,000 participants and ongoing recruitment. Participants who provide electronic informed consent to join the UAS panel are invited to complete various surveys on a monthly basis until they decide to leave the panel or are lost to follow-up. The study was approved by the University of Southern California's Institutional Review Board.

Measures

Formal Processing Speed Test

As mentioned earlier, the Figure Identification task involved matching figures as quickly and accurately as possible.[19] Its score is the number of correctly matched figures within the preset time limit. If respondents failed to provide correct matches to at least 70% of trials (indicating likely carelessness and/or inattention), the score was set to missing.[19] Of 10,114 participants, 221 (2%) had <70% correct. The Figure Identification task was initially administered twice to UAS participants in the same sitting as part of a study investigating the effect of device type on test performance,[19] but more recent UAS members completed the task only once. A majority of the participants represented in this study (7,422/9,893=75%) completed the Figure Identification test twice, but we used only their score in the first test administration in analyses as the second test may have been impacted by fatigue and/or practice effects.

Survey Item RTs

After each administration of the Figure Identification task, the same six questions were asked about the experience of completing the test. This included items about ease of seeing figures on the screen, location where the test was completed, level of technical difficulties experienced, and presence of distractors (items shown in Supplementary Table S1). Five of the items were multiple choice and one had checkboxes. Only these six items were used in the computation of average survey RTs because they were the only ones administered at the same time as the Figure Identification task to all participants. RTs were recorded for each item, log transformed to normalize their distribution, and multiplied by negative one so that higher values could be interpreted as faster performance. If RTs were less than or equal to 1 second, they were set to missing as prior literature has suggested that such fast answers were likely indicative of careless responding.[24] RTs greater than or equal to 1 minute were also set to missing as they were deemed outliers that may have been caused by disruptions in completing the survey. For the 75% of participants represented in this study who completed the Figure Identification task twice, RTs from up to 12 survey items were recorded (i.e., from two administrations of the same six items). RTs from all 12 items were used in these cases for a more reliable RT average. The remaining 25% of participants had RTs from up to 6 survey items recorded.

Cognitive Measures

Aside from the Figure Identification task, several other formal cognitive tests were utilized in the analyses that were typically administered on different days (though only scores within a year of the Figure Identification task completion date were used). The Stop and Go

Switching task consists of four subscores, which assessed choice reaction time, response inhibition, task switching, and executive functioning, respectively.[19] Other measures used were Serial 7s (attention and working memory),[25,26] Number Series (fluid intelligence),[27] Verbal Analogies (fluid intelligence),[27] Picture Vocabulary (word knowledge),[27] and Word Recall (episodic memory).[25,26] For all tests, scores were coded such that greater values meant better cognitive status.

Functional Status

Participants were asked whether or not they had difficulties with activities of daily living (ADLs) or instrumental activities of daily living (IADLs). The ADLs assessed were dressing, eating, walking, getting in/out of bed, using the toilet, and bathing. The IADLs assessed were meal preparation, making a phone call, shopping for groceries, taking medications, house/yard work, and money management. Participants indicating difficulty or an inability to perform any one ADL were classified as having an ADL impairment, and an identical coding was applied for IADL impairment.[28,29]

Statistical Analyses

Correlation between Survey RTs and a Formal Processing Speed Test

To account for potential unreliability in respondents' mean survey RT, latent average survey item RTs were computed using two-level multilevel models in M*plus* version 8.11.[30] In these models, survey item RTs were specified at both level 1 (within-person) and 2 (between person level, where latent average RT was represented), while the formal processing speed test score was specified at level 2. Covariances between the variables at level 2 were estimated, and the standardized versions of these covariances were reported as correlations. Correlations were adjusted by type of device (i.e., computer, mobile phone, or tablet) used to complete surveys and device used for the processing speed test (which could at times be different) by regressing survey RTs and formal processing speed test scores on both.

Participant Characteristics and Survey RTs

To compare age differences in survey item RTs and Figure Identification scores, both the Figure Identification score and the person specific latent average of log transformed survey item RTs were regressed on age in separate linear regression models. Model predicted values for both by age were then outputted and converted to z-scores, with predicted values for survey RTs multiplied by negative one to have the same direction of interpretation as the Figure Identification score. The z-scores were then plotted against age.

For all other categorical participant characteristics (i.e., education, income, race, chronic condition, and gender) the same procedures as above were followed except that survey item RTs and Figure Identification scores were regressed on the dummy codes for categorical variables. The model-predicted means specific to each group were then computed using the "model constraint" statement in M*plus.*

Correlations between Survey RTs and Other Measures

To examine correlations between survey RTs and other measures a multilevel setup identical to the one described above for correlations with the formal processing speed test was used, except that more variables (e.g., cognitive test scores, functional status measures, etc.) and covariances were specified at the between-person level. As done in prior studies, a false discovery rate (FDR) adjustment was applied to the p-values of different groups of hypotheses that were tested (e.g., association between the Figure Identification task and other cognitive measures, association between average survey RTs and other functional measures, etc.).[31] Additionally, we conducted tests to determine the significance of differences in the magnitudes of the (Fisher-z transformed) correlations between survey RTs and other measures and correlations between Figure Identification score and other measures. The differences between the correlations were specified using the "model constraint" statement in M*plus*, allowing for testing whether the differences were statistically significantly different from zero using the delta method.[30]

Reliability

To compute the number of survey RTs required for a reliable estimate of a person's average RT, the formula reliability = *Var(BP)/(Var(BP) + Var(WP)/n)* was used, where *Var(BP)* is the between-person variance in the average of survey item RTs, *Var(WP)* is the variance of survey item RTs within a person, and *n* is the number of measurement occasions.[32]

Exploratory analyses

We examined whether the results of validity testing would differ by use of the following: 1) raw average RTs of items administered at the same time as the formal processing speed test; 2) latent mean of survey RTs within a month of the formal processing speed test; and 3) the raw mean of surveys RTs within a month of the processing speed test. Note that in the UAS, participants had some freedom to choose when and what surveys to complete, so there could be heterogeneity in the survey items they completed within a given time span. Thus, using RTs from surveys completed within a month of the Figure Identification Task had the potential issues of being based on different sets of items for each participant and not being immediately proximal to the Figure Identification task, but the advantage of potentially greater reliability from being based on a larger number of survey RTs. For these exploratory analyses, associated p-values and confidence intervals were not adjusted for multiple comparisons with the understanding that future studies would be needed to replicate the findings.[33]

We also examined the robustness of both the survey RT and Figure Identification correlations across five-year age groups and across self-reported diagnoses. This was to investigate whether correlations with survey RTs approximated correlations with the Figure Identification scores not only in a heterogeneous general population sample, but also within selected subgroups of the population that may be of relevance to clinical researchers and practitioners (e.g., older versus younger adults, people with versus without diabetes).

Finally, prior methodological work has used percentage correct on the Figure Identification Task as part of a screening instrument to detect careless survey responders.[34] Thus, we investigated whether the 221 participants who did not meet the criterion of correct responses on the Figure Identification task would have survey RTs less relevant to cognitive performance. For these careless participants, the correlations between survey RTs and formal cognitive tests were computed.

Results

Data on 9,893 participants were included in the analyses. Participant characteristics are shown in Table 1. See supplementary Figure S1 for the age and diagnostic group distribution of the sample. Across all respondents, the median average RT for the six survey items was 5.63 seconds, the minimum was 4.94 seconds, and the maximum was 8.19 seconds. In terms of the number of items each participant completed, 9,765 (99%) completed at least 6 items and 6,863 (69%) completed 12 items. Average within month survey RTs (used in exploratory analyses) were based on median of 24 items across participants, and had a median average RT of 8.50 seconds.

Survey RTs and a Formal Processing Speed Test

Person average RTs (reverse scored) had a 0.61 ($p<0.001$) correlation with the formal processing speed test. Figure 1 shows a scatterplot between standardized versions of the two variables.

Participant Characteristics and Survey RTs

The mean formal processing test z-score for different subgroups of participant characteristics was nearly identical to the corresponding mean survey item RT z-scores (Figure 2). The formal processing speed test had a -0.54 (p<0.001) correlation with age while survey item RTs had a -0.56 (p<0.001) correlation with age.

Correlations between Survey RTs and Other Measures

Correlations between the Figure Identification task score/Survey RTs and different cognitive measures are shown in Table 2. All correlations were statistically significant at $p<0.05$ (with most $p<0.001$) except for one between the Figure Identification task and word knowledge. The survey RTs had lower correlations with other cognitive functions than the Figure Identification task for all cognitive domains except choice reaction time.

Correlations between the Figure Identification task score/Survey RTs and functional status measures are shown in Table 3. All correlations were statistically significant at $p<0.05$ with the exception of those involving difficulty making phone calls and difficulty managing money. Remarkably, correlations with functional status measures were statistically significantly stronger with average survey item RTs than with the formal processing speed test for 8 of 14 measures.

Reliability

RTs from 4 items were necessary to compute a person average survey RT with a reliability of at least 0.7. Supplementary Figure S2 shows reliability as a function of the number of items used to compute a person's average survey RT.

Exploratory analyses

Raw average survey RTs from items administered at the same time as the formal processing speed test, latent mean of survey RTs within a month of the formal processing speed test, and the raw mean of surveys RTs within a month had correlations with the Figure Identification task of r=0.56 (p<0.001), r=0.59 (p<0.001), and r=0.49 (p<0.001) respectively. The raw average RT of items concurrent with the formal processing speed test and the latent mean of survey RTs within a month of the processing speed test had correlations with cognitive and functioning measures very similar to those seen in the primary analyses (Tables 2 and 3), while the correlations with the raw mean of all survey RTs within a month were slightly attenuated. Average within month survey RTs had differences by participant characteristics nearly identical to those found in the primary analyses (i.e., Figure 2), so they were not shown.

Overall, the correlations between the formal processing test and other measures (e.g., other cognitive domains, functioning) were similar to the corresponding correlations with survey RTs, when considering various participant characteristics (e.g., age, diagnoses; Supplementary Figures S3 through S6). This was evidenced by the often overlapping 95% confidence intervals of the processing speed and survey RT correlations. There were some differences in correlations with the two for some domains of cognitive functioning, in younger individuals (Supplementary Figure S3) and a few diagnostic groups (Supplementary Figure S4).

The 221 "careless" participants had faster response times than the general sample (e.g., a median RT of 4.75 seconds compared to 5.63 seconds for participants included in the primary analyses). Supplementary Table S2 shows the correlations between average survey response times and other cognitive measures for the careless sample, beside the correlations for the primary sample for comparison. Generally, the correlations for the careless sample with other formal cognitive tests were attenuated.

Discussion

Overall, evidence from this study supports the validity and reliability of person average survey RTs from a single survey as a digital biomarker for processing speed.

Validity

As hypothesized, survey RTs had a strong correlation with a formal processing speed test (Figure 1), and sensitivity to various participant characteristics was nearly identical to the formal processing speed test (Figure 2). Also as anticipated, survey RTs were significantly associated with a range of cognitive and functional status measures. However, the formal processing speed test had greater correlations with other cognitive measures as compared to survey RTs. One potential reason for this is that the formal processing speed test and other cognitive measures had explicit performance expectations, while responding to survey items did not. In contrast, survey RTs had stronger associations with functional status measures than the formal processing speed test for half of the ADL/IADL impairment measures. Survey RTs theoretically capture processing speed without explicit performance expectations and hence may be more relevant to typically non-performance pressured daily living tasks than the Figure Identification test.

Reliability

Results of reliability testing suggested that at least four survey RTs were needed for a reliable raw average RT. Since nearly all participants completed at least 6 items, it is unsurprising that the latent and raw average survey RTs showed great similarity in their associations with other variables.

Exploratory Analyses

The latent average of survey RTs within a month of the formal processing speed had correlations with other measures nearly identical to those with the latent average of survey RTs concurrent with the processing speed test. This may point to the flexibility in the types of items that can be used to compute average survey RTs for the purpose of inferring processing speed. In our past work and in this study, items addressing a wide array of constructs (e.g., mood, pain, presence of distractors, etc.) and various response formats (e.g., multiple choice, check all that apply, etc.) yielded average RTs highly correlated with processing speed.[17] Items examined in this paper and our prior work consisted of self-report questions that generally required a modest amount of thought and reflection. Thus, it appears the average of RTs of any “easy” items is sufficient for a proxy of processing speed, but this requires further investigation.

For the subsample of participants who answered carelessly on the Figure Identification task, correlations between average survey RTs and other cognitive measures were attenuated, likely because their carelessness was also present to an extent when answering survey questions.[34] Thus, to reduce the noise in survey RTs from carelessness that can obscure associations with cognitive performance, researchers may consider excluding careless participants and unreasonable values of survey RTs, both approaches utilized in this study.

Limitations

Several factors aside from processing speed may also impact survey RTs including level of fine motor skills, visual acuity, carelessness, and type of device used. Thus, a high average survey RT could indicate low processing speed, poor fine motor skills, poor visual acuity, or some combination thereof. Note however that this limitation applies to formal processing speed tests as well.

More refined methods for careless responder detection exist than those used here, such as latent mixture modelling approaches that consider both survey RTs and response patterns,[35] but exploration and application of these more involved approaches were beyond the scope of this study.

Conclusion

Overall, study results supported the reliability and validity of internet survey RTs as a digital biomarker for processing speed. RTs to online surveys are an untapped resource for information on cognitive performance for participants in large panel studies that may have great potential utility for researchers. For example, they could facilitate examination of questions such as differences in longitudinal trends in processing speed by variables such as job type and health status, even if a formal web-based processing speed test was not repeatedly administered.

References

1. Vasudevan S, Saha A, Tarver ME, Patel B. Digital biomarkers: Convergence of digital health technologies and biomarkers. *npj Digit Med*. 2022;5(1):1-3. doi:10.1038/s41746-022-00583-z

2. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digit Med*. 2019;2(1):1-5. doi:10.1038/s41746-019-0090-4

3. Shim J, Fleisch E, Barata F. Wearable-based accelerometer activity profile as digital biomarker of inflammation, biological age, and mortality using hierarchical clustering analysis in NHANES 2011–2014. *Sci Rep*. 2023;13(1):9326. doi:10.1038/s41598-023-36062-y

4. Dagum P. Digital biomarkers of cognitive function. *npj Digital Med*. 2018;1(1):1-3. doi:10.1038/s41746-018-0018-4

5. Jin H, Junghaenel DU, Orriens B, Lee PJ, Schneider S. Developing Early Markers of Cognitive Decline and Dementia Derived From Survey Response Behaviors: Protocol for Analyses of Preexisting Large-scale Longitudinal Data. *JMIR Research Protocols*. 2023;12(1):e44627. doi:10.2196/44627

6. Gorus E, De Raedt R, Lambert M, Lemper JC, Mets T. Reaction Times and Performance Variability in Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *J Geriatr Psychiatry Neurol*. 2008;21(3):204-218. doi:10.1177/0891988708320973

7. Junghaenel DU, Schneider S, Orriens B, et al. Inferring Cognitive Abilities from Response Times to Web-Administered Survey Items in a Population-Representative Sample. *J Intell*. 2022;11(1):3. doi:10.3390/jintelligence11010003

8. Tourangeau R. The survey response process from a cognitive viewpoint. *Quality Assurance in Education*. 2018;26(2):169-181. doi:10.1108/QAE-06-2017-0034

9. Schneider S, Junghaenel DU, Meijer E, et al. Using Item Response Times in Online Questionnaires to Detect Mild Cognitive Impairment. Taler V, ed. *The Journals of Gerontology: Series B*. 2023;78(8):1278-1283. doi:10.1093/geronb/gbad043

10. Callegaro M, Baker R, Bethlehem J, Göritz AS, Krosnick JA, Lavrakas PJ. Online panel research. In: Callegaro M, Baker R, Bethlehem J, Göritz AS, Krosnick JA, Lavrakas PJ, eds. *Online Panel Research*. 1st ed. Wiley; 2014:1-22. doi:10.1002/9781118763520.ch1

11. Johnson D, McGonagle K, Freedman V, Sastry N. Fifty Years of the Panel Study of Income Dynamics: Past, Present, and Future. *Ann Am Acad Pol Soc Sci*. 2018;680(1):9-28. doi:10.1177/0002716218809363

12. Fisher GG, Ryan LH. Overview of the Health and Retirement Study and Introduction to the Special Issue. Wang M, ed. *Work, Aging and Retirement*. 2018;4(1):1-9. doi:10.1093/workar/wax032

13. Alattar L, Messel M, Rogofsky D. An Introduction to the Understanding America Study Internet Panel. *Soc Sec Bull*. 2018;78(2):13-28.

14. Salthouse T. The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological review*. 1996;103:403-428. doi:10.1037//0033-295X.103.3.403

15. Schubert AL, Frischkorn GT. Neurocognitive Psychometrics of Intelligence: How Measurement Advancements Unveiled the Role of Mental Speed in Intelligence Differences. *Curr Dir Psychol Sci*. 2020;29(2):140-146. doi:10.1177/0963721419896365

16. Fry AF, Hale S. Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*. 2000;54(1-3):1-34. doi:10.1016/S0301-0511(00)00051-X

17. Hernandez R, Hoogendoorn C, Gonzalez JS, et al. Reliability and Validity of Noncognitive Ecological Momentary Assessment Survey Response Times as an Indicator of Cognitive Processing Speed in People’s Natural Environment: Intensive Longitudinal Study. *JMIR mHealth and uHealth*. 2023;11(1):e45203. doi:10.2196/45203

18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Routledge; 1988. doi:10.4324/9780203771587

19. Liu Y, Schneider S, Orriens B, et al. Self-administered Web-Based Tests of Executive Functioning and Perceptual Speed: Measurement Development Study With a Large Probability-Based Survey Panel. *Journal of Medical Internet Research*. 2022;24(5):e34347. doi:10.2196/34347

20. Salthouse TA. Aging and measures of processing speed. *Biological Psychology*. 2000;54(1-3):35-54. doi:10.1016/S0301-0511(00)00052-1

21. Carlozzi NE, Beaumont JL, Tulsky DS, Gershon RC. The NIH Toolbox Pattern Comparison Processing Speed Test: Normative Data. *Archives of Clinical Neuropsychology*. 2015;30(5):359-368. doi:10.1093/arclin/acv031

22. Eckert M. Slowing Down: Age-Related Neurobiological Predictors of Processing Speed. *Frontiers in Neuroscience*. 2011;5. doi:10.3389/fnins.2011.00025

23. Wadley VG, Bull TP, Zhang Y, et al. Cognitive Processing Speed Is Strongly Related to Driving Skills, Financial Abilities, and Other Instrumental Activities of Daily Living in Persons With Mild Cognitive Impairment and Mild Dementia. Newman AB, ed. *The Journals of Gerontology: Series A*. 2021;76(10):1829-1838. doi:10.1093/gerona/glaa312

24. Jaso BA, Kraus NI, Heller AS. Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychological Methods*. Published online September 16, 2021. doi:10.1037/met0000312

25. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol*. 1988;1(2):111-117.

26. Gatz M, Schneider S, Meijer E, et al. Identifying Cognitive Impairment Among Older Participants in a Nationally Representative Internet Panel. *The Journals of Gerontology: Series B*. 2023;78(2):201-209. doi:10.1093/geronb/gbac172

27. Mather N, Jaffe LE. *Woodcock-Johnson IV: Reports, Recommendations, and Strategies*. John Wiley & Sons; 2016.

28. Edwards RD, Brenowitz WD, Portacolone E, et al. Difficulty and help with activities of daily living among older adults living alone with cognitive impairment. *Alzheimer's & Dementia*. 2020;16(8):1125-1133. doi:10.1002/alz.12102

29. Knoll K, Rhee Y, Hamm JM, et al. The Prevalence and Trends of Instrumental Activities of Daily Living Impairments in the United States from 2008–2018. *ADR*. 2023;7(1):271-278. doi:10.3233/ADR-220107

30. Muthén LK, Muthén BO. Mplus user's guide. Eighth Edition. *Los Angeles CA: Muthén & Muthén*. Published online 2017 1998.

31. Veazie PJ. When to Combine Hypotheses and Adjust for Multiple Tests. *Health Serv Res*. 2006;41(3 Pt 1):804-818. doi:10.1111/j.1475-6773.2006.00512.x

32. Nezlek JB. A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*. 2017;69:149-155. doi:10.1016/j.jrp.2016.06.020

33. Althouse AD. Adjust for Multiple Comparisons? It's Not That Simple. *The Annals of Thoracic Surgery*. 2016;101(5):1644-1645. doi:10.1016/j.athoracsur.2015.11.024

34. Schneider S, May M, Stone AA. Careless responding in internet-based quality of life assessments. *Qual Life Res*. 2018;27(4):1077-1088. doi:10.1007/s11136-017-1767-2

35. Ulitzsch E, Pohl S, Khorramdel L, Kroehne U, Von Davier M. A Response-Time-Based Latent Response Mixture Model for Identifying and Modeling Careless and Insufficient Effort Responding in Survey Data. *Psychometrika*. 2022;87(2):593-619. doi:10.1007/s11336-021-09817-7

Tables

Table 1. Study sample characteristics (n=9,893).

Characteristic	n	Mean (SD) or Percentage (%)
Age (years)	9,893	49.4 (16.2)
Gender		
Female	6025	60.9
Male	3867	39.1
Race		
White	7483	76.2
Black	915	9.3
American Indian/Alaska Native	206	2.1
Asian	598	6.1
Pacific Islander	64	0.7
Multi-racial	556	5.7
Hispanic/Latino		
Yes	1629	16.5
No	8269	83.5
Employment status		
Currently working	5452	55.1
Unemployed (laid off/looking)	805	8.1
Retired	1733	17.5
Disabled	566	5.7
Other	1333	13.5
Education		
High school grad or less	2106	21.3
Some college	2188	22.1
Associate's degree	1340	13.5
Bachelor's degree	2480	25.1
Graduate degree	1777	18
Annual household income		
<$25,000	1931	19.5
$25,000-$49,999	2096	21.2
$50,000-$74,999	1805	18.3
$75,000-$99,999	1295	13.1
$100,000-$149,999	1458	14.8
≥$150,000	1287	13.1

Table 2. Correlations between the Figure Identification task/average survey response time and formal cognitive tests. RTs were multiplied by -1 so that higher values were associated with greater processing speed. Significance of differences in the (z-transformed) correlations were tested using the delta method. Correlations were adjusted by type of device used, and false discovery rate adjustments were applied to p-values separately for each column of distinct statistical tests that were part of the primary analyses.

	Corr. with FigID[a]	**Corr. with RT[a]**	**Difference in corr.[a]**	**Sig. diff. ?**	**Corr. with raw mean RT**	**Corr. with RTs within month of FigID**	**Corr. with raw RTs within month**
Stop and Go baseline (choice reaction time, baseline)	0.35 (p<0.001)	0.35 (p<0.001)	0.01 (p=0.483)	N	0.33 (p<0.001)	0.35 (p<0.001)	0.30 (p<0.001)
Stop and Go reverse (response inhibition)	0.41 (p<0.001)	0.39 (p<0.001)	0.03 (p=0.028)	Y	0.38 (p<0.001)	0.39 (p<0.001)	0.35 (p<0.001)
Stop and Go non-switch (task switching)	0.51 (p<0.001)	0.49 (p<0.001)	0.03 (p=0.028)	Y	0.47 (p<0.001)	0.50 (p<0.001)	0.44 (p<0.001)
Stop and Go switch (task switching, executive function)	0.44 (p<0.001)	0.38 (p<0.001)	0.07 (p<0.001)	Y	0.36 (p<0.001)	0.38 (p<0.001)	0.33 (p<0.001)
Serial 7 (attention and working memory)	0.13 (p<0.001)	0.09 (p<0.001)	0.04 (p<0.001)	Y	0.07 (p<0.001)	0.09 (p<0.001)	0.06 (p<0.001)
Number series (fluid intelligence)	0.24 (p<0.001)	0.14 (p<0.001)	0.11 (p<0.001)	Y	0.12 (p<0.001)	0.14 (p<0.001)	0.10 (p<0.001)
Verbal analogies (fluid intelligence)	0.30 (p<0.001)	0.21 (p<0.001)	0.10 (p<0.001)	Y	0.18 (p<0.001)	0.2 (p<0.001)	0.16 (p<0.001)
Picture vocabulary (word knowledge)	0.00 (p=0.982)	-0.10 (p<0.001)	-0.11 (p<0.001)	Y	-0.12 (p<0.001)	-0.11 (p<0.001)	-0.11 (p<0.001)
Immediate word recall	0.20 (p<0.001)	0.09 (p<0.001)	0.11 (p<0.001)	Y	0.09 (p<0.001)	0.08 (p<0.001)	0.08 (p<0.001)
Delayed word recall	0.19 (p<0.001)	0.08 (p<0.001)	0.12 (p<0.001)	Y	0.07 (p<0.001)	0.07 (p<0.001)	0.06 (p<0.001)

Note. Corr.=correlation; FigID= Figure Identification task; RT= survey item response time; Sig. diff.=significant difference
[a]False discovery rate adjusted p-values

Table 3. Correlations between the Figure Identification task/average survey response time and functional status. RTs were multiplied by -1 so that higher values were associated with greater processing speed. Significance of differences in the (z-transformed) correlations were tested using the delta method. Correlations were adjusted by type of device used, and false discovery rate adjustments were applied to p-values separately for each column of distinct statistical tests that were part of the primary analyses.

	Corr. with FigID[a]	**Corr. with RT[a]**	**Difference in corr.[a]**	**Sig. diff. ?**	**Corr. with raw mean RT**	**Corr. with RTs within month of FigID**	**Corr. with raw RTs within month**
ADLs Difficulty dressing	-0.13 (p<0.001)	-0.14 (p<0.001)	-0.02 (p=0.132)	N	-0.13 (p<0.001)	-0.14 (p<0.001)	-0.13 (p<0.001)
Difficulty eating	-0.05 (p<0.001)	-0.04 (p=0.002)	0.01 (p=0.52)	N	-0.04 (p=0.004)	-0.04 (p=0.001)	-0.03 (p=0.017)
Difficulty walking	-0.1 (p<0.001)	-0.14 (p<0.001)	-0.03 (p=0.002)	Y	-0.13 (p<0.001)	-0.14 (p<0.001)	-0.12 (p<0.001)
Difficulty in bed transfers	-0.10 (p<0.001)	-0.12 (p<0.001)	-0.02 (p=0.048)	Y	-0.12 (p<0.001)	-0.12 (p<0.001)	-0.11 (p<0.001)
Difficulty using toilet	-0.09 (p<0.001)	-0.12 (p<0.001)	-0.03 (p=0.004)	Y	-0.11 (p<0.001)	-0.12 (p<0.001)	-0.12 (p<0.001)
Difficulty bathing	-0.12 (p<0.001)	-0.14 (p<0.001)	-0.03 (p=0.015)	Y	-0.13 (p<0.001)	-0.14 (p<0.001)	-0.13 (p<0.001)
At least 1 ADL difficulty	-0.16 (p<0.001)	-0.20 (p<0.001)	-0.03 (p=0.002)	Y	-0.19 (p<0.001)	-0.20 (p<0.001)	-0.17 (p<0.001)
IADLs Difficulty with meal preparation	-0.09 (p<0.001)	-0.10 (p<0.001)	-0.01 (p=0.474)	N	-0.09 (p<0.001)	-0.10 (p<0.001)	-0.08 (p<0.001)
Difficulty making phone calls	-0.01 (p=0.154)	-0.01 (p=0.25)	0.00 (p=0.985)	N	-0.01 (p=0.342)	-0.02 (p=0.191)	-0.01 (p=0.384)
Difficulty grocery shopping	-0.10 (p<0.001)	-0.14 (p<0.001)	-0.04 (p<0.001)	Y	-0.13 (p<0.001)	-0.14 (p<0.001)	-0.12 (p<0.001)
Difficulty taking medications	-0.03 (p=0.01)	-0.04 (p=0.001)	-0.01 (p=0.472)	N	-0.03 (p=0.002)	-0.04 (p<0.001)	-0.03 (p=0.028)
Difficulty with house/yard work	-0.14 (p<0.001)	-0.20 (p<0.001)	-0.05 (p<0.001)	Y	-0.18 (p<0.001)	-0.20 (p<0.001)	-0.17 (p<0.001)
Difficulty managing money	-0.02 (p=0.047)	-0.02 (p=0.116)	0.00 (p=0.964)	N	-0.01 (p=0.262)	-0.02 (p=0.087)	-0.02 (p=0.146)
At least 1 IADL difficulty	-0.15 (p<0.001)	-0.20 (p<0.001)	-0.06 (p<0.001)	Y	-0.19 (p<0.001)	-0.20 (p<0.001)	-0.17 (p<0.001)

Note. ADL= activities of daily living; Corr.=correlation; FigID= Figure Identification task; IADLs= instrumental activities of daily living; RT= survey item response time; Sig. diff.=significant difference
[a]False discovery rate adjusted p-values

Figures

Figure 1. Scatterplot between Figure Identification task z-score and average survey item response time z-score (multiplied by negative one) where r=0.61, p<0.001

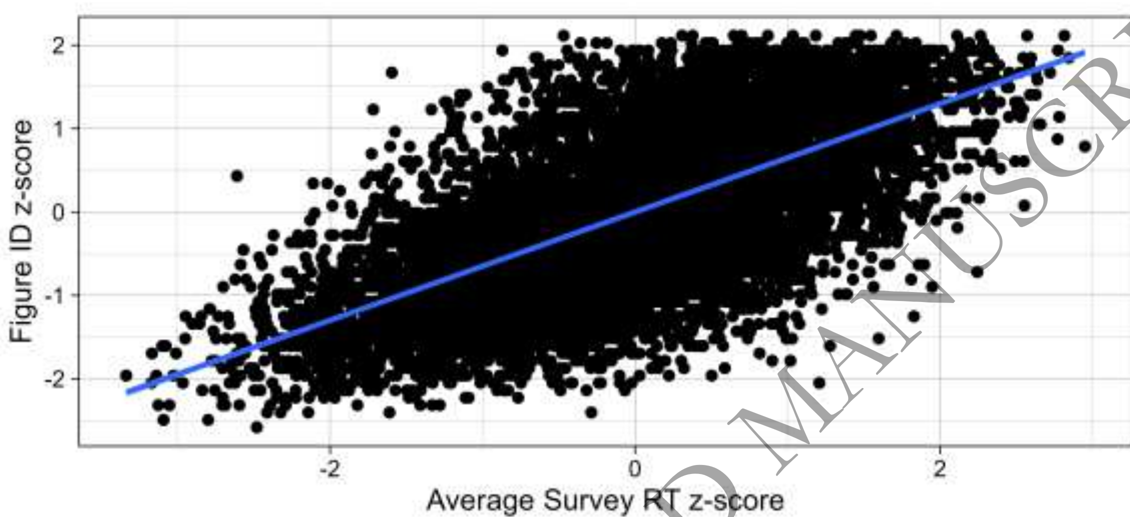

Figure 2. Plots of mean Figure Identification (Figure ID) score/survey item response time (RT) by various participant characteristics. Both the Figure ID and survey item RTs were converted to z scores, and the RTs were also multiplied by negative one (so that higher values were associated with faster processing speed). The dots are observed averages, while the lines and crosses are model derived averages. Confidence intervals (95%) are indicated by colored bands or error bars.