scientific reports

OPEN

Comparing ChatGPT and validated questionnaires in assessing loneliness and online social support among college students: a cross-sectional study

Juan Gu[1], Jiali Liu[1], Lijuan Zeng[1], Yiqing Yu[1], Yufei Qiu[1], Yake Yue[1], Mengjie Tong[1], Fen Yang[1,2,3✉] & Xiaohong Zhang[2,3,4,5✉]

The capability of ChatGPT to understand and generate human-readable text has prompted the investigation of its potential as mental health assessment tools. This study aims to explore the validity of ChatGPT in assessing loneliness and online social support among college students by comparing scoring consistency between ChatGPT and the validated questionnaires. This was a cross-sectional study between June and August 2024. We pre-trained ChatGPT-4 based on the validated University of California Los Angeles Loneliness Scale-6 (ULS-6) and Chinese Youth Version of the Online Social Support Scale (OSSS-CS), creating a structured interview questionnaire. Participants were invited to complete both the ChatGPT-created questionnaire and the validated questionnaires. We used Spearman correlation analysis, Intra-class correlation coefficients (ICC), and Bland–Altman plots to assess the agreement between the scores from ChatGPT and the validated questionnaires. In addition, we evaluated ceiling and floor effects. A total of 216 college students participated the survey. The results demonstrated a good consistency between the scores obtained from ChatGPT and the validated questionnaires, with ICC of 0.81 (95% CI 0.75–0.85, $p < 0.001$) for ULS-6 and 0.95 (95% CI 0.94–0.96, $p < 0.001$) for OSSS-CS. The Spearman correlation coefficients were 0.64 ($p < 0.001$) for ULS-6 and 0.89 ($p < 0.001$) for OSSS-CS, indicating a moderate correlation. No ceiling or floor effects were observed. The ChatGPT-created questionnaire demonstrated acceptable consistency with the validated questionnaires. Future studies can further explore the performance of ChatGPT in different populations and domains, as well as how to integrate it with validated questionnaires to enhance the accessibility of assessments.

Keywords ChatGPT, Large language models, Artificial intelligence, Loneliness, Online social support, Questionnaires

In today's society, mental health issues among college students are increasingly becoming a focus of public concern[1]. Against this background, loneliness and online social support are considered two vital dimensions affecting the mental health of college students, and they continue to be the focus of researchers' attention[2–4]. Loneliness is a negative social and emotional experience characterized by an unacceptable or unpleasant sense of lack in interpersonal relationships[5,6]. As one of the most common emotional experiences among humans, loneliness is typically associated with a variety of negative outcomes[7]. Extensive studies have indicated that loneliness significantly increases the risks of stress, distress, anxiety, depression, anger, insomnia, self-harm, and suicidal ideation[8–11]. College students are at an important stage of developing peer relationships, pursuing autonomy and individuality, and are susceptible to loneliness, significantly affecting their mental health[12,13].

[1]College of Nursing, Hubei University of Chinese Medicine, Wuhan 430065, Hubei Province, China. [2]Hubei Shizhen Laboratory, Wuhan 430065, Hubei Province, China. [3]Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan 430061, Hubei Province, China. [4]Hubei Key Laboratory of Theory and Application Research of Liver and Kidney in Traditional Chinese Medicine, Affiliated Hospital of Hubei University of Chinese Medicine, Wuhan 430061, Hubei Province, China. [5]Hubei Institute of Traditional Chinese Medicine, Wuhan 430061, Hubei Province, China. ✉email: fenyang@hbucm.edu.cn; 1239702016@qq.com

Online social support refers to the communication that occurs within the virtual realm, where individuals are understood and respected through the sharing of emotions, information, and materials, thereby attaining a sense of identity and belonging[14]. The rapid proliferation of the internet and social media has increased the online interactions among college students, making online social support an important channel for them to obtain emotional support and communicate information[15,16]. Studies have shown that online social support is associated with lower negative outcomes[17]. Increased online social support can reduce levels of loneliness and improve mental health to some extent[18–20]. Therefore, it is crucial to select appropriate methods to accurately and efficiently assess loneliness and online social support among college students.

Existing assessments often rely on validated questionnaires. The ULS-6 (University of California Los Angeles Loneliness Scale-6) is a unidimensional scale with good reliability and validity, capable of concisely and swiftly measuring loneliness in the Chinese adult population, including college students[21]. The OSSS-CS (Chinese Youth Version of the Online Social Support Scale) is specifically designed for the Chinese adolescent population to assess the social support of individuals in the online environment, and it has demonstrated good reliability and validity[22,23]. The ULS-6 and OSSS-CS, two widely recognized and validated questionnaires, provide us with in-depth insights into college students' loneliness and how they seek and obtain social support in the online environment[24]. However, there are some shortcomings in the use of validated questionnaires in mental health assessment, including low response rates and efficiency, and susceptibility to the subjectivity of respondents[25]. With the advancement of artificial intelligence technology, ChatGPT, as an emerging assessment tool, shows great potential in the field of mental health, which may compensate some of the shortcomings of validated questionnaires and provide more efficient and objective assessments.

ChatGPT is a GPT-based chatbot that is trained to generate human-understanding texts from inputs (prompts). Since its release in November 2022, ChatGPT has rapidly garnered significant attention from the society[26]. Data showed that ChatGPT has attracted more than 1 million users after just 5 days of launch and has gained 100 million monthly active users within just 2 months of launch, making it the fastest-growing consumer-grade app to date[27]. With its powerful language generation and comprehension capabilities, ChatGPT has already been applied to mental health-related tasks[28]. In the field of mental health, ChatGPT can be used to provide mental health education, conduct mental status assessments, and assist in the implementation of psychological interventions[28,29]. ChatGPT can at times predict mental health symptoms and diagnoses accurately[28]. By pre-training ChatGPT with mental health information, it has proven to be accurate to varying degrees in mental health assessments[30,31]. A study found that ChatGPT's ability to assess suicide risk was similar to that of mental health professionals, especially when using the ChatGPT-4 version[32]. In addition, the ChatGPT demonstrated greater accuracy in recognizing suicidal ideation[32]. These studies suggest that the ChatGPT has potential in mental health assessment, especially in providing initial screening and intervention recommendations. However, current research in exploring the evaluation capabilities of ChatGPT mostly involves comparing ChatGPT's output with the assessment results from professionals, with less involvement in comparing with the results of validated questionnaire assessments. Given that most current mental health assessments use validated questionnaires, so it was necessary to compare the output of ChatGPT with the results of a validated questionnaire assessment. Whether ChatGPT can assist conducting mental health assessments, which could improve medical efficiency.

This study combined the core elements and scoring criteria of the ULS-6 and OSSS-CS to pre-train ChatGPT-4, creating a structured interview questionnaire. Participants were invited to complete both the ChatGPT-created questionnaire and the validated questionnaires. By comparing the consistency of scores from ChatGPT and the validated questionnaires, evaluating the validity of ChatGPT as a mental health assessment tool to measure college students' loneliness and online social support. This study has significant implications by applying ChatGPT to the field of mental health assessment, especially for the assessment of loneliness and online social support in the college student population. First, the study provides an innovative approach to mental health assessment, allowing ChatGPT to assist healthcare professionals to conduct assessments more efficiently. Second, due to the accessibility of ChatGPT, it enables more people to break through time and geographic constraints to access real-time mental health assessment services.

Methods
Participants
From June 2024 to August 2024, 220 college students were enrolled in a cross-sectional study at several universities and colleges in Wuhan, Hubei Province, China, using a convenience sampling method. The inclusion criteria were: (1) aged 18–30 years; (2) current students with regular full-time enrollment (specifically including junior college students, undergraduates, master's and doctoral degree students); (3) informed consent and were willing to participate in this study.

Materials and measures
An information sheet was designed by ourselves to gather demographic data from participants, which included queries regarding gender, age, student leaders' experience, high school origin place, father's education, mother's education, family monthly income, and monthly living expenses.

ChatGPT-4 was pre-trained using the items from the validated questionnaires (ULS-6 and OSSS-CS) and tailored to reflect the daily lives of college students. Specifically, the complete items of these questionnaires were input into ChatGPT-4, and the model was instructed to generate new questions that are more relevant to the everyday experiences of college students. This process involved using the original questionnaire items as prompts to guide ChatGPT-4 in generating contextually appropriate questions without additional fine-tuning. The ChatGPT-4-generated questionnaire was systematically adapted from the original items to align with collegiate cultural and linguistic contexts. To ensure contextual appropriateness, the generated questions

were reviewed by three researchers specializing in psychological assessment, who confirmed that the language, examples, and relevance aligned with the daily experiences of university students (e.g., academic workload, social media interactions, and campus life). Although no formal quantitative evaluation was conducted, this expert review process aimed to enhance the ecological validity of the AI-generated questions. The output of ChatGPT-4 is adjusted by computer experts to ensure effective interaction with validated questionnaires. The ChatGPT-4 constructed four scenarios to assess college students' loneliness and online social support, including campus activities, online recognition, online interactions, and online help-seeking (see Supplementary Information). These scenarios included most of the items in the ULS-6 and OSSS-CS. It is worth noting that although the ChatGPT-4 questionnaire was generated based on two validated questionnaires, the ChatGPT-4 differed from the two previous questionnaires in terms of question content. These differences are mainly in the form of changes in wording and presentation, and the ChatGPT-4 questionnaire is tailored to the cultural background of college students, which is conducive to enhancing their understanding and participation. In order to better accommodate the Chinese-speaking college students, the items were developed, answered, and evaluated in Chinese. After completing these scenarios on the ChatGPT-4, the responses were automatically scored by the software based on the criteria of the ULS-6 and OSSS-CS. This automated scoring process ensured that the generated responses were evaluated for both accuracy and relevance.

The ULS-6 is a valid instrument for measuring loneliness in the Chinese adult population, with good reliability and applicability, and can be assessed concisely and quickly. The ULS-6 scale was validated in Chinese adults with a Cronbach's α coefficient of 0.891[33]. The questionnaire is composed of 6 items and subjects rated the frequency of the 6 symptoms over the past 3 weeks on a 4-point Likert scale (1 = never, 2 = sometimes, 3 = rarely, and 4 = often). The total scores of the ULS-6 ranged from 6 to 24, with higher scores indicating that the individual is more lonely.

The OSSS-CS, a scale developed specifically for Chinese adolescents to assess the level of social support they receive online, has been validated in the Chinese adolescent population. The study showed that the OSSS-CS and each dimension had good internal consistency with a Cronbach's α coefficient of 0.969 and Cronbach's α coefficients of 0.898 to 0.935 for each dimension[22]. The OSSS-CS scale is composed of four dimension of esteem/emotional support (EE), social companionship (SC), informational support (INF), and instrumental support (INS), each of which consists of five items, and subjects are asked to rate each entry on a 5-point Likert scale based on their experiences within the past 2 months, reflecting the frequency of social support they have received from interacting with others online (1 = never, 2 = sometimes, 3 = rarely, 4 = often, and 5 = always). The total scores of the scale ranged from 20 to 100, with higher scores indicating that individuals receive more social support online.

Data collection

Firstly, college students were asked to complete the assessment for the four scenarios adjusted by ChatGPT-4, ChatGPT-4 will output a score for each item as the ChatGPT-created questionnaire scores. Additionally, college students were asked to complete the validated ULS-6 and OSSS-CS questionnaires to produce the validated questionnaire scores. All data were collected anonymously, and the investigators of the questionnaire were not aware of the participants' responses. Participants had the right to withdraw from the study at any time without suffering any adverse consequences.

Statistical analysis

The collected data were statistically analyzed using SPSS 27.0 and Origin 2024. A p value of < 0.05 was considered statistically significant. In this study, Cronbach's α coefficient was calculated to assess the internal consistency of the ChatGPT-created questionnaire and the validated questionnaires.

Spearman correlation analysis was used to explore the correlation between the ChatGPT-created questionnaire and the validated questionnaires. Correlation (r) levels were defined as: high (r ≥ 0.6), moderate (0.6 < r ≤ 0.3), low (r < 0.3)[34]. We hypothesized that the scores from ChatGPT-4-generated questionnaires would exhibit strong positive correlations (r > 0.6) with the validated ULS-6 and OSSS-CS. In this study, Spearman's correlation was chosen because it is a non-parametric measure that assesses the monotonic relationship between two variables without assuming a normal distribution, making it suitable for non-normally distributed continuous data[35].

Intra-class correlation coefficients (ICC) were computed for each item and the total score to assess agreement between the ChatGPT-created questionnaire and the validated questionnaires. ICC was selected because it is a widely used measure for assessing the reliability and consistency of continuous data, providing valuable insights into the agreement between measurements even when the data do not strictly follow a normal distribution[36]. The analysis employed a one-way random effects model with absolute agreement based on single measurements. The Intra class correlation estimates and their 95% Confidence Interval (CI) could range from 0 to 1, where 0 signifies no agreement and 1 denotes perfect agreement. Concerning the interpretation of the ICC, those < 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and > 0.9 are indicative of poor, moderate, good and excellent reliability, respectively[37].

Consistency was quantitatively assessed using Bland–Altman plots. By comparing the differences in scores of the ChatGPT-created questionnaire and the validated questionnaires ULS-6 and OSSS-CS, the bias and 95% consistency boundaries can be visualized. On these scatter plots, the y-axis represents the difference between scores from the ChatGPT-created questionnaire and the validated questionnaires, with a positive difference indicating that the ChatGPT-created questionnaire scored higher than the validated questionnaires. The x-axis represents the mean of the scores of the two questionnaire. According to the Bland–Altman analysis, at least 95% of the difference values should be within the agreement limit, and compared with the professionally acceptable limit value. If the agreement limit is clinically acceptable, it can be considered that the agreement between the two raters is good and can be exchanged.

The percentage of dimension/total scores that reach the highest or lowest score is calculated to evaluate the ceiling or floor effect. If the ratio exceeds 15%, the dimension, domain and total score are considered to have a ceiling or floor effect[35].

Ethical statement

This research protocol was conducted in accordance with the Declaration of Helsinki and has received ethical approval from the Nursing College at Hubei University of Chinese Medicine, and all participants clearly knew the purpose of the study and have signed an informed consent form before participating in the study. In our study, we took stringent measures to ensure that the samples' information could not be reached by the third parties or individuals not involved in this research project. In addition, a series of measures were implemented to protect the samples' information in our study. Firstly, we anonymized all data collected from participants to remove any personally identifiable information. Secondly, only authorized personnel involved in the research project had access to the data. Finally, we adhered to relevant data protection regulations to ensure that all data handling practices were compliant.

Results

A total of 220 college students were recruited, and excluding unanswered questionnaires, 216 valid questionnaires were returned, with an effective recovery rate of 98.18%. All 216 college students were able to complete the administered assessments without assistance. The age distributed between 18 and 27 years old, the mean age was 21.09 years (SD = 1.28), and 149 (69.00%) of the participants were female (Table 1). The median results of the questionnaires, with interquartile ranges and total scores, were as follows: GPT-ULS-6: 8 (IQR 6–11 [range 6–19]); ULS-6: 11.50 (IQR 8–13 [range 6–20]); GPT-OSSS-CS: 67 (IQR 56–74 [range 27–100]); OSSS-CS: 67.50 (IQR 58–78 [range 33–100]).

Spearman correlation analysis between the ChatGPT questionnaire and the validated loneliness and online social support questionnaires revealed both positive intra and inter-correlations. The analysis showed a high correlation between the total scores of the ChatGPT questionnaire and the validated questionnaires: ULS-6 (r = 0.64, $p < 0.001$) and OSSS-CS (r = 0.89, $p < 0.001$), both of which were consistent with our hypothesis. Also most of the other items on the ChatGPT questionnaire were significantly correlated with the validated questionnaires. Moreover, the correlation for OSSS-CS was observed to be stronger than that for ULS-6 (Fig. 1).

The Cronbach's α for the GPT-ULS-6 scale was 0.87, indicating a good reliability. For the GPT-OSSS-CS, the overall Cronbach's α was 0.96, also demonstrating a good internal consistency.

The ICC estimates and the corresponding 95% CI can be found in Table 2. The ICC for the ULS-6 was 0.81 (95% CI 0.75–0.85; $p < 0.001$), indicating slightly lower consistency compared to the OSSS-CS, which had an ICC of 0.95 (95% CI 0.94–0.96; $p < 0.001$). These findings suggest good agreement between the total scores of the ChatGPT questionnaire and those of the validated questionnaires.

Bland–Altman plots showed that the ULS-6 (limits of agreement, 95% CI − 7.27 to 3.23) and OSSS-CS (limits of agreement, 95% CI − 12.95 to 10.08) demonstrated adequate and acceptable concordance at the 95% level, suggesting that these scales could be used interchangeably in clinical settings or research. A slight bias was observed between the GPT-ULS-6 and ULS-6, with a mean difference of − 2.02 (SD: 2.68), and between the GPT-OSSS-CS and OSSS-CS, with a mean difference of − 1.44 (SD: 5.88) (Fig. 2).

ULS-6 had the lowest score of 13.40% and the highest score of 0.40%; OSSS-CS had the lowest score of 0.40% and the highest score of 1.30%; ChatGPT-ULS-6 had the lowest score of 11.50% and the highest score of 0.40%; and ChatGPT-OSSS-CS had the lowest score of 0.40% and the highest score of 0.40%. There is almost no floor/ceiling effect for the four.

	Number	Percent	Groups	Number	Percent
Gender			Age (years)		
Male	67	31.00	18–21	137	63.40
Female	149	69.00	22–30	79	26.60
High school origin place			Student leaders' experience		
Cities	63	29.20	Yes	122	56.50
Towns	63	29.20	No	94	43.50
Rural areas	90	41.70			
Father's education			Mother's education		
Elementary schools	44	20.40	Elementary schools	56	25.90
Middle schools	81	37.50	Middle schools	89	41.20
High schools/Junior colleges	64	29.60	High schools/Junior colleges	50	23.10
Colleges and above	27	12.50	Colleges and above	21	9.70
Family monthly income			Monthly living expenses		
≤ 5000	89	41.20	≤ 1500	121	56.00
> 5000	127	58.80	> 1500	95	44.00

Table 1. Socio-demographic characteristics of college students (n = 216).

(a)

(b)
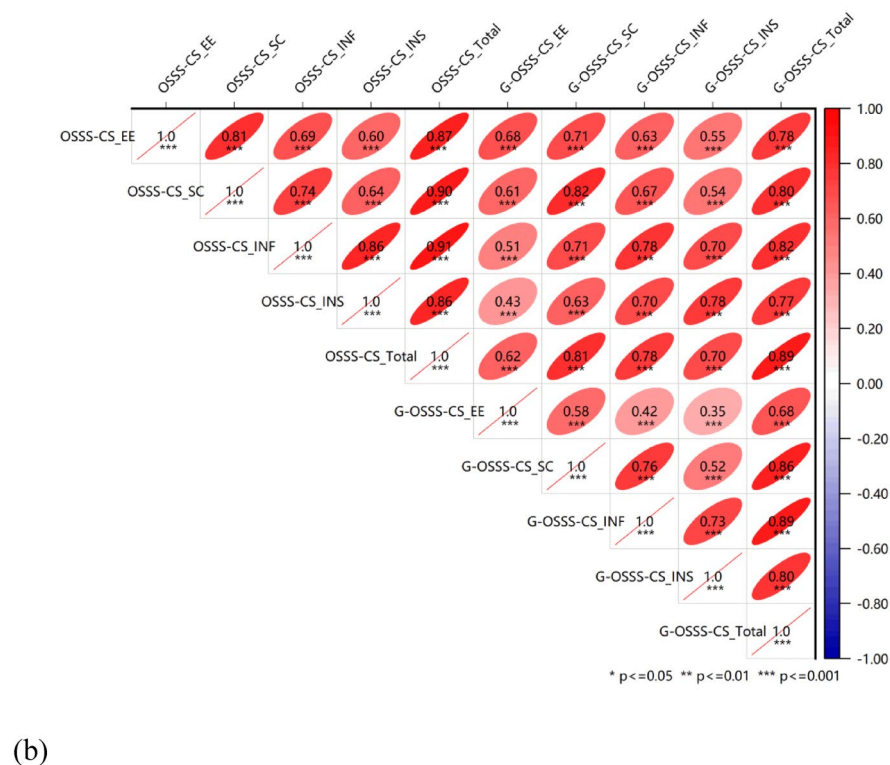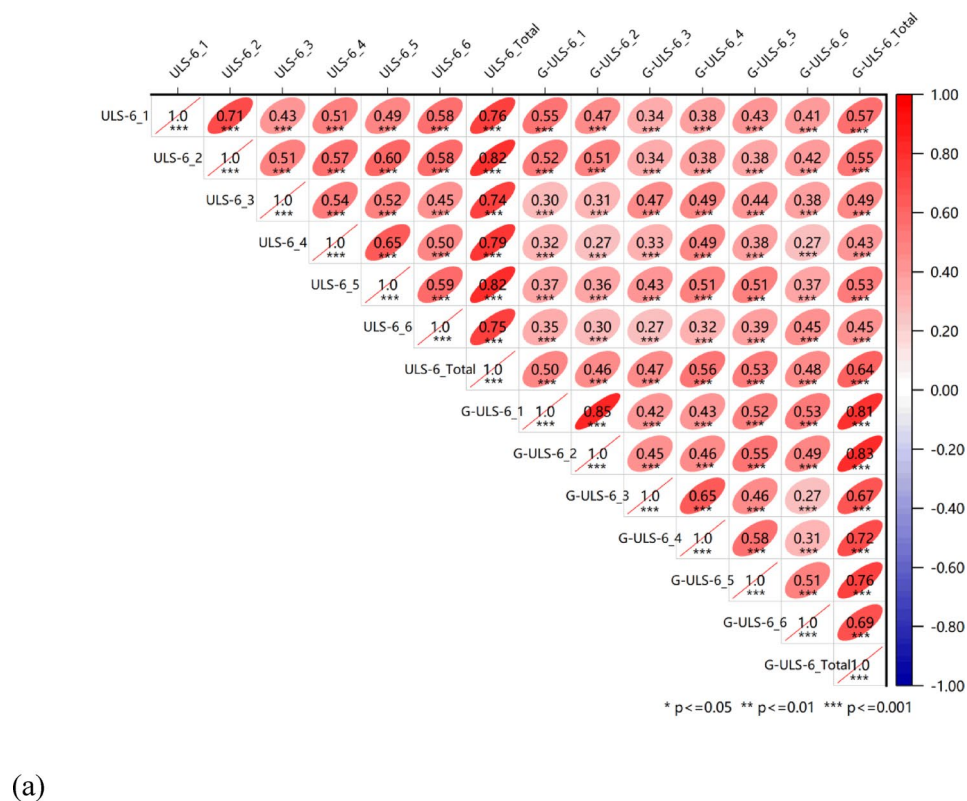
Fig. 1. Spearman correlation matrix of the ChatGPT questionnaire and the validated loneliness and online social support questionnaires scores: the heatmap visualizes the Spearman correlation matrix for the ChatGPT questionnaire and the validated questionnaires scores. Heatmap (**a**) compares the Spearman correlation coefficients and the degree of correlation between GPT-ULS-6 and ULS-6; Heatmap (**b**) compares the Spearman correlation coefficients and the degree of correlation between GPT-OSSS-CS and OSSS-CS. Each cell in the heatmap represents the Spearman correlation coefficient between two variables, with the color intensity and the value indicating the strength and direction of the correlation. Asterisks indicates significant values: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

	ICC (95%)	F-test value	p	Interpretation
ULS-6	0.81 (0.75–0.85)	F (215, 215) = 5.1	< 0.001	Good
OSSS-CS	0.95 (0.94–0.96)	F (215, 215) = 20.8	< 0.001	Excellent

Table 2. ICC to evaluate the consistency of the total scores between the ChatGPT-created questionnaire and the validated questionnaires.

Discussion

This study compared scoring consistency between ChatGPT and the validated questionnaires in assessing college students' loneliness and online social support. The results showed that ChatGPT may be effective in assessing loneliness and online social support among college students, and may also be applicable to the assessment of a broader range of mental health dimensions.

In assessing loneliness and online social support among college students, we found acceptable significant consistency between the results of ChatGPT and the validated questionnaires, which is consistent with previous research[38]. This consistency may be related to two factors. First, the ChatGPT adapted the questionnaire to the purpose of the study and the characteristics of the college student population, ensuring that the questionnaire was both relevant to loneliness and online social support items and appropriated for college students, with a high reliability. Second, college students typically have more open and positive attitudes toward emerging technologies[39]. The study found that perceived usefulness plays a central role in shaping students' trusting attitudes toward ChatGPT and is critical in predicting and encouraging successful technology adoption[40]. Higher acceptance of the new technology ChatGPT among college students may lead them to hold a positive psychological prediction of ChatGPT in favor of new technology adoption[40,41]. College students may be more willing to accept and trust an artificial intelligence tool like ChatGPT. This trust may motivate them to be more open in their interactions with ChatGPT, thus allowing ChatGPT to more accurately assess their level of loneliness and online social support. The results suggest that ChatGPT-created questionnaire, specifically designed to assess loneliness and online social support among college students, shows promise in generating relevant items for mental health evaluations, though further validation is needed to establish its independent validity.

The results of the questionnaire scores showed lower assessment scores on the ChatGPT compared to the validated questionnaire, which may indicated that the ChatGPT is more inclined to reflect lower levels of loneliness when measuring loneliness and online social support among college students. This difference may stem from ChatGPT's unique assessment methodology or from the positive experiences college students have when interacting with ChatGPT[42,43]. First, ChatGPT's interactive design and natural language processing capabilities allow it to communicate with users in a way that is closer to everyday conversation[44]. This interactivity may make respondents feel more comfortable and at ease, and thus more willing to share their true feelings[45]. One study noted that the ChatGPT showed positive results in increasing students' behavioral engagement, which may indicate that the ChatGPT is more effective in promoting student engagement and interaction when assessing college students' loneliness and online social support[45]. Second, the ChatGPT's assessment may be more focused on respondents' immediate responses, which may result in the ChatGPT being able to capture more subtle emotional changes when assessing loneliness and online social support[46]. This real-time feedback mechanism may allow the ChatGPT to more accurately reflect respondents' psychological states than traditional questionnaires in some cases[47]. Overall, the ChatGPT's lower scores in assessing college students' feelings of loneliness and online social support may not be due to the fact that it underestimates these feelings, but rather to the fact that it provides a more interactive and personalized assessment, which may be more in line with the communication habits and psychological needs of the college student population. This difference emphasizes the importance of considering the characteristics of the target group when choosing an assessment tool.

The consistency in scores between ChatGPT and validated questionnaires when assessing loneliness and online social support among college students suggests that ChatGPT may offer a level of reliability in the realm of mental health assessment. This consistency suggests that the ChatGPT may be useful as a complementary tool, especially in capturing subtle changes in an individual's psychological state[48]. Its real-time feedback and adaptive nature may provide a nuanced assessment perspective in specific contexts, offering a possible resource for clinical practice[49]. At the same time, the convenience and accessibility of ChatGPT may encourage more individuals to seek help, opening up new possibilities for mental health service delivery[47]. These findings provide a new perspective to explore the use of ChatGPT in mental health.

Additionally, the high correlation between the ChatGPT-created questionnaire and the validated questionnaires may be influenced by the pre-training process, which aligned the AI-generated content with the constructs measured by existing questionnaires. While this ensures the relevance of the generated questions, it also introduces a potential tautological relationship. Future research should explore additional validation methods to assess the unique contributions of AI-generated questionnaires, ensuring that these tools provide meaningful insights and enhance the measurement of psychological constructs.

Limitations

There are potential limitations of this study. First, due to time and effort limits, this study only surveyed college students in Wuhan, Hubei Province, China, resulting in a small sample size and potentially insufficient representation, which may affect the generalizability of the results. Future studies should expand the sample
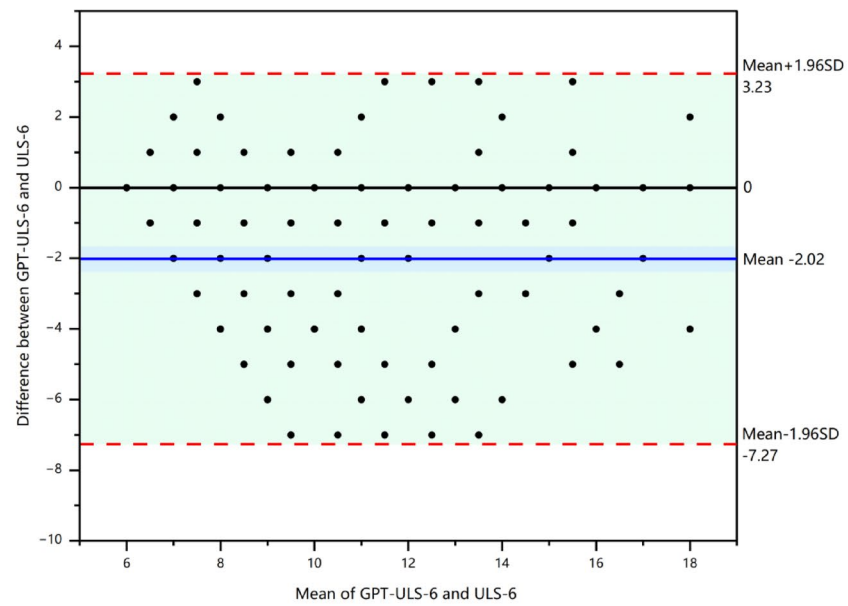
(a)
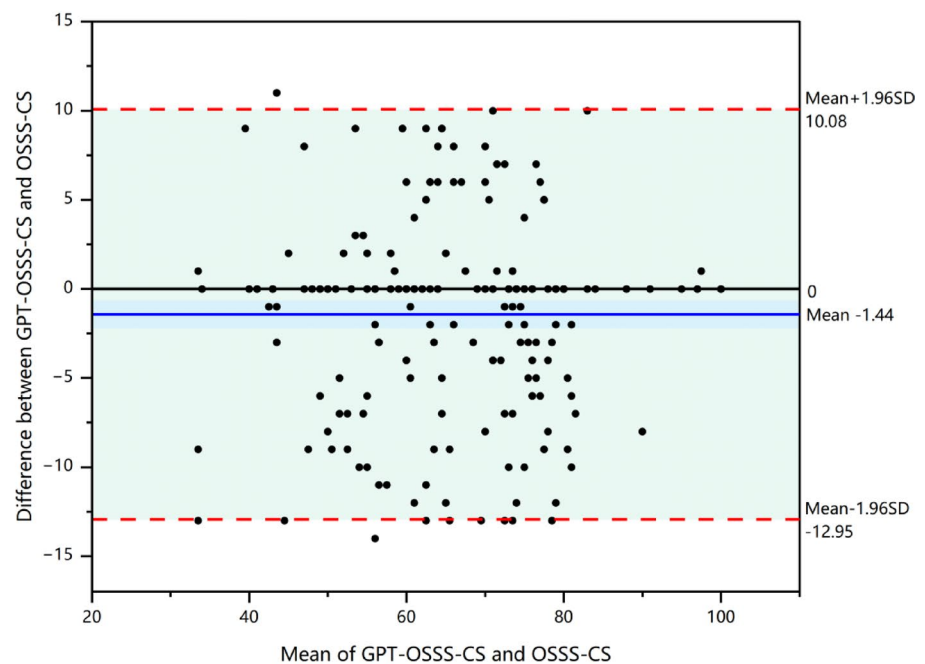
(b)

Fig. 2. (**a**) Bland–Altman plot of difference between GPT-ULS-6 and ULS-6 change total scores (n = 216). (**b**) Bland–Altman plot of difference between GPT-OSSS-CS and OSSS-CS change total scores (n = 216).

range, considering various demographic characteristics, to verify the validity and applicability of the ChatGPT multidimensional assessment in a broader social context.

Second, current ChatGPT-4 access is limited by a paywall, and access to the professional version of Bing BETA requires a $20 per month subscription fee, which may increase the difficulty of promoting in clinical assessments in China. Third, this study did not ask the respondents about the feasibility between the ChatGPT-4-created and the validated questionnaires, nor did it ask which one is better and more practical for them. Future studies should fully consider the respondents' feedback. Fourth, while we claimed that the ChatGPT-4-generated questions were contextually appropriate for college students, this assessment relied on expert judgment rather than standardized quantitative measures. Future research should employ systematic approaches (e.g., cognitive interviews or item relevance ratings) to objectively evaluate contextual appropriateness. Additionally, because our study did not include subgroups of college students with varying levels of loneliness and online social support, it was not possible to measure known-group validity; nor did we have the opportunity to assess responsiveness due to the cross-sectional nature of this study. Future research should emphasize the importance of known cluster effects and reactivity in psychometric assessments.

Conclusions

The results of this cross-sectional study indicated substantial agreement between the questionnaire created by ChatGPT and the validated questionnaires assessing loneliness and online social support among college students. This finding not only heralds the great potential of ChatGPT technology in mental health assessment, but also provides new ideas for our future research directions. In the future, we expect ChatGPT to be further integrated into more clinical assessments, improve the efficiency and accuracy of assessments through intelligent questionnaire design, and provide more comprehensive and in-depth data support for mental health research and clinical practice. At the same time, we should also think deeply about how to make better use of ChatGPT technology to meet the mental health assessment needs of different populations and situations, and to promote the personalized and precise development of mental health assessment.

Data availability

The datasets generated and/or analyzed in the study are not currently publicly available but are available from the corresponding authors of this study upon reasonable request.

References

1. Lipson, S. K. et al. Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. *J. Affect. Disord.* **306**, 138–147. https://doi.org/10.1016/j.jad.2022.03.038 (2022).
2. Lee, D. S., Stahl, J. L. & Bayer, J. B. Social resources as cognitive structures: Thinking about a dense support network increases perceived support. *Soc. Psychol. Q.* **83**, 405–422. https://doi.org/10.1177/0190272520939506 (2020).
3. Expósito, F. & Moya, M. Soledad y apoyo social. *Rev. Psicol. Soc.* https://doi.org/10.1174/021347499760260000 (1999).
4. Perlman, D. & Peplau, L. A. Toward a social psychology of loneliness. *Pers. Relat.* **3**, 31–56 (1981).
5. Cunningham, K. B., Kroll, T. & Wells, M. First steps in identifying and addressing loneliness in the context of COVID-19. *Perspect. Public Health* **141**, 200–201. https://doi.org/10.1177/1757913920975793 (2021).
6. Sipowicz, K., Podlecka, M., Mokros, Ł & Pietras, T. Lonely in the city-sociodemographic status and somatic morbidities as predictors of loneliness and depression among seniors-preliminary results. *Int. J. Environ. Res. Public Health* **18**, 7213. https://doi.org/10.3390/ijerph18147213 (2021).
7. Fan, Z., Shi, X., Yang, S., Sun, Y. & Chen, R. Reliability and validity evaluation of the stigma of loneliness scale in Chinese college students. *BMC Public Health* **24**, 238. https://doi.org/10.1186/s12889-024-17738-0 (2024).
8. Espinoza, G. & Hernandez, H. L. Adolescent loneliness, stress and depressive symptoms during the COVID-19 pandemic: The protective role of friends. *Infant Child Dev.* **31**, e2305. https://doi.org/10.1002/icd.2305 (2022).
9. Lewis, K. C., Roche, M. J., Brown, F. & Tillman, J. G. Attachment, loneliness, and social connection as prospective predictors of suicidal ideation during the COVID-19 pandemic: A relational diathesis-stress experience sampling study. *Suicide Life Threat. Behav.* **53**, 64–74. https://doi.org/10.1111/sltb.12922 (2023).
10. Rodriguez, M., Osborn, T. L., Gan, J. Y., Weisz, J. R. & Bellet, B. W. Loneliness in Kenyan adolescents: Socio-cultural factors and network association with depression and anxiety symptoms. *Transcult. Psychiatry* **59**, 797–809. https://doi.org/10.1177/13634615221099143 (2022).
11. Rönkä, A. R., Taanila, A., Koiranen, M., Sunnari, V. & Rautio, A. Associations of deliberate self-harm with loneliness, self-rated health and life satisfaction in adolescence: Northern Finland Birth Cohort 1986 Study. *Int. J. Circumpolar Health* **72**, 21085. https://doi.org/10.3402/ijch.v72i0.21085 (2013).
12. Wang, G. & Hu, W. Peer relationships and college students' cooperative tendencies: Roles of interpersonal trust and social value orientation. *Front. Psychol.* **12**, 656412. https://doi.org/10.3389/fpsyg.2021.656412 (2021).
13. Conti, C., Lanzara, R., Rosa, I., Müller, M. M. & Porcelli, P. Psychological correlates of perceived loneliness in college students before and during the COVID-19 stay-at-home period: A longitudinal study. *BMC Psychol.* **11**, 60. https://doi.org/10.1186/s40359-023-01099-1 (2023).
14. Liang, X. & Wei, L. A preliminary study on the evaluation of online social support for college students. *Psychol. Sci.* **3**, 671–689 (2008).
15. Sun, D., Lin, Y., Liao, C. & Pan, L. Online interpersonal trust and online altruistic behavior in college students: The chain mediating role of moral identity and online social support. *Front. Psychol.* **15**, 1452066. https://doi.org/10.3389/fpsyg.2024.1452066 (2024).
16. Jiang, Y., Chi, J., Wang, L. & Geng, X. Online communication and positive psychological capital of college students in China: The mediating role of online social support. *BMC Psychol.* **11**, 277. https://doi.org/10.1186/s40359-023-01324-x (2023).
17. Zhou, Z. & Cheng, Q. Relationship between online social support and adolescents' mental health: A systematic review and meta-analysis. *J. Adolesc.* **94**, 281–292. https://doi.org/10.1002/jad.12031 (2022).
18. Barnum, D. D., Snyder, C. R., Rapoff, M. A., Mani, M. M. & Thompson, R. Hope and social support in psychological adjustment of children who have survived burn injuries and their matched controls. *Child. Health Care* **27**, 15–30. https://doi.org/10.1207/s15326888chc2701_2 (1998).

19. Liu, G., Li, S. & Kong, F. Association between social support, smartphone usage and loneliness among the migrant elderly following children in Jinan, China: A cross-sectional study. *BMJ Open* **12**, e060510. https://doi.org/10.1136/bmjopen-2021-060510 (2022).

20. Luo, Q., Huang, L. & Wu, N. The relationship between internet use preference and loneliness among college students during COVID-19: The chain mediating effect of online social support and self-esteem. *Front. Psychol.* **13**, 1058944. https://doi.org/10.3389/fpsyg.2022.1058944 (2022).

21. Xiao, R. & Du, J. Reliability and validity of the 6-item UCLA Loneliness Scale (ULS-6) for application in adults. *Nan fang yi ke da xue xue bao J. South. Med. Univ.* **43**, 900–905. https://doi.org/10.12122/j.issn.1673-4254.2023.06.04 (2023).

22. Zhou, Z. & Cheng, Q. Measuring online social support: Development and validation of a short form for Chinese adolescents. *Int. J. Environ. Res. Public Health* **19**, 14058. https://doi.org/10.3390/ijerph192114058 (2022).

23. Qiu, Y. F., Yang, F., Li, C. Y., Liu, J. L. & Zeng, L. J. Reliability and validation of Chinese version of online social support scale for older adults. *Yixue Xinzhi Zazhi* **34**, 33–42. https://doi.org/10.12173/j.issn.1004-5511.202309068 (2024) (**in Chinese**).

24. Fang, J. Cybervictimization and loneliness among Chinese college students: A moderated mediation model of rumination and online social support. *Child. Youth Serv. Rev.* **115**, 105085 (2020).

25. Shields, R. E. et al. Brief mental health disorder screening questionnaires and use with public safety personnel: A review. *Int. J. Environ. Res. Public Health* **18**, 3743. https://doi.org/10.3390/ijerph18073743 (2021).

26. Eppler, M. et al. Awareness and Use of ChatGPT and large language models: A prospective cross-sectional global survey in urology. *Eur. Urol.* **85**, 146–153. https://doi.org/10.1016/j.eururo.2023.10.014 (2024).

27. Cheng, K. et al. Potential use of artificial intelligence in infectious disease: Take ChatGPT as an example. *Ann. Biomed. Eng.* **51**, 1130–1135. https://doi.org/10.1007/s10439-023-03203-3 (2023).

28. Lawrence, H. R. et al. The opportunities and risks of large language models in mental health. *JMIR Ment. Health* **11**, e59479. https://doi.org/10.2196/59479 (2024).

29. Ayers, J. W. et al. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* **183**, 589–596. https://doi.org/10.1001/jamainternmed.2023.1838 (2023).

30. Lamichhane, B. *Evaluation of ChatGPT for NLP-Based Mental Health Applications. arXiv,* Preprint posted online on https://doi.org/10.48550/arXiv.2303.15727 (2023).

31. Amin, M. M., Cambria, E. & Schuller, B. W. *Will Affective Computing Emerge from Foundation models and General AI? A First Evaluation on ChatGPT.* arXiv, Preprint posted online on https://doi.org/10.48550/arXiv.2303.03186 (2023).

32. Levkovich, I. & Elyoseph, Z. Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: Vignette study. *JMIR Ment. Health* **10**, e51232 (2023).

33. Xiao, R. & Du, J. W. Reliability and validity of the 6-item UCLA loneliness scale(ULS-6)for application in adults. *J. South. Med. Univ.* **43**, 900–905. https://doi.org/10.12122/j.issn.1673-4254.2023.06.04 (2023).

34. Andresen, E. M. Criteria for assessing the tools of disability outcomes research. *Arch. Phys. Med. Rehabil.* **81**, S15-20. https://doi.org/10.1053/apmr.2000.20619 (2000).

35. McHorney, C. A. & Tarlov, A. R. Individual-patient monitoring in clinical practice: Are available health status surveys adequate?. *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.* **4**, 293–307. https://doi.org/10.1007/bf01593882 (1995).

36. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropract. Med.* **15**, 155–163 (2016).

37. Koo, T. K. & Li, M. Y. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012 (2016).

38. Coraci, D. et al. ChatGPT in the development of medical questionnaires. The example of the low back pain. *Eur. J. Transl. Myol.* **33**, 12114. https://doi.org/10.4081/ejtm.2023.12114 (2023).

39. Novikova, I. A., Bychkova, P. A. & Novikov, A. L. Attitudes towards digital educational technologies among Russian university students before and during the COVID-19 pandemic. *Sustainability* **14**, 6203 (2022).

40. Abdaljaleel, M. et al. A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Sci. Rep.* **14**, 1983. https://doi.org/10.1038/s41598-024-52549-8 (2024).

41. Ajlouni, A. O., Wahba, F.A.-A. & Almahaireh, A. S. Students' attitudes towards using ChatGPT as a learning tool: The case of the University of Jordan. *Int. J. Interact. Mob. Technol.* **17**, 99 (2023).

42. Suriano, R., Plebe, A., Acciai, A. & Fabio, R. A. Student interaction with ChatGPT can promote complex critical thinking skills. *Learn. Instr.* **95**, 102011. https://doi.org/10.1016/j.learninstruc.2024.102011 (2025).

43. Acosta-Enriquez, B. G., Arbulú Pérez Vargas, C. G., Huamaní Jordan, O., Arbulú Ballesteros, M. A. & Paredes Morales, A. E. Exploring attitudes toward ChatGPT among college students: An empirical analysis of cognitive, affective, and behavioral components using path analysis. *Comput. Educ. Artif. Intell.* **7**, 100320. https://doi.org/10.1016/j.caeai.2024.100320 (2024).

44. Jagnade, G., Sable, S. & Ikar, M. In *Control and Information Sciences* (eds Thirunavukkarasu, I. & Kumar, R.) 271–300 (Springer, 2023).

45. Lo, C. K., Hew, K. F. & Jong, M.S.-Y. The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Comput. Educ.* **219**, 105100 (2024).

46. Schaaff, K., Reinig, C. & Schlippe, T. Exploring ChatGPT's empathic abilities. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8 (2023).

47. Siddals, S., Torous, J. & Coxon, A. "It happened to be the perfect thing": experiences of generative AI chatbots for mental health. *Ment. Health Res.* **3**, 48 (2024).

48. Alanezi, F. Assessing the effectiveness of ChatGPT in delivering mental health support: A qualitative study. *J. Multidiscip. Healthc.* **17**, 461–471. https://doi.org/10.2147/jmdh.S447368 (2024).

49. Liu, J., Wang, C. & Liu, S. Utility of ChatGPT in clinical practice. *J. Med. Internet Res.* **25**, e48568. https://doi.org/10.2196/48568 (2023).

Author contributions

FY: Conceptualization, methodology, project administration, resource, supervision, validation. XH Z: Conceptualization, methodology, project administration, resource, supervision, validation. JG: Data curation, investigation, software, visualization, writing—original draft preparation. JL L: Investigation, methodology, supervision. LJ Z: Methodology, supervision. YQ Y: Methodology, supervision. YF Q: Methodology, supervision. YK Y: Investigation. MJ T: Investigation.

Funding

Declarations

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/s41598-025-06358-2.

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

Reprints and permissions information is available at www.nature.com/reprints.

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