



OPEN

DATA DESCRIPTOR

# Temporal dynamics in psychological assessments: a novel dataset with scales and response times

Zhao Su<sup>1,2</sup>, Rongxun Liu<sup>1,3</sup>, Yange Wei<sup>1,4</sup>, Ran Zhang<sup>1</sup>, Xiao Xu<sup>2</sup>, Yang Wang<sup>1</sup>, Yue Zhu<sup>1</sup>, Lifei Wang<sup>1</sup>, Lijuan Liang<sup>1,5</sup>, Fei Wang<sup>1</sup> <sup>✉</sup> & Xizhe Zhang<sup>1,2</sup> <sup>✉</sup>

The growing prevalence of mental health issues underscores the need for innovative screening methods. Large-scale, internet-based psychological screening has emerged as a vital tool to accurately determine morbidity rates and facilitate early diagnosis of mental disorders. However, conventional psychological screening methods often struggle with non-genuine responses and lack objective metrics. To bridge this gap, we have compiled a novel dataset derived from an expansive screening initiative at Xinxiang Medical University. The study, conducted from February 27 to March 17, 2021, yielded a dataset comprising responses from 24,292 students to four well-established psychological scales—PHQ-9, GAD-7, ISI, and PSS. A distinctive feature of this dataset is the inclusion of response time data, which captures the temporal dynamics of participants' interactions with the scales, offering valuable insights into their response behaviour. The release of this dataset offers a substantial opportunity for researchers in the domains of psychology and public health to explore new insights into mental health, scale reliability, and the dynamics of psychological assessment.

## Background & Summary

In recent years, the incidence of psychological disorders has been on the rise, prompting widespread societal concern<sup>1</sup>. Concurrently, with advancements in technology and the ubiquity of the internet, online psychological assessments have emerged as a practical tool for preliminary screenings of mental health status in the general population<sup>2</sup>. These online evaluations not only offer individuals a convenient and swift avenue to understand their psychological well-being, but they also effectively cover a broader demographic. Additionally, early detection and diagnosis through such platforms can lead to timely interventions and treatments, potentially mitigating the progression of the disorder and its impact on an individual's quality of life<sup>3</sup>. Thus, leveraging online psychological assessments as an early screening mechanism holds significant promise in addressing the escalating challenges of mental health disorders.

While large-scale psychological screenings are essential, the assessment scales they utilize inherently pose challenges. Due to influences such as the environment, internet-based assessment systems still exhibit a certain gap in their results compared to clinical diagnoses<sup>4</sup>. For instance, studies have shown that for screening purposes, using internet-based assessment systems can yield reliable diagnostic outcomes, but often at the expense of sensitivity<sup>5</sup>. During large-scale screenings, where questionnaires are answered in a free environment, a certain percentage of individuals might not answer the questions earnestly. The proportion of careless response (CR) varies widely due to numerous factors, ranging from 1% to 50%<sup>6–10</sup>. A meta-analysis study on alcohol indicated that 11.7% of the sample comprised CRs<sup>11</sup>. Such a rate of CRs can also significantly impact the research outcomes. A particular study revealed that partialling out CRs significantly changed the observed relationships among substantive variables<sup>12</sup>. Maniaci and Rogge found a notable increase in the power of regression after removing samples with CRs<sup>13</sup>. Therefore, identifying and handling CR is a crucial step before analysis.

<sup>1</sup>Early Intervention Unit, Department of Psychiatry, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing, China. <sup>2</sup>School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China. <sup>3</sup>School of Psychology, Xinxiang Medical University, Xinxiang, Henan, China. <sup>4</sup>Department of Psychiatry, The Second Affiliated Hospital of Xinxiang Medical University, Henan Mental Hospital, Xinxiang, China. <sup>5</sup>Laboratory of Psychology, The First Affiliated Hospital of Hainan Medical University, Haikou, Hainan, China. ✉e-mail: [fei.wang@yale.edu](mailto:fei.wang@yale.edu); [zhangxizhe@njmu.edu.cn](mailto:zhangxizhe@njmu.edu.cn)

Response time is a frequently used metric in cognitive psychology experiments, represents the duration required for a participant to react to a stimulus or complete a given task<sup>14</sup>. In the realm of cognitive psychology, response time often serves as a dependent variable, influenced by other independent variables like stimulus exposure duration<sup>15</sup>. Numerous datasets from such psychological experiments are publicly accessible<sup>16,17</sup>. In contemporary internet-based scale screening systems, it's feasible to collect the time participants take to answer each query, thus deriving a response time sequence as a form of objective behavioural data. In a 2019 Kaggle competition centred around a survey in the data science domain, the total response time participants took to complete the questionnaire was documented as a pivotal variable<sup>18</sup>. This data affords researchers a fresh perspective, enabling exploration from a novel, more objective dimension than solely the options chosen by participants. Without this additional layer of data, the subtle intricacies between individuals might remain elusive. Research has indeed delved into the interplay between response time and participants' selections<sup>19</sup>.

By recording response times when participants answer scales, this data can be utilized to identify CRs<sup>20</sup>. Furthermore, since self-assessment scales often face issues with participant subjectivity bias, response times can serve as an indicator of a participant's level of certainty regarding an answer. To delve deeper, since traditional mental health measurement methods based on aggregate or dimensional scores tend to neglect heterogeneity<sup>21</sup>, recent years have witnessed the advent of many computerized adaptive testing scale forms rooted in Item Response Theory (IRT)<sup>22</sup>. Response times, as objective data, can also supplement such computerized adaptive testing. In our prior studies, we employed response time data to architect machine learning models predicting Insomnia Index (ISI) scale outcomes, achieving an accuracy of 0.743<sup>23</sup>. We observed a conspicuous absence of large-scale datasets meticulously recording response times for each question in the field of psychology. This gap poses a hindrance to the in-depth exploration of datasets related to extensive psychological screenings.

In 2021, a detailed psychological screening was conducted at Xinxiang Medical University, encompassing over 20,000 participants. This study involved the collection of data using four established scales - the Patient Health Questionnaire-9 (PHQ-9), the General Anxiety Disorder-7 (GAD-7), the Insomnia Severity Index (ISI), and the Perceived Stress Scale (PSS) - while meticulously recording the response times for each question administered to the subjects. This dataset serves several pivotal roles: 1. Investigate the mental health status of Chinese university students during the COVID-19 pandemic; 2. Identifying invalid responses, which can be gleaned from irregularities in response times<sup>24-27</sup>; 3. Classifying subtypes based on patterns of answering behaviour; 4. Informing the redesign of scales based on response patterns, thereby enhancing the efficiency and accuracy of the questionnaires. This endeavour bridges traditional assessment methodologies with modern data-driven approaches, promising a more nuanced understanding of psychological health metrics.

## Methods

In the pursuit of collecting a large-sample dataset, this study was meticulously designed to gauge the emotional states of university students across four distinct dimensions: depression, anxiety, sleep quality, and stress. Additionally, response times to survey questions were recorded to provide an objective measure of behavioural data. The data collection process spanned 3 weeks, ensuring a comprehensive capture of the targeted metrics.

**Data collection.** The data was collected from a psychological screening project organized by Xinxiang Medical University, which was conducted across the entire campus. Notifications were disseminated to students in each class via their counselors. Participation in the project was voluntary for the students. This study has been approved by the Ethics Committee of Xinxiang Medical College of Henan Province (XYLL-2020235). Informed consent was obtained online from all participants prior to survey administration, including agreement for data sharing and dissemination of their data for scientific research. We ensured adherence to all necessary ethical considerations and informed consent procedures throughout the study.

The dataset targeted the university's enrolled undergraduate and graduate students over approximately three weeks. Due to the constraints imposed by the COVID-19 pandemic, all data collection activities were limited to campus premises. Students were given the flexibility to choose the timing and location for completing the questionnaires. At the conclusion of the data collection phase, a total of 24,367 participants had successfully completed the survey.

**Collection settings.** *Data collection apparatus.* A web-based application was developed for the purpose of this study. Participants were able to access the survey via a QR code or a URL in their web browser. This application served as the medium for both administering the questionnaires and collecting the ensuing data. Upon entering the application, participants first completed the informed consent form. Following this, they completed the survey in the following order: demographic questions - > PHQ-9 - > GAD-7 - > PSS - > ISI.

*Measurement instruments.* The psychological scales employed in this study included the Patient Health Questionnaire-9 (PHQ-9)<sup>28,29</sup>, the General Anxiety Disorder-7 (GAD-7)<sup>30,31</sup>, the Perceived Stress Scale (PSS)<sup>32,33</sup> and the Insomnia Severity Index (ISI)<sup>34,35</sup>. We used the Chinese versions of these scales. Detailed descriptions are shown in Supplement Table. The complete content of the scales, including the text of the questions, response options, and scoring details, can be found in the GitHub repository.

*Data collection setting.* Students were granted the autonomy to select the time and place of their choosing to complete the survey, which allowed for a free and unstructured data collection environment.

*Data generation process.* During the survey, the web application diligently recorded each participant's response to every item. This was accompanied by the precise measurement of the response time for each question. The

File name	context
demographic.csv	<b>export_id</b> : A unique identifier for each participant.
	<b>gender</b> : Gender of the participant.
	<b>age</b> : Age of the participant represented as a number.
	<b>edu</b> : Education level of the participant.
	<b>smoke</b> : Smoking habits of the participant.
	<b>drink</b> : Drinking habits of the participant.
phq9.csv	<b>export_id</b> : A unique identifier matching the demographic.csv.
	<b>score</b> : The aggregate score of the PHQ9 questionnaire.
	<b>questionX</b> : Score for the Xth question.
	<b>timeX</b> : Time taken (in seconds) for the participant to answer the Xth question.
gad7.csv	<b>export_id</b> : A unique identifier matching the demographic.csv.
	<b>score</b> : The aggregate score of the GAD7 questionnaire.
	<b>questionX</b> : Score for the Xth question.
	<b>timeX</b> : Time taken (in seconds) for the participant to answer the Xth question.
pss.csv	<b>export_id</b> : A unique identifier matching the demographic.csv.
	<b>score</b> : The aggregate score of the PSS questionnaire.
	<b>questionX</b> : Score for the Xth question.
	<b>timeX</b> : Time taken (in seconds) for the participant to answer the Xth question.
isi.csv	<b>export_id</b> : A unique identifier matching the demographic.csv.
	<b>score</b> : The aggregate score of the ISI questionnaire.
	<b>questionX</b> : Score for the Xth question.
	<b>timeX</b> : Time taken (in seconds) for the participant to answer the Xth question.

**Table 1.** The content and structure of dataset files.

	Coefficient	Std. Error	z-value	p-value	95% CI
(Intercept)	2.662	0.011	233.506	<0.001	2.640–2.684
Word Count	0.075	0.001	117.010	<0.001	0.074–0.076

**Table 2.** The results from the Generalized Linear Mixed Model (GLMM) analysing the effect of word count on response time, with subject as a random effect.

methodology involved logging a timestamp upon the participant’s arrival at a new question, followed by a subsequent timestamp upon selection of a response. The interval between these two timestamps, denoted in seconds and precise to two decimal places, constituted the response time for the respective question.

**Data quality control.** Participants were required to complete all demographic information and fully respond to all four scales. Out of the total 24,367 collected responses, 75 participants were excluded due to missing values. Consequently, the final dataset includes 24,292 participants. It was acknowledged that network conditions could potentially influence response time recordings, however, given that data collection was conducted within the confines of a single campus, regional disparities were considered negligible, thereby minimizing the impact of network variability on the data collection process.

Ultimately, the dataset comprised 24,292 participants, including 8,747 males and 15,545 females, with an average age of 20.65 (SD = 2.4) years. Each participant provided complete responses to all four scales, resulting in a collective total of 898,804 item responses.

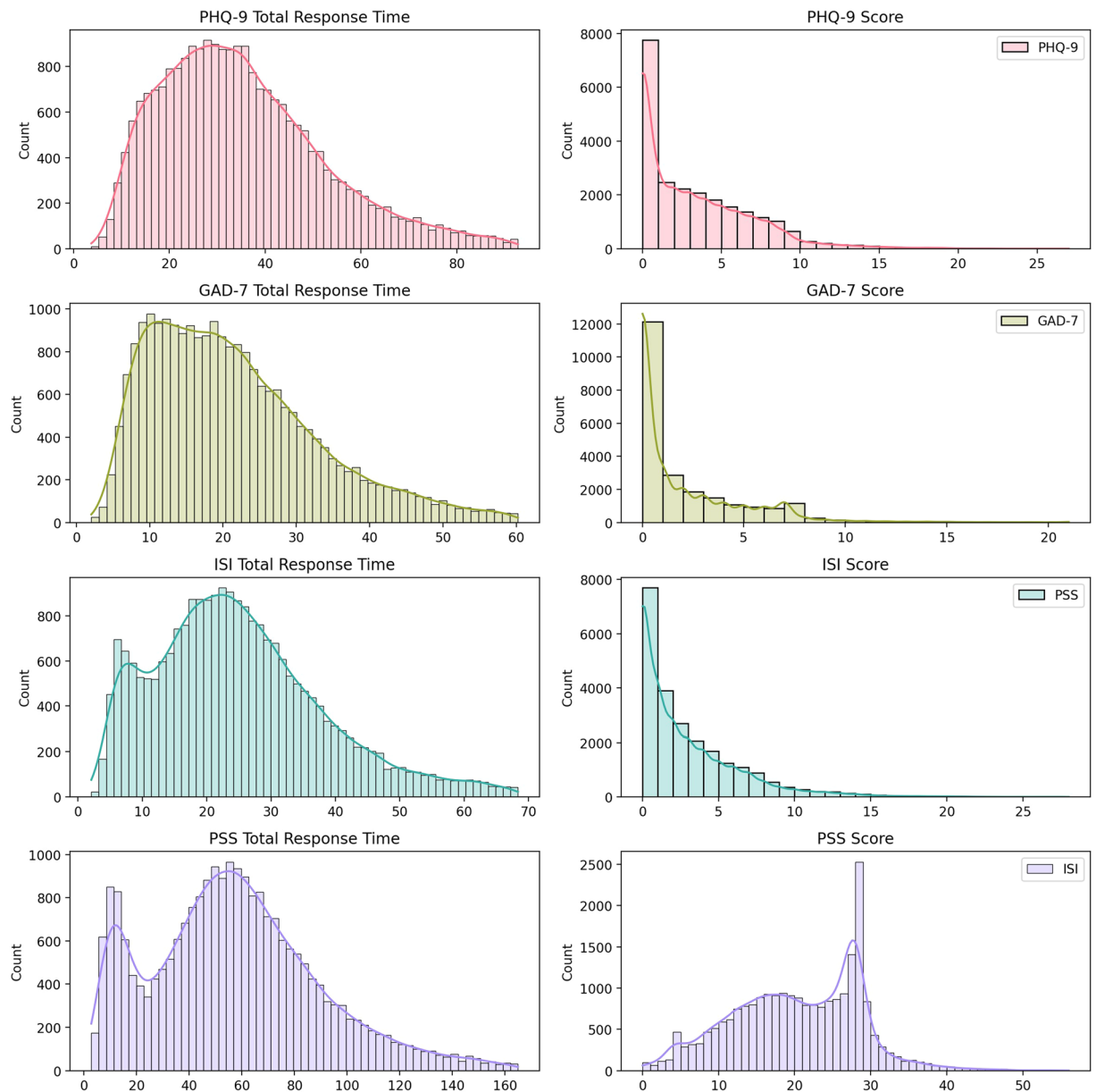
**Data Records**

All files (demographics.csv, phq9.csv, gad7.csv, pss.csv and isi.csv) have been anonymized and both the raw and cleaned data are available in CSV (non-proprietary) formats on the Zenodo platform<sup>36</sup>. All these materials are available at the following link: <https://zenodo.org/records/10423537>.

In the provided dataset, five distinct CSV files capture the data collected from the research participants. Below, each file’s content and structure list in Table 1.

**Technical Validation**

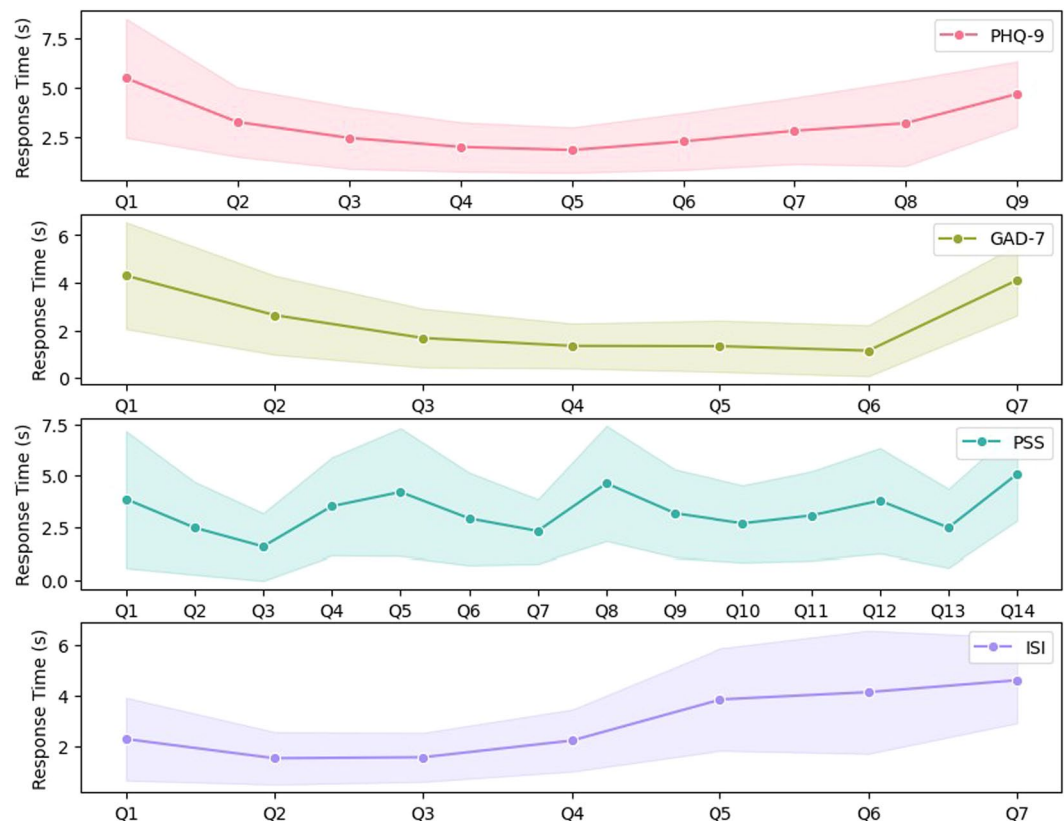
**Descriptive validation.** In the comprehensive psychological screening conducted, data from a total of 24,292 individuals were collected across four distinct scales: PHQ-9, GAD-7, ISI, and PSS. For the PHQ-9 scale, the distribution is as follows: 16,886 individuals (69.51%) reported minimal symptoms, 6,090 (25.07%) displayed mild symptoms, 957 (3.94%) indicated moderate symptoms, 262 (1.08%) had moderately severe symptoms, and 97 (0.40%) exhibited severe symptoms. Within the GAD-7 scale, 20,170 participants (83.03%) had minimal



**Fig. 1** Distributions of total response times and scores for four different psychological assessments: PHQ-9, GAD-7, ISI, and PSS. Each subplot presents a histogram overlaid with a kernel density estimate (KDE) to visualize the distribution of data. (a,b) show the total response time and scores for the PHQ-9, respectively; (c,d) for the GAD-7; (e,f) for the ISI; and (g,h) for the PSS. Different colours are used to distinguish between the four assessments. Legends indicate the respective assessment tool in each subplot.

anxiety symptoms, 3,576 (14.72%) reported mild symptoms, 332 (1.37%) had moderate symptoms, 157 (0.65%) showcased severe symptoms, and a minority of 57 individuals (0.23%) had very severe symptoms. For the ISI scale, a significant majority of 22,090 individuals (90.94%) exhibited no clinically significant insomnia, while 1,922 (7.91%) had subthreshold insomnia, 242 (1.00%) had clinical insomnia (moderate severity), and 38 (0.16%) displayed clinical insomnia (severe). Lastly, for the PSS scale, 6,607 respondents (27.20%) showed low stress, 14,884 (61.27%) exhibited moderate stress, 2,687 (11.06%) indicated high perceived stress, and 114 (0.47%) were classified under the very high perceived stress category.

For the subsequent analyses, we calculated every scale response time's median and median absolute deviation (MAD) and then excluded participants whose standardized deviation exceeded five times the MAD. Figure 1 illustrates the response time distributions (Fig. 1a,c,e,g) and score distributions (Fig. 1b,d,f,h) for the four scales. It can be observed that the response time distributions for the ISI and PSS scales exhibit distinct bimodal distributions, with clear inflection points at approximately 12 seconds for the ISI and 23 seconds for the PSS. For participants with shorter response times, we tend to consider the presence of careless responding behaviour, and this group should be treated with caution during analysis. Figure 2 shows the response times for individual items on the four scales and the overall variation in response times as the questions progress.



**Fig. 2** Median response times with error bands for individual questions across four psychological assessment scales: PHQ-9, GAD-7, PSS, and ISI. Each panel presents a line plot of the median response time (in seconds) for each question within the respective scale. The error bands represent standard deviations (scaled by 0.5) around the median estimates.

**Quantitative validation.** We quantitatively validated our dataset via a word counts effect on response time. To address the issue of non-negative and right-skewed distribution of response time, we performed generalized linear mixed model (GLMM) analyses with word count as an independent continuous variable, and RT as the dependent variable, incorporating subject as random effects. GLMM allows fitting data without the normality assumptions and the need for transformation. It allows differences between individuals to be properly assessed<sup>37</sup>. Analysis was performed using the statsmodels<sup>38</sup> package in the Python 3.9 environment.

The fixed effect of the intercept was estimated at 2.662 (SE = 0.011,  $z = 233.506$ ,  $p < 0.001$ ), indicating the baseline response time. The coefficient for word count was 0.075 (SE = 0.001,  $z = 117.010$ ,  $p < 0.001$ ), suggesting a significant positive relationship between word count and response time. Specifically, for each additional word in the count, the response time increased by approximately 0.075 seconds. The random effects included a group variance of 0.849 and a covariance between group intercept and word count of 0.040, with a variance for word count of 0.002. These results indicate that while there is significant between-group variability, the relationship between word count and response time remains robust. The detailed results are shown in Table 2.

## Usage Notes

**Data overview.** This dataset provides a comprehensive snapshot of the psychological evaluations conducted between February 27, 2021, and March 17, 2021, at Xinxiang Medical University. It comprises data from 24,292 students who utilized a mobile application to respond to items from key psychological scales. These scales include the PHQ9, GAD7, ISI, and PSS. Alongside the primary responses, we also captured data related to the time participants took to respond, offering an added dimension to the dataset.

**Potential applications.** The depth and breadth of this dataset, with its robust sample size of  $N = 24,292$ , make it an invaluable asset for numerous research initiatives. The dataset offers opportunities to explore various research questions, from examining the prevalence of specific mental health symptoms among university students to probing correlations between different psychological conditions. Importantly, the data was collected during the COVID-19 pandemic, providing a unique context to investigate the impact of this global crisis on mental health.

Furthermore, the inclusion of response time data adds a novel dimension to traditional psychological evaluations. It allows for an investigation into participants' engagement during assessments and comprehension of their thought processes. In large-scale psychological screenings, careless responding is likely present in all survey data<sup>6</sup>. Factors such as survey design and incentive methods can contribute to this issue, and response time is an



effective means of identifying such behaviour<sup>6</sup>. Additionally, by analysing continuous response time sequences, researchers can model and classify participants' response behaviours, enabling more precise interpretation of psychological screening results. Furthermore, this analysis can inform the development of more efficient scales and questionnaires, thereby improving the overall accuracy and effectiveness of psychological screenings.

**Limitations.** While the dataset presents a robust resource, its utilization comes with several caveats. Firstly, the inherent nature of self-report questionnaires introduces potential biases, such as social desirability or recall bias, where participants might either opt for socially acceptable answers or face challenges in accurately recalling past experiences. Additionally, the data is predominantly derived from students at Xinxiang Medical University, potentially limiting its generalizability to broader populations or students from different cultural and socio-economic backgrounds. Lastly, the inclusion of response time data, though insightful, brings its own challenges. The factors influencing response times can range from genuine contemplation to mere distractions, necessitating that researchers handle this data judiciously, ensuring they account for potential variables in their analyses.

### Code availability

All code for formatting, cleaning, and quality assurance was written in Python (python.org) with use of the NumPy (numpy.org) and Pandas (pandas.pydata.org) libraries. Code is available on <https://github.com/njnklab/response-time-dataset>.

Received: 25 December 2023; Accepted: 17 September 2024;

Published online: 27 September 2024

### References

1. Organization, W. H. *World Mental Health Report: Transforming Mental Health for All*. <https://archive.hshsl.umaryland.edu/handle/10713/20295> (2022).
2. Barak, A. Chapter Twelve - Internet-based Psychological Testing and Assessment. in *Online Counseling (Second Edition)* (eds. Kraus, R., Stricker, G. & Speyer, C.) 225–255. <https://doi.org/10.1016/B978-0-12-378596-1.00012-5> (Academic Press, San Diego, 2011).
3. Kelly, C. M., Jorm, A. F. & Wright, A. Improving mental health literacy as a strategy to facilitate early intervention for mental disorders. *Medical Journal of Australia* **187**, S26–S30 (2007).
4. Andersson, G. & Titov, N. Advantages and limitations of Internet-based interventions for common mental disorders. *World Psychiatry* **13**, 4–11 (2014).
5. Nguyen, D. P., Klein, B., Meyer, D., Austin, D. W. & Abbott, J.-A. M. The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders. *Journal of Medical Internet Research* **17**, e4195 (2015).
6. Ward, M. K. & Meade, A. W. Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annual Review of Psychology* **74**, 577–596 (2023).
7. Brühlmann, F., Petralito, S., Aeschbach, L. F. & Opwis, K. The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology* **2**, 100022 (2020).
8. Goldammer, P., Annen, H., Stöckli, P. L. & Jonas, K. Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly* **31**, 101384 (2020).
9. Johnson, R., Barnhardt, J. & Zhu, J. Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research* **24**, 386–404 (2005).
10. Meade, A. W. & Craig, S. B. Identifying careless responses in survey data. *Psychological Methods* **17**, 437–455 (2012).
11. Jones, A. *et al.* Careless responding in crowdsourced alcohol research: A systematic review and meta-analysis of practices and prevalence. *Experimental and Clinical Psychopharmacology* **30**, 381–399 (2022).
12. Huang, J. L., Liu, M. & Bowling, N. A. Insufficient effort responding: examining an insidious confound in survey data. *J Appl Psychol* **100**, 828–845 (2015).
13. Maniaci, M. R. & Rogge, R. D. Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality* **48**, 61–83 (2014).
14. Gale, C. R., Harris, A. & Deary, I. J. Reaction time and onset of psychological distress: the UK Health and Lifestyle Survey. *J Epidemiol Community Health* **70**, 813–817 (2016).
15. Kosinski, R. J. A literature review on reaction time. *Clemson University* **10**, 337–344 (2008).
16. Lewis, R. Moral Stroop, experimental psychology - dataset by robertjoellewis. (2017).
17. Brůha, P. *et al.* Collection of human reaction times and supporting health related data for analysis of cognitive and physical performance. *Data in Brief* **17**, 469–511 (2018).
18. Mooney, P. 2019 kaggle machine learning & data science survey. (2019).
19. Chung, K., Park, J. Y., Joung, D. & Jhung, K. Response Time as an Implicit Self-Schema Indicator for Depression Among Undergraduate Students: Preliminary Findings From a Mobile App-Based Depression Assessment. *JMIR Mhealth Uhealth* **7**, e14657 (2019).
20. Bowling, N. A. *et al.* Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *J Pers Soc Psychol* **111**, 218–229 (2016).
21. Gibbons, R. D., Weiss, D. J., Frank, E. & Kupfer, D. Computerized Adaptive Diagnosis and Testing of Mental Health Disorders. *Annual Review of Clinical Psychology* **12**, 83–104 (2016).
22. Reise, S. P. & Waller, N. G. Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology* **5**, 27–48 (2009).
23. Su, Z. *et al.* Exploring the relationship between response time sequence in scale answering process and severity of insomnia: A machine learning approach. *Heliyon* **10** (2024).
24. Börger, T. Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment. *Environ Resource Econ* **65**, 389–413 (2016).
25. Bunji, K. & Okada, K. Item response and response time model for personality assessment via linear ballistic accumulation. *Jpn J Stat Data Sci* **2**, 263–297 (2019).
26. Gogami, M., Matsuda, Y., Arakawa, Y. & Yasumoto, K. Detection of Careless Responses in Online Surveys Using Answering Behavior on Smartphone. *IEEE Access* **9**, 53205–53218 (2021).
27. 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* **87**, 593–619 (2022).
28. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9. *Journal of General Internal Medicine* **16**, 606–613 (2001).

29. Wang, W. *et al.* Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry* **36**, 539–544 (2014).
30. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* **166**, 1092–1097 (2006).
31. Gong, Y. *et al.* Validation of the 7-item Generalized Anxiety Disorder scale (GAD-7) as a screening tool for anxiety among pregnant Chinese women. *Journal of Affective Disorders* **282**, 98–103 (2021).
32. Cohen, S., Kamarck, T. & Mermelstein, R. A Global Measure of Perceived Stress. *Journal of Health and Social Behavior* **24**, 385–396 (1983).
33. Huang, F. *et al.* Psychometric properties of the perceived stress scale in a community sample of Chinese. *BMC Psychiatry* **20**, 130 (2020).
34. Bastien, C. H., Vallières, A. & Morin, C. M. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine* **2**, 297–307 (2001).
35. Yu, D. S. F. Insomnia Severity Index: psychometric properties with Chinese community-dwelling older people. *Journal of Advanced Nursing* **66**, 2350–2359 (2010).
36. Su, Z. Data from: Temporal dynamics in psychological assessments: a novel dataset with scales and response times. *Zenodo* <https://doi.org/10.5281/zenodo.10423537> (2023).
37. Lo, S. & Andrews, S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* **6**, (2015).
38. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. in 92–96, <https://doi.org/10.25080/Majora-92bf1922-011> (Austin, Texas, 2010).

## Acknowledgements

This study is funded by National Natural Science Foundation of China (62176129 to Xizhe Zhang), National Science Fund for Distinguished Young Scholars (81725005 to Fei Wang), the National Natural Science Foundation Regional Innovation and Development Joint Fund (U20A6005 to Fei Wang), Jiangsu Provincial Key Research and Development Program, China (BE2021617 to Fei Wang), Henan Provincial Research and Practice Project for Higher Education Teaching Reform (2021SJGLX189Y to Rongxun Liu).

## Author contributions

Xizhe Zhang and Fei Wang: Conceived and designed the study, oversaw data collection, reviewed and revised the manuscript. Rongxun Liu, Yange Wei, Ran Zhang, Xiao Xu, Yang Wang, Yue Zhu, Lifei Wang, and Lijuan Liang: Collected the data. Zhao Su: Analysed the data and wrote the manuscript. All authors assisted in drafting and editing the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03888-8>.

**Correspondence** and requests for materials should be addressed to F.W. or X.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024