Proceedings of the 7th International Conference on Education Innovation and Philosophical Insights
DOI: 10.54254/2753-7048/2026.PE38243

# *Analysis of Mental Health Assessment of Junior High School Students Based on AIGC Human-Computer Interaction Model*

**Dingchen Zhu**

*Kang Chiao International School, Kunshan, China*
*kriszhu1215@163.com*

***Abstract.*** Adolescent mental health challenges are worsening. Traditional paper- or web-based screening tools lack generalizability, timeliness, and objectivity—and are especially vulnerable to social desirability bias and stigma, causing students to conceal distress. AI-generated content (AIGC) offers a promising alternative for mental health assessment. Yet, the validity and acceptability of AIGC-based conversational assessments in real school settings remain untested. This study involved 584 seventh- and eighth-grade students from Chunjiang Middle School (Hangzhou, Zhejiang) to: (1) assess agreement between the AIGC conversational model and the school's existing observation list; (2) test its ability to identify at-risk students missed by routine observation; and (3) evaluate student acceptance. Results show the AIGC model aligns well with current practices, detects overlooked cases, and is well accepted—making it a viable supplement to traditional screening and teacher observation. It supports a shift from passive intervention to active prevention in school mental health. This study provides empirical evidence and practical guidance for scaling AIGC-based screening in K–12 settings and advances efforts to build a destigmatized adolescent mental health service system.

***Keywords:*** Adolescents, AIGC, human-computer dialogue, mental health screening

## 1. Introduction

Adolescent mental health is a growing global public health concern, with the WTO estimating that one in seven 10–19-year-olds experiences mental health problems—many undetected and untreated due to stigma and system gaps [1, 2]. The Healthy China 2030 Planning Outline prioritizes strengthening mental health services, improving population-level literacy, and expanding early detection for high-risk groups. Current school-based assessments rely on self-report scales and teacher observations—methods limited by poor accessibility, delay, subjectivity, and privacy risks [3]. Adolescents often underreport due to stigma or cognitive immaturity, social desirability bias further distorts responses [4]. Meanwhile, shortages of trained professionals hinder scalable, routine screening—so much so that high-pressure cohorts are frequently excluded, creating a "neglect-of-the-most-vulnerable" paradox [5]. Artificial intelligence (AI), particularly advances in natural language processing and multimodal analysis, enables unobtrusive, continuous, and dynamic mental

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.PE38243

state assessment through language, voice, and behavioral patterns [6, 7]. This "accompanying assessment" shifts focus from reactive intervention to proactive prevention [8]. Yet real-world validation in schools remains scarce: evidence on validity, implementation conditions, and scalability of AIGC-driven conversational assessment is still lacking [9].

This pilot study at Chunjiang Middle School (Hangzhou) evaluates two core questions: (1) How consistently does the AIGC conversational assessment align with the school's existing at-risk student list ? (2) Does it identify at-risk students missed by current observation practices ? Findings provide empirical grounding for deploying AIGC tools in school and inform policy design for intelligent mental health services.

## 2. Methods

### 2.1. Research design

This study adopted a mixed-method design. The quantitative analysis compared the AI system's risk identification results with the school's existing observation list and the MHT scale to examine screening consistency and the AI system's additional identification capability. The qualitative part collected contextualized information through homeroom teacher interviews and student feedback questionnaires to help explain the mechanisms behind the quantitative results. The school's "secondary observation list," routinely formed in the school's daily mental health management, was used as a school-based real-world reference rather than a clinical diagnostic gold standard. The list was used to evaluate the consistency between AI screening results and the school's already-monitored students and to examine the AI system's additional identification capability among students not included on the school's observation list.

### 2.2. Research participants

This study was conducted at Chunjiang Middle School in Hangzhou, and all participating students participated under the school's unified research arrangement with parental informed consent. Participants comprised all students from Classes 1–8 of Grade 7 and Classes 1–8 of Grade 8, totaling approximately 600 students. A final valid sample of 584 students was obtained, including about 305 males and 279 females, aged 12–15 years (mean 13.4±0.7 years). The school's original secondary observation list included 57 students. This list was compiled by the school through routine mental health management processes, combining homeroom teacher observations, psychological teacher interviews, and other information, representing the school's internal high-attention group rather than a clinically diagnosed group.

### 2.3. Measurement tools

The Mental Health Test (MHT) Scale scale was revised and developed by the Department of Psychology at East China Normal University, suitable for group screening of mental health status among primary and secondary school students in China [10]. The scale contains 100 items covering eight content subscales (learning anxiety, social anxiety, loneliness tendency, self-blame tendency, allergic tendency, physical symptoms, fear tendency, impulsive tendency) and one validity subscale (lie scale). The validity items are Nos. 82, 84, 86, 88, 90, 92, 94, 96, 98, and 100, if the total score on these items is too high, the questionnaire is invalidated. Criteria: a total score of ≥65 falls into the "professional assessment" range (requiring further clinical interviews), 56–64 falls into the

DOI: 10.54254/2753-7048/2026.PE38243

"suboptimal" range (requiring continued attention), 0 indicates "not administered", others are "normal." In this sample, the Cronbach's α coefficient of the MHT scale was 0.82, indicating acceptable internal consistency.

Jueqing AI system builds a digital human based on AIGC technology to conduct voice-based, semi-structured interactions with students. The digital human's appearance adopts a youth-friendly, neutral style, and its language style matches junior high school students' communication habits. The system's backend analyzes students' linguistic content, paralinguistic features, and interaction depth. Interaction depth is classified according to the actual number of exchanges (See Table 1).

Table 1. Operational criteria for AI-based assessment

| Indicator | Operational definition | Analytical role |
|---|---|---|
| Linguistic features | Keywords, topic choices, emotional word frequency | Identification of potential psychological risk signals |
| Paralinguistic features | Speech rate variation, pause patterns | Identification of potential interactional/emotional signals |
| Interaction depth | Number of exchanges during the interaction | Stratification of assessment reliability |
| <20 exchanges | Low reliability | Considered insufficient interaction for effective assessment |
| 20-35 exchanges | Medium-low reliability | Used for stratified analysis |
| 35-50 exchanges | Medium-high reliability | Used for stratified analysis |
| >50 exchanges | High reliability | Used for stratified analysis |

Each student-machine interaction lasted approximately 15–20 minutes. To ensure interaction effectiveness, the system provided a prompt list of 100 common psychological issues for junior high students, covering academic pressure, interpersonal relationships, emotional distress, family problems, etc. Students could refer to this list for topic selection or initiate topics freely. Students logged in using school-assigned accounts (school code + student ID + preset password), with on-site operational guidance provided by Teacher Xu. After the interaction, the school organized students to complete a satisfaction survey questionnaire (voluntary). The MHT scale and AI interaction were administered at different time slots to avoid fatigue effects and cross-contamination. To clarify the relationship among the different data sources, the main measures used in the analysis are summarized in Table 2.

Table 2. Measures and data sources

| Measure | Data source | Main role |
|---|---|---|
| MHT total score | MHT scale | Conventional screening reference |
| MHT classification | MHT scale | Comparison with AI identification |
| AI risk classification | Jueqing AI Assessment System | Primary AI-based identification outcome |
| Interaction depth | AI interaction records | Stratified analysis of identification consistency |
| School observation status | School secondary observation list | Real-world educational reference |
| Students feedback | Post-interaction questionnaire | Descriptive assessment of user experience |

DOI: 10.54254/2753-7048/2026.PE38243

### 2.4. Procedure

The experiment took place in February 2025 in Chunjiang Middle School's computer lab, with classes rotating in sessions of ~40 students. To ensure independent responses, students sat at spaced intervals, wore headphones, and interacted individually with an AIGC digital avatar for 15–20 minutes. A curated list of 100 common adolescent psychological concerns—covering academic pressure, peer relationships, emotional distress, and family issues—was provided as optional conversation prompts; students could select from the list or initiate topics freely. Login used school-issued accounts (school ID + student number + preset password), with on-site support from Teacher Xu Cheng. Immediately after the interaction, students voluntarily completed a school-administered satisfaction survey. The MHT scale and AI assessment were administered at separate time points to prevent fatigue and cross-contamination.

### 2.5. Data analysis methods

Descriptive statistics were used to calculate AI identification rates, consistency rates between AI and the school's secondary observation list, AI's additional identification rates outside the school's list, and student feedback percentages. A cross-classification table was constructed to present the consistency and discrepancies between AI screening results and the school's secondary observation list. Consistency between AI and the school's list was stratified by interaction depth. MHT scale data were used as a reference validity indicator for cross-comparison with AI assessment results.

### 2.6. Ethical considerations and data security

This study involves mental health information of minors, so special attention was paid to informed consent, privacy protection, and risk management during implementation. All participating students participated under the school's unified arrangement and obtained parental informed consent. Research data were anonymized before analysis, and no personally identifiable information is presented in the paper.

## 3. Results

### 3.1. Descriptive statistics & reliability and validity

A total of 584 valid samples were obtained. The school's original secondary observation list had 57 students, after excluding 30 ninth-grade students who did not participate and 5 who did not log in, 23 students on the observation list actually participated. MHT scale results showed that among all valid samples, 42 students (7.2%) scored ≥65, and 78 (13.4%) scored 56–64, with an overall proportion requiring attention of about 20.6%. Interaction depth distribution: <20 exchanges: 86 (14.7%), 20–35: 203 (34.8%), 35–50: 218 (37.3%), >50: 77 (13.2%).

The MHT scale demonstrated acceptable internal consistency in this sample (Cronbach's $\alpha$=0.82), with subscale $\alpha$ coefficients ranging from 0.64 to 0.79, consistent with reliability indices in previous studies. Using the school observation list as a real-world criterion, the concurrent validity of AI assessment significantly improved with increasing interaction depth.

DOI: 10.54254/2753-7048/2026.PE38243

### 3.2. Comparison between AI identification and school observation list

Among the 23 students on the school observation list, stratified analysis showed: (1) Low-reliability range: 8 students, no reliable assessment due to insufficient valid information. (2) Medium-low range: 5 students, AI identified 3 (60% identification rate). (3) Medium-high and high ranges combined: 10 students, AI identified 9 (90% identification rate).

The AI system identified 13 out of the 23 students on the school's secondary observation list who participated, yielding an overall consistency rate of 56.5%. This overall result was influenced by students with low interaction depth. Within the medium-high and high-reliability interaction ranges, there were 10 students on the school's secondary observation list, and the AI identified 9 of them, reaching a consistency rate of 90%. This suggests that the reference value of AI screening results may depend on sufficient human-computer interaction information, insufficient interaction depth may reduce the stability of the model's judgment.

### 3.3. Additional identification capability

Table 3. Examples of students additionally identified by the AI system

| MHT Assessment | MHT Score | AI Mean Score | AI Assessment Frequency | AI Reliability | High-Risk Frequency | Summary of System-Displayed Content |
|---|---|---|---|---|---|---|
| Not Implemented | 0 | 7.67 | 50 | High | 25 | Problems identified in all dimensions |
| Suboptimal | 59 | 7.22 | 49 | Moderate | 27 | Problems identified in all dimensions |
| Suboptimal | 57 | 4.89 | 73 | High | 22 | Learning, Social, Self-Blame, Loneliness |
| Suboptimal | 62 | 6.33 | 250 | High | 66 | Problems identified in all dimensions |
| Professional assessment | 70 | 4.89 | 95 | High | 19 | Learning, Social, Self-Blame, Loneliness |
| Suboptimal | 61 | 7.11 | 37 | Moderate | 15 | Learning, Self-Blame, Depression, Impulsivity |
| Professional assessment | 65 | 8.89 | 19 | Insufficient Data | 15 | Depression, Self-Blame, Loneliness |
| Suboptimal | 59 | 8.83 | 31 | Low | 20 | Social, Loneliness, Impulsivity, Depression |
| Professional assessment | 66 | 7.67 | 61 | High | 29 | Learning, Social, Fear in Specific Situations |
| Suboptimal | 64 | 7.56 | 38 | Low | 29 | Loneliness, Self-Blame, Impulsivity, Learning, Social |
| Professional assessment | 73 | 6.44 | 99 | High | 50 | Depression, Fear in Specific Situations, Physical Symptoms, Self-Blame, Loneliness, Learning, Social |

DOI: 10.54254/2753-7048/2026.PE38243

Table 3. (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| Professional assessment | 65 | 7.44 | 45 | Moderate | 20 | Learning, Social, Self-Blame, Physical Symptoms |
| Suboptimal | 61 | 4.89 | 67 | High | 22 | Learning, Fear in Specific Situations, Physical Symptoms |
| Suboptimal | 56 | 7.11 | 47 | Moderate | 24 | Depression, Self-Blame, Learning, Social, Loneliness |
| Professional assessment | 71 | 7.67 | 22 | Low | 17 | Loneliness, Social |

More critically, beyond the school's secondary attention list, the AI additionally provided a list of 56 students requiring attention, including 15 "priority attention (A) check" students. Partial data are shown in Table 3.

The research team conducted verification interviews with the corresponding homeroom teachers: among the 15 priority-check students, 8 (53.3%) showed identifiable tendencies toward external environmental stress, pressure, or personal characteristics (e.g., recent family changes, interpersonal conflicts, sudden grade drops, etc.), 2 were reported to have no obvious external issues observed, 5 were judged as neutral, with teachers noting they had not paid much attention before but, upon reflection, recalled noteworthy details.

Overall, the information value rate of the AI's additional identification list exceeded 70%. It should be noted that teachers' judgments are influenced by their subjective experience, daily observation acuity, and personal perspectives, even experienced teachers find it difficult to accurately grasp every student's true psychological state. Nevertheless, the preliminary integrative judgments still hold important reference value.

### 3.4. Cross-comparison between MHT scale and AI assessment

Comparing the AI priority-check list (15 students) with MHT results: 9 of the 15 (60%) scored in the "suboptimal" or "professional assessment" range on the MHT, while 6 (40%) had MHT scores in the "normal" range but were still marked as priority checks by the AI. Among these 6 students, teacher verification confirmed that 4 had significant external pressures. This cross-analysis suggests that AI conversational assessment may capture issues not truthfully reflected in static scales, helping to compensate for the under-reporting limitations of traditional self-report scales.

### 3.5. Student feedback

The school conducted a post-class survey among 211 students, with the following results: 98% expressed "satisfied" or "very satisfied" with the conversational psychological assessment mode, 100% indicated willingness to continue using it. In open-ended responses, many students mentioned that interacting with the AI digital human felt "more relaxed," "stress-free," "not worried about being judged," and "able to speak honestly" compared to traditional paper questionnaires. This sharply contrasts with the practical dilemma of reduced accuracy due to concealment in traditional scale assessments. The 211 feedback responses accounted for 36.1% of all participants, the voluntary sample may have a more positive attitude toward AI interaction.

DOI: 10.54254/2753-7048/2026.PE38243

# 4. Discussion

## 4.1. Main findings

This paper provides preliminary empirical evidence for the application of AIGC-based conversational psychological risk screening in secondary school mental health work.

Under sufficient interaction depth, the matching rate between AI assessment and the school's existing observation list reached 90%, indicating that the model achieves practically useful concurrent validity. This finding supports previous research conclusions that AI can capture psychological distress markers through paralinguistic and linguistic features.

The most groundbreaking finding is that the AI independently identified 56 students not noticed by routine school methods, among the priority-check group, teacher verification showed that over 70% had genuine external pressures or personal characteristics, addressing the under-reporting bottleneck caused by social desirability bias in traditional scales, and suggesting that dynamic, conversational AI may effectively circumvent the limitations of static scales.

In terms of student acceptance, 98% satisfaction and 100% willingness to continue using are important practical advantages. This indicates that the digital native generation perceives AI mental health interaction as an attractive, safe, and destigmatizing experience, in stark contrast to the negative perceptions often associated with paper-and-pencil screening.

## 4.2. Practical implications and theoretical contributions

From a practical standpoint, this study offers an "active + accompanied" supplementary solution for school mental health screening. Beyond routine semester assessments, the AI system can conduct broader-scale risk screening at relatively low human cost, helping psychological teachers and homeroom teachers focus their limited attention on students who truly need care.

From a theoretical perspective, this study extends the concept of "destigmatized assessment." One core dilemma in traditional mental health screening is the respondent's defensiveness and masking, AIGC dialogues, by simulating a "non-judgmental listener," may reduce social desirability motivation and bring assessment closer to true states. This mechanism warrants further dissection and validation in subsequent research.

## 4.3. Limitations

This study has several limitations. First, it was conducted in only one school, with a sample mainly from urban public junior high schools, limiting external validity to different types of schools (rural, private, vocational, etc.). Second, the "school observation list" as a criterion is imperfect, being influenced by teachers' subjective biases, attention preferences, and observational ability differences, it cannot be equated with clinical "mental disorder criteria." Third, AI assessment effectiveness depends on sufficient interaction depth, accuracy dropped to 60% at low interaction depths, and interaction willingness itself may correlate with students' psychological state and personality traits – the most vulnerable yet extremely silent students may be precisely those with the fewest exchanges, creating a "silent-student paradox." Fourth, the acceptance survey covered only 36% of participants, and voluntary responses may introduce positive bias.

DOI: 10.54254/2753-7048/2026.PE38243

## 5. Conclusion

This empirical study conducted at Chunjiang Middle School in Hangzhou, Zhejiang Province, shows that the AIGC-based conversational mental health assessment model demonstrates good consistency with the school's observation list (90% matching rate under effective interaction depth), and can independently identify a substantial number of at-risk students not routinely noticed (56 additional identifications, with an information value rate exceeding 70%). Meanwhile, the model enjoys high acceptance among students (98% satisfaction, 100% willingness to continue, though the voluntary sampling bias should be noted). These findings preliminarily indicate that AIGC interactive assessment can serve as an effective, acceptable, and large-scale implementable supplementary screening approach, effectively overcoming the under-reporting bottleneck caused by social desirability bias in traditional scales.

Current shortcomings are: (1) Limited sample representativeness–conducted in only one school, repeated validation across multiple schools in urban/rural, different socioeconomic levels, and different grade levels is needed. (2) Criterion not ideal–the "school observation list" relies on teacher subjective judgment, not a clinical gold standard, it is recommended to use structured clinical interviews (e.g., K-SADS) in sub-samples for criterion validation. (3) Interaction depth dependence–accuracy is only 60% at low depth, it is recommended to design guided protocols with minimum interaction rounds, or use light gamification to enhance initial participation among silent students, while also exploring interaction depth itself as a supplementary risk indicator. (4) Acceptance survey bias–only 36% feedback, future assessments should embed brief experience evaluations at the end of each conversation for full-sample acceptance collection.

Future research should further advance multimodal data fusion – with informed consent, combining voice features, facial expression recognition, and behavioral metadata such as response time, correction frequency, and topic avoidance patterns – to build more robust and camouflage-resistant unobtrusive assessment models. Longitudinal predictive validity studies should also be conducted, tracking the 56 additionally identified students over one to two years, validating the true preventive value of screening through actual mental health outcomes such as counseling help-seeking rates, psychological teacher referral rates, and clinical diagnosis rates.

Furthermore, efforts should be made to bridge the screening-to-intervention pathway, directly linking AI assessment results to personalized, evidence-based digital micro-intervention modules, such as self-help conversational robots based on cognitive behavioral therapy, forming a closed-loop service system of "initial screening – triage – light intervention – referral." For the low-interaction-depth silent group, alternative non-verbal assessment strategies – such as emoji selection, card sorting, or drawing description – should be developed to ensure that the most vulnerable children are not missed.

## References

[1] Patel, V., Saxena, S., Lund, C., et al. (2018). The Lancet Commission on global mental health and sustainable development. *Lancet, 392*(10157), 1553-1598.
[2] World Health Organization. (2021). Adolescent mental health. WHO. Geneva.
[3] Deighton, J., Tymms, P., Vostanis, P., et al. (2013). The development of a school-based measure of child mental health. *J Psychoeduc Assess, 31*(3), 247-257.
[4] Lien, Y. J., Kao, R. H., Liu, Y. E., & Chang, H. A. (2020). Social desirability bias in adolescents' self-reported mental health: The role of stigma. *Child Indic Res, 13*(4), 1405-1422.
[5] Jiang, H., Li, Q., Huang, T., et al. (2021). Mental health status and related factors among middle school students in Qianjiang District, Chongqing. *Chinese Journal of School Health, 42*(8), 1212-1215.

DOI: 10.54254/2753-7048/2026.PE38243

[6] Graham, S., Depp, C., Lee, E. E., et al. (2019). Artificial intelligence for mental health and mental illness: An overview. *Current Psychiatry Reports, 21*(11), 116.
[7] Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *NPJ Digital Medicine, 3*, 43.
[8] Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology, 13*, 23–47.
[9] Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology, 5*(6), 1002–1013.
[10] Zhou, B. (1991). *Manual of mental health test (MHT)*. East China Normal University Press. Shanghai.