Development and Psychometric Testing of a Questionnaire for Evaluating the SINA (Sahabat Ibu Anti Anemia) Maternal Health Chatbot

Erli Zainal[1,4], Sri Sulistyowati[2], Agung Wibowo[3], and Widiyanto Widiyanto[3]

[1] Graduate Program in Community Empowerment and Development Communication, Universitas Sebelas Maret, Surakarta, Indonesia
[2] Faculty of Medicine, Universitas Sebelas Maret, Surakarta, Indonesia
[3] Faculty of Agriculture, Universitas Sebelas Maret, Surakarta, Indonesia
[4] Midwifery Study Program, Sapta Bakti College of Health Sciences, Bengkulu, Indonesia
zainalerli@gmail.com

Abstract. Background: Iron-deficiency anemia (IDA) in pregnancy remains a critical health concern in Indonesia, particularly in Seluma Regency, Bengkulu Province, where the prevalence remains high. Limited compliance with iron–folic acid supplementation underscores the need for innovative maternal health education strategies. The SINA (Sahabat Ibu Anti Anemia) chatbot was designed to deliver interactive, digital-based health education for pregnant women. This study aimed to develop and psychometrically test a questionnaire to evaluate the effectiveness of the SINA chatbot. Methods: Instrument development was based on Green's PRECEDE–PROCEED model. The questionnaire initially contained 62 items across three domains: knowledge (22), attitudes (30), and practices (10). Content validity was evaluated by experts using the Content Validity Index (CVI), readability was tested among five pregnant women, and construct validity and reliability were examined among 30 pregnant women. Validity testing employed item-total correlation ($r > 0.30$), and reliability was assessed using Cronbach's Alpha. Results: The overall CVI was 0.89, indicating strong content validity. Readability testing showed good comprehension in the knowledge (4.36) and practice (4.72) domains and moderate comprehension in attitudes (3.82). After item analysis, 46 items were valid, with 3 items revised. Cronbach's Alpha for the overall instrument was 0.87, with subscale reliabilities of 0.82 (knowledge), 0.85 (attitudes), and 0.80 (practices). Conclusion: The SINA questionnaire demonstrated strong validity and reliability and can serve as a valid instrument for preliminary evaluation of chatbot-based maternal health interventions. Further testing with larger samples and confirmatory factor analysis is recommended.

Keywords: Maternal health; chatbot; questionnaire; validity; reliability; psychometric testing

H. Saptaningtyas et al. (eds.), *Proceedings of the 4th International Conference on Multidisciplinary Studies 2025 (ICOMSI 2025)*, Advances in Social Science, Education and Humanities Research 1012,
https://doi.org/10.2991/978-2-38476-571-3_16

1 Introduction

Iron-deficiency anemia (IDA) is one of the most prevalent nutritional disorders among pregnant women globally. The World Health Organization (WHO) reports that approximately 40% of pregnant women suffer from anemia, with the highest prevalence found in Southeast Asia [1]. In Indonesia, anemia affects more than 40% of pregnant women [2]. In Bengkulu Province, the prevalence remains high, particularly in Seluma Regency (11.61% in 2023) [3].

Anemia during pregnancy has been linked to several adverse maternal and neonatal outcomes, including fatigue, preterm birth, and low birth weight [4]. Although iron–folic acid supplementation (IFA) programs have been implemented nationwide, adherence remains low due to poor knowledge, misconceptions, and limited access to health education [5].

The rise of digital health technologies, including mobile health (mHealth) and conversational agents or chatbots, provides new opportunities for interactive health promotion [6], [7]. The SINA (Sahabat Ibu Anti Anemia) chatbot was developed to deliver educational messages on anemia prevention via WhatsApp. To evaluate its effectiveness, a psychometrically tested questionnaire is required to measure changes in knowledge, attitudes, and practices among pregnant women.

This study aimed to develop and validate such an instrument through comprehensive psychometric testing involving content validity, readability, construct validity, and reliability analyses.

2 Methods

2.1 Study Design

This study employed a cross-sectional design for the development and validation of a questionnaire intended to assess the SINA maternal health chatbot.

2.2 Instrument Development

Item construction was guided by Green's PRECEDE–PROCEED model [8], covering predisposing (knowledge, attitudes) and reinforcing (practices) factors in anemia prevention. The initial version comprised 62 items distributed across three domains: knowledge (22), attitudes (30), and practices (10).

2.3 Content Validity

Three subject-matter experts in maternal health, health promotion, and public health assessed each item's relevance, clarity, and representativeness using a 4-point Likert scale. The overall Content Validity Index (CVI) was 0.89, reflecting excellent expert agreement (Table 1).

Table 1. Content Validity Index (CVI)

Domain	No. of Items	CVI	Interpretation
Knowledge	22	0.89	High validity
Attitudes	30	0.88	High validity
Practices	10	0.90	High validity
Overall	62	0.89	Excellent

2.4 Readability Testing

The readability test was conducted with five pregnant women of varied education levels (junior high to college). Each participant rated the clarity and understandability of each domain on a 5-point scale. Results are shown in Table 2.

Table 2. Readability Test Results

Domain	Mean Score	Interpretation
Knowledge	4.36	Easy to understand
Attitudes	3.82	Moderate; minor revision required
Practices	4.72	Very easy to understand

2.5 Construct Validity and Reliability

The revised questionnaire was administered to 30 pregnant women attending antenatal care in Seluma Regency. Construct validity was assessed using Pearson's item-total correlation (cut-off $r \geq 0.30$) [9]. Internal consistency reliability was analyzed using Cronbach's Alpha (acceptable ≥ 0.70) [10].

Table 3. Construct Validity and Reliability Results

Domain	Initial Items	Valid Items	Revised Items	Cronbach's Alpha
Knowledge	22	16	0	0.82
Attitudes	30	20	0	0.85
Practices	10	10 (7 valid, 3 re-vised)	3	0.80
Total	62	46 valid items	3 revised items	0.87 (overall)

2.6 Ethical Considerations

Ethical approval was obtained from the Ethics Committee of Sapta Bakti College of Health Sciences, Bengkulu, Indonesia. All participants provided informed consent.

3 Results

3.1 Content Validity Results

Content validity of the questionnaire was evaluated by three experts with backgrounds in maternal health, health promotion, and public health. The overall Content Validity Index (CVI) of the instrument was 0.89, indicating a high level of agreement among experts regarding item relevance, clarity, and representativeness.

Across domains, the CVI values were consistently high, with knowledge items achieving a CVI of 0.89, attitude items 0.88, and practice items 0.90. These results suggest that the initial pool of items adequately reflected key constructs related to anemia prevention and maternal health education within the context of chatbot-based interventions. Minor revisions were recommended by experts, primarily involving simplification of technical terms and rephrasing of negatively worded statements to enhance clarity.

3.2 Readability Test Results

Readability testing was conducted among five pregnant women with diverse educational backgrounds, ranging from junior high school to higher education. The mean readability score for the knowledge domain was 4.36, indicating that most items were easy to understand. The practice domain showed the highest readability score (4.72), suggesting that behavior-related questions were highly comprehensible and directly aligned with respondents' daily experiences.

In contrast, the attitude domain yielded a lower mean score of 3.82, indicating moderate readability. Participants reported difficulty understanding items that used abstract expressions or negative sentence constructions. Based on this feedback, several attitude items were revised to improve sentence structure and simplify wording, while maintaining their conceptual meaning. Overall, the readability findings indicate that the questionnaire was generally well understood by the target population, with targeted improvements required mainly for attitudinal statements.

3.3 Construct Validity Results

Construct validity was assessed using item–total correlation analysis based on responses from 30 pregnant women. A correlation coefficient threshold of $r \geq 0.30$ was applied to determine item validity. Of the 62 initial items, 46 items met the validity criteria, while the remaining items were either excluded or revised.

Within the knowledge domain, 16 out of 22 items demonstrated adequate item–total correlations and were retained in the final instrument. In the attitude domain, 20 of the 30 items met the validity threshold, indicating that several attitudinal statements did not consistently align with the overall construct. All 10 practice items were retained, although three items underwent minor revisions to improve clarity based on both statistical results and qualitative feedback.

These findings suggest that knowledge and practice items exhibited stronger construct alignment compared to attitude items, which required more refinement to capture consistent respondent perceptions.

3.4 Reliability Results

Internal consistency reliability of the revised questionnaire was assessed using Cronbach's Alpha. The overall reliability coefficient for the final instrument was $\alpha = 0.87$, indicating excellent internal consistency.

Subscale reliability analysis showed Cronbach's Alpha values of 0.82 for the knowledge domain, 0.85 for the attitude domain, and 0.80 for the practice domain. All values exceeded the commonly accepted threshold of 0.70, demonstrating that the retained items within each domain consistently measured their respective constructs.

Taken together, the validity and reliability analyses resulted in a final questionnaire consisting of 46 valid items, with three practice items revised for clarity. These results indicate that the instrument possesses strong psychometric properties and is suitable for preliminary evaluation of maternal health chatbot interventions.

4 Discussion

This study aimed to develop and psychometrically test a questionnaire designed to evaluate the SINA (Sahabat Ibu Anti Anemia) maternal health chatbot. Overall, the findings demonstrate that the instrument possesses strong content validity, satisfactory construct validity, and high internal consistency reliability, indicating its suitability for preliminary evaluation of chatbot-based maternal health education interventions.

4.1 Content Validity and Conceptual Adequacy

The high overall Content Validity Index (CVI = 0.89) indicates strong agreement among experts regarding the relevance and representativeness of the questionnaire items. This finding suggests that the instrument adequately captures key constructs related to anemia prevention and maternal health education. According to Polit and Beck, a CVI value above 0.80 reflects excellent content validity and indicates that items are conceptually aligned with the intended measurement domain [8].

The consistently high CVI scores across knowledge, attitude, and practice domains further support the conceptual robustness of the instrument. This is particularly important in the context of digital health interventions, where evaluation tools must reflect both biomedical content and behavioral dimensions of health education [9]. The use of Green's PRECEDE–PROCEED model as the theoretical foundation likely contributed to this strong conceptual alignment, as the model emphasizes the role of predisposing and reinforcing factors in shaping health behaviors [10].

4.2 Readability and Comprehension Among Pregnant Women

Readability testing revealed that the knowledge and practice domains were generally easy to understand, whereas the attitude domain showed comparatively lower readability. This pattern is consistent with previous studies reporting that attitude-related items tend to be more abstract and cognitively demanding than factual knowledge or behavioral questions, particularly among populations with varied educational backgrounds [1].

Pregnant women participating in the readability test reported difficulty with negatively worded and abstract attitudinal statements. Similar findings have been reported in maternal health surveys, where complex sentence structures and negative phrasing were associated with reduced comprehension and response accuracy [4]. The revisions made to these items were therefore essential to enhance clarity without altering the underlying construct. These findings underscore the importance of linguistic simplicity and cultural adaptation in developing instruments for maternal health education, especially in low- and middle-income settings [6].

4.3 Construct Validity Across Domains

Construct validity analysis indicated that not all initial items demonstrated adequate item–total correlations, particularly within the attitude domain. This result suggests that attitudes toward anemia prevention may be more heterogeneous and context-dependent than knowledge or practice behaviors. Attitudinal constructs are influenced by personal beliefs, social norms, and cultural expectations, which can vary substantially among pregnant women [7].

In contrast, knowledge and practice items exhibited stronger construct alignment, with most items meeting the validity threshold. This finding aligns with previous research indicating that factual knowledge and observable behaviors tend to yield more stable psychometric properties than attitudinal measures [11]. The retention of all practice items, albeit with minor revisions, suggests that behavioral indicators related to iron tablet consumption and anemia prevention are more directly interpretable and consistently perceived by respondents.

4.4 Reliability and Internal Consistency

The reliability analysis demonstrated high internal consistency for the overall instrument (Cronbach's Alpha = 0.87) and for each subscale ($\alpha = 0.80–0.85$). According to Tavakol and Dennick, Cronbach's Alpha values above 0.70 are considered acceptable, while values above 0.80 indicate good reliability for newly developed instruments [12]. The reliability coefficients obtained in this study therefore suggest that the questionnaire items consistently measure their respective constructs.

High reliability is particularly important for instruments intended to evaluate digital health interventions, as repeated assessments are often required to monitor changes in

knowledge, attitudes, and practices over time [13]. The strong internal consistency observed in this study supports the use of the questionnaire as a stable measurement tool in future chatbot-based maternal health evaluations.

4.5 Implications for Chatbot-Based Maternal Health Evaluation

The final 46-item questionnaire provides a structured and psychometrically sound instrument for evaluating the educational impact of the SINA maternal health chatbot. Unlike many existing evaluations that rely on ad hoc or non-validated measures, this instrument offers a theoretically grounded and empirically tested approach to assessing digital health education outcomes [9,14].

By capturing knowledge, attitudes, and practices simultaneously, the questionnaire enables a comprehensive evaluation of behavioral change mechanisms targeted by chatbot interventions. This is particularly relevant in maternal health promotion, where improvements in knowledge alone may not translate into sustained behavioral change without corresponding shifts in attitudes and practices [15].

4.6 Study Limitations and Future Directions

Despite its strengths, this study has several limitations. The psychometric testing was conducted with a relatively small sample size (n = 30), which limits the generalizability of the findings. As recommended in psychometric literature, larger samples are required for advanced analyses such as exploratory and confirmatory factor analysis [8,11]. Future studies should therefore validate the questionnaire using larger and more diverse populations to further strengthen construct validity.

Additionally, this study focused on preliminary validation. Longitudinal studies are needed to assess the sensitivity of the instrument in detecting changes following chatbot-based interventions. Nevertheless, the present findings provide a strong foundation for further instrument refinement and application.

5 Limitations

This study was limited by a small sample (n=30), below the typical psychometric standard of five respondents per item. As a result, the findings should be considered preliminary. Additionally, data were collected from a single district, which may limit cultural representativeness.

6 Conclusion

This study successfully developed and conducted preliminary psychometric testing of a questionnaire designed to evaluate the SINA (Sahabat Ibu Anti Anemia) maternal health chatbot. The findings demonstrate that the instrument possesses strong content

validity, acceptable construct validity, and high internal consistency reliability, indicating that it is suitable for use as an evaluation tool in digital maternal health education.

The overall Content Validity Index (CVI = 0.89) confirms that the questionnaire items are conceptually relevant and representative of key domains related to anemia prevention, namely knowledge, attitudes, and practices. Readability testing further indicated that the instrument is generally understandable by pregnant women, although attitudinal items required minor linguistic refinement. These adjustments improved clarity without compromising the underlying constructs.

Construct validity and reliability analyses resulted in a final instrument consisting of 46 valid items, with three practice items revised for clarity. The overall Cronbach's Alpha coefficient ($\alpha = 0.87$) and strong subscale reliabilities support the internal consistency of the instrument and its potential utility for repeated assessments in chatbot-based interventions.

Importantly, this questionnaire provides a theoretically grounded and empirically tested tool for evaluating maternal health chatbots, addressing a critical gap in digital health research where evaluation often relies on non-validated measures. By capturing knowledge, attitudes, and practices simultaneously, the instrument enables a more comprehensive assessment of behavioral change mechanisms targeted by chatbot-based education.

Although this study represents a preliminary validation with a limited sample size, the results provide robust evidence supporting the feasibility and reliability of the instrument. Future research should involve larger and more diverse populations and apply advanced psychometric analyses, such as factor analysis, to further strengthen construct validation. Nevertheless, the present findings support the use of the questionnaire as a promising evaluation instrument for maternal health chatbot interventions aimed at anemia prevention.

Acknowledgments. The authors thank the pregnant women who participated in this study and the health workers in Seluma Regency. Appreciation is extended to the expert reviewers for their contributions and to Universitas Sebelas Maret for institutional support.

References

1. World Health Organization: Daily Iron and Folic Acid Supplementation During Pregnancy. WHO, Geneva (2012)
2. Indonesian Ministry of Health: Indonesia Health Profile 2023. Ministry of Health of the Republic of Indonesia, Jakarta (2024)
3. Bengkulu Provincial Health Office: Maternal Health Report 2023. Provincial Health Office of Bengkulu, Bengkulu (2024)
4. Haider, B.A., Olofin, I., Wang, M., Spiegelman, D., Ezzati, M., Fawzi, W.W.: Anaemia, prenatal iron use, and risk of adverse pregnancy outcomes: systematic review and meta-analysis. BMJ 346, f3443 (2013)
5. Notoatmodjo, S.: Promosi Kesehatan dan Ilmu Perilaku. Rineka Cipta, Jakarta (2012)

6. Palanica, A., Flaschner, P., Thommandram, A., Li, M., Fossat, Y.: Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. JMIR Medical Education 5(1), e12887 (2019)
7. Brixey, J., Johnson, T.R., Zachary, W.: A usability evaluation of an intelligent virtual assistant for patient education. Journal of Biomedical Informatics 95, 103208 (2019)
8. Polit, D.F., Beck, C.T.: Nursing Research: Generating and Assessing Evidence for Nursing Practice. 11th edn. Wolters Kluwer, Philadelphia (2021)
9. Tavakol, M., Dennick, R.: Making sense of Cronbach's alpha. International Journal of Medical Education 2, 53–55 (2011)
10. Green, L.W., Kreuter, M.W.: Health Program Planning: An Educational and Ecological Approach. 4th edn. McGraw-Hill, New York (2005)
11. DeVellis, R.F.: Scale Development: Theory and Applications. 4th edn. Sage Publications, Thousand Oaks, CA (2017)
12. Osborne, J.W., Costello, A.B.: Sample size and subject to item ratio in principal components analysis. Practical Assessment, Research & Evaluation 9(11), 1–9 (2004)
13. Nutbeam, D.: Health literacy as a public health goal: A challenge for contemporary health education. Health Promotion International 15(3), 259–267 (2000)
14. Lupton, D.: Health promotion in the digital era: A critical commentary. Health Promotion International 30(1), 174–183 (2014)
15. Fitzpatrick, R., Davey, C., Buxton, M.J., Jones, D.R.: Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment 2(14), 1–74 (1998)