

Psychometric Evaluation of a START Triage Psychomotor Observation Instrument: Content Validity and Interrater Reliability

Julfikar Adnan Irnu¹, Novia Faizatiwahida², Superzeki Zaidatul Fadilah²,
Desy Listyaningrum²



¹ Politeknik Karya Husada Jakarta, Indonesia

² Faculty of Nursing, Universitas Jember, Indonesia

Correspondence should be addressed to:

Julfikar Adnan Irnu

julfikaradnan07@gmail.com

Abstract:

Accurate psychomotor assessment in Simple Triage and Rapid Treatment (START) is critical during mass-casualty incidents. Observation-based evaluations often yield inconsistent judgments when utilizing unvalidated instruments, compromising competency assessments. This study aimed to establish the content validity and interrater reliability of a newly developed psychomotor observation sheet for START triage among healthcare professionals. This methodological study evaluated a 12-item dichotomous observation sheet derived from standard START protocols. Content validity was determined by three emergency and disaster nursing experts using a four-point relevance scale, analyzed with Aiken's V. Interrater reliability was assessed with 36 medical staff members, independently evaluated by two trained raters. Agreement metrics included percent agreement and Cohen's kappa. Expert evaluation yielded a perfect Aiken's V of 1.00 across all 12 items, indicating exceptional content validity. Interrater reliability testing demonstrated percent agreement ranging from 94.44% to 100%, alongside Cohen's kappa coefficients between 0.85 and 1.00. These metrics significantly exceeded minimum acceptability thresholds, reflecting outstanding consistency and minimal scoring ambiguity among raters. The START triage psychomotor observation sheet exhibits robust content validity and excellent interrater reliability. This validated instrument provides educators with a standardized, objective tool for evaluating clinical competencies in disaster preparedness training, though future validation in high-fidelity simulation environments is recommended.

Article info:

Submitted:

21-05-2026

Revised:

29-06-2026

Accepted:

07-07-2026

Published:

03-08-2026

Keywords:

clinical competence, disaster nursing, mass casualty incidents, psychometrics, psychomotor performance

DOI: <https://doi.org/10.53713/htechj.v4i4.739>

This work is licensed under CC BY-SA License.



INTRODUCTION

Mass-casualty incidents demand rapid, structured triage decisions to optimize the allocation of limited healthcare resources during sudden surges in patient volume (Armstrong, 2024). Emergency responders must quickly prioritize patients requiring urgent intervention to prevent the inappropriate distribution of personnel, equipment, and referral capacity (Gabbe et al., 2022). Recent literature on mass-casualty triage emphasizes that effective tools must remain simple, fast, and highly applicable in resource-limited contexts where healthcare professionals must manage overwhelming numbers of victims (Abdul-Nabi & Hitti, 2025). This global necessity underscores the critical importance of standardized triage protocols in modern disaster response frameworks (Javid et al., 2025).

The Simple Triage and Rapid Treatment (START) protocol requires healthcare professionals to translate complex clinical assessments into immediate physical actions within a strict 30- to 60-

second timeframe per victim (Franc et al., 2022). This approach classifies victims by rapidly evaluating walking ability, respiration, perfusion, and mental status (Lin et al., 2022). The instantaneous translation of assessment rules into decisive physical actions makes psychomotor competence a central component of safe triage performance. Cognitive understanding alone is insufficient to ensure accurate patient categorization in the chaotic environment of a mass casualty incident (Abdul-Nabi & Hitti, 2025).

Recent educational interventions have successfully demonstrated that targeted START triage training significantly enhances the cognitive, affective, and psychomotor domains of medical staff and healthcare professionals (Hosseinzadeh et al., 2024). Training studies indicate substantial improvements in the execution of practical skills following structured disaster preparedness programs (Faizatiwahida et al., 2025). Previous instrument development efforts have also established that knowledge and attitude instruments related to START triage can be effectively validated for medical and health personnel (Faizatiwahida et al., 2024). These advancements highlight a strong institutional focus on improving disaster response capabilities through comprehensive educational strategies.

Observation-based psychomotor assessments remain highly vulnerable to rater variation and subjective bias when the underlying measurement items lack sufficient clarity and standardized interpretation (Shabani et al., 2025). Direct observation of physical skills requires absolute precision in item formulation to prevent divergent scoring among different evaluators (Wisnesky et al., 2022). This methodological issue is critically important because modern healthcare education increasingly demands highly valid instruments for accurately assessing observable clinical skills (Bourassa et al., 2024). The absence of rigorously validated psychomotor tools creates a significant barrier to objectively measuring actual clinical competence in disaster scenarios.

Establishing robust content validity ensures that assessment items accurately represent the intended clinical construct, while evaluating interrater reliability confirms that different evaluators assign consistent scores to identical performances (Huabbangyang et al., 2024). Content validity is a fundamental step in instrument development, evaluating whether items adequately represent the construct being measured and remain understandable to users (Mokkink et al., 2025). Researchers commonly utilize Aiken's V to quantify expert agreement on item relevance and clarity (Aiken, 1985). Interrater reliability metrics, including Cohen's kappa and percent agreement, are frequently used in health research to assess observational consistency (Zhao et al., 2022).

The primary novelty of the current investigation lies in the development and rigorous psychometric validation of a dichotomous observation sheet specifically designed to evaluate START triage psychomotor execution. Existing literature predominantly focuses on cognitive and affective assessments, leaving a distinct gap in tools designed for direct physical observation. This study introduces a specialized 12-item instrument that captures the essential, observable steps of victim assessment, including initial sorting, respiration, perfusion, and mental status evaluation (Moura et al., 2022). The dichotomous scoring format specifically mitigates ambiguity during direct observation.

The present study aims to examine the content validity and interrater reliability of this newly developed psychomotor observation sheet, providing educators with a standardized, objective tool for competency assessment in disaster preparedness training. Addressing this methodological gap is urgently needed to ensure that healthcare professionals possess the verified physical skills required in real-world mass-casualty incidents (Carenzo et al., 2022). The study contributes vital psychometric evidence for an observation-based instrument that can be seamlessly integrated into START triage training, competency assessment, and broader disaster preparedness education initiatives.

METHOD

Research Design

This study used a methodological research design to evaluate the psychometric quality of a START triage psychomotor observation sheet. The validation process consisted of two stages: content validity testing through expert judgment and interrater reliability testing through direct observation of respondents' START triage performance.

Participants and Sampling

The content validity phase involved three experts with academic and practical expertise in emergency care, disaster nursing, and triage. For the interrater reliability phase, 36 medical staff and healthcare professionals in Indonesia performed START triage procedures, which were independently evaluated by two trained raters. All data were collected during structured psychomotor skills assessment sessions.

Instrument

The instrument was a 12-item psychomotor observation sheet developed from the standard START triage procedure. The items represented essential steps in victim assessment, including initial sorting, respiration, perfusion, and mental status assessment. Each item used a dichotomous scoring format, namely, performed or not performed, to reduce ambiguity during direct observation.

Content Validity Procedure

Experts assessed each item using four criteria: relevance, clarity, simplicity, and ambiguity. A four-point scale was used, ranging from 1 to 4. Content validity was analyzed using Aiken's V formula: $V = \sum s / [n(c - 1)]$, where s is the difference between the expert rating and the lowest possible score, n is the number of experts, and c is the number of rating categories. Aiken's V values were interpreted as very high when they ranged from 0.80 to 1.00, high when they ranged from 0.60 to 0.79, moderate when they ranged from 0.40 to 0.59, and low when they were below 0.40.

Interrater Reliability Procedure

Interrater reliability was examined by having two trained raters independently assess the same 36 respondents using the same observation sheet. Agreement between raters was analyzed using percent agreement and Cohen's kappa. An agreement of 80% or higher was considered acceptable for observational assessment in health research, while kappa values of 0.81 to 1.00 were interpreted as indicating very good agreement.

Ethical Declaration

The Ethics Committee of Politeknik Karya Husada Jakarta approved ethical clearance for this study. All assessments used de-identified performance data, focusing exclusively on clinical competencies rather than patient-specific information.

RESULTS

The findings are presented in two parts: content validity based on expert judgment and interrater reliability based on agreement between two independent raters.

Content Validity of the Observation Sheet

All 12 items obtained an Aiken's V value of 1.00. This result indicates very high content validity for every item in the psychomotor observation sheet. The findings suggest that the experts considered the items relevant, clear, simple, and sufficiently unambiguous for assessing START triage psychomotor skills.

Table 1. Content validity results of the START triage psychomotor observation sheet

Item	Aiken's V	Interpretation
1	1.00	Very high
2	1.00	Very high
3	1.00	Very high
4	1.00	Very high
5	1.00	Very high
6	1.00	Very high
7	1.00	Very high
8	1.00	Very high
9	1.00	Very high
10	1.00	Very high
11	1.00	Very high
12	1.00	Very high

Interrater Reliability of the Observation Sheet

The interrater reliability results showed consistently high agreement between the two raters across all indicators. Percent agreement ranged from 94.44% to 100%, while Cohen's kappa ranged from 0.85 to 1.00. Therefore, all indicators met the acceptance criteria and were interpreted to indicate very good agreement.

Table 2. Interrater reliability results of the START triage psychomotor observation sheet

Indicator	Percent Agreement (%)	PA Decision	Cohen's Kappa	Kappa Interpretation
Indicator 1	97.22	Accepted	0.92	Very good
Indicator 2	100.00	Accepted	1.00	Very good
Indicator 3	97.22	Accepted	0.92	Very good
Indicator 4	94.44	Accepted	0.85	Very good
Indicator 5	100.00	Accepted	1.00	Very good
Indicator 6	97.22	Accepted	0.85	Very good
Indicator 7	100.00	Accepted	1.00	Very good
Indicator 8	100.00	Accepted	1.00	Very good
Indicator 9	100.00	Accepted	1.00	Very good
Indicator 10	100.00	Accepted	1.00	Very good
Indicator 11	97.22	Accepted	0.89	Very good
Indicator 12	100.00	Accepted	1.00	Very good

Note. Percent agreement (PA)

DISCUSSION

The present study established exceptional content validity for the newly developed START triage psychomotor observation sheet, with all twelve items achieving a perfect Aiken's V score of 1.00. This perfect score reflects unanimous consensus among the expert panel regarding the relevance, clarity, and representativeness of the assessment items. Content validity extends beyond mere surface relevance to ensure that the instrument comprehensively captures the underlying construct being measured while remaining comprehensible to the intended users (Mokkink et al., 2025). The START triage construct encompasses rapid, observable clinical actions, including victim

sorting, respiratory evaluation, perfusion assessment, and mental status assessment. Expert agreement confirms that the selected items accurately map onto these critical psychomotor domains without introducing unnecessary complexity. The operational nature of START triage demands that assessment tools mirror the rapid translation of clinical signs into definitive triage categories. Quantitative validation through Aiken's V provides robust empirical evidence that expert judgment consistently supports this structural alignment (Aiken, 1985). This rigorous validation process confirms that the observation sheet functions effectively as a comprehensive psychomotor assessment tool rather than a superficial checklist of theoretical knowledge.

Interrater reliability testing further confirmed the robust consistency of the observation sheet, yielding percent agreement values above 94% and Cohen's kappa coefficients ranging from 0.85 to 1.00 across all indicators. These statistical outcomes indicate very good agreement beyond chance, demonstrating that independent raters interpret and score the psychomotor performances in a highly uniform manner. Observation-based assessment in clinical education depends heavily on rater interpretation, making the mitigation of subjective bias a critical priority. Inconsistency between raters can severely undermine the credibility of competency decisions and compromise the fairness of high-stakes evaluations (Zhao et al., 2022). The exceptionally high agreement observed in this study suggests that the operational definitions within the observation sheet are sufficiently clear to prevent divergent interpretations. The use of a dichotomous scoring format, which required raters to classify actions strictly as performed or not performed, played a pivotal role in reducing scoring ambiguity. This binary approach forces raters to focus on the presence or absence of specific, observable psychomotor steps, thereby minimizing the influence of subjective halo effects or personal grading leniency. The resulting statistical consistency proves that the instrument can be reliably deployed across different evaluators without sacrificing measurement precision.

The psychometric strengths of this observation instrument align seamlessly with recent empirical evidence emphasizing the critical need for competency-based disaster triage training. Faizatiwahida et al. (2025) demonstrated that targeted START triage training significantly improves the psychomotor execution skills of medical staff and healthcare professionals during simulated mass casualty incidents. The current study builds upon these educational outcomes by providing a specific psychometric tool to objectively measure these psychomotor improvements. Previous literature has also highlighted the broader impacts of disaster preparedness education on healthcare workers. Setyawati et al. (2025) found that structured disaster triage training programs substantially enhance nurses' knowledge, practical skills, and psychological resilience in emergency scenarios. The validated observation sheet from the present research serves as a crucial missing link in this educational ecosystem, providing educators with a reliable means to quantify the practical skill acquisition reported in those training studies. By supplying a rigorously tested assessment instrument, this research enables future educational interventions to move beyond self-reported confidence and instead measure actual, observable clinical competence.

This investigation also successfully bridges a significant methodological gap in the existing literature by shifting the focus from cognitive and attitudinal assessments to the direct evaluation of observable clinical skills. Prior instrument development efforts in this domain have predominantly concentrated on measuring the theoretical understanding and disaster-related attitudes of healthcare personnel. Faizatiwahida et al. (2024) successfully validated knowledge and attitude instruments for START triage among medical personnel, establishing a strong foundation for cognitive assessment. Cognitive knowledge and positive attitudes, although necessary, do not automatically translate into flawless psychomotor execution during the chaotic environment of a mass casualty incident. The current study addresses this translational gap by introducing a validated tool specifically designed to capture the physical execution of triage protocols. Evaluating

psychomotor performance requires direct observation under standardized conditions, an approach that captures the physical dexterity, speed, and sequential accuracy demanded by real-world triage. The provision of this psychomotor observation sheet completes the triad of assessment tools, allowing researchers and educators to comprehensively evaluate the cognitive, affective, and psychomotor domains of disaster nursing competence (Pandawa et al., 2025).

The establishment of strong content validity and interrater reliability carries profound implications for disaster preparedness education and clinical competency evaluation within healthcare institutions. Validated psychomotor instruments are absolute prerequisites for ensuring that training outcomes are measured fairly, consistently, and objectively across diverse educational settings. Nursing educators and hospital training coordinators can now utilize this observation sheet to conduct standardized evaluations during simulation-based learning exercises. The tool enables the identification of specific psychomotor deficits, allowing instructors to provide targeted corrective feedback to individual learners before they encounter real mass-casualty incidents (Elshami et al., 2025). Healthcare institutions can integrate this instrument into their routine disaster drill protocols to objectively assess the readiness of their medical staff. Regulatory bodies and credentialing organizations may also adopt this validated sheet to standardize the certification processes for emergency and disaster nursing competencies. The availability of a reliable, dichotomous observation format significantly reduces the administrative burden on evaluators while maintaining high scientific rigor, facilitating the widespread implementation of competency-based assessments in resource-constrained training environments.

The robust psychometric properties demonstrated in this study are accompanied by methodological limitations that warrant careful interpretation of the findings. The interrater reliability testing phase was conducted with a relatively modest sample size of 36 respondents evaluated by only two trained raters. A larger and more diverse sample of evaluatees might introduce greater variability in performance execution, potentially challenging the universal applicability of the current reliability coefficients. The assessment took place in a controlled training environment, which inherently lacks the extreme physiological stress, environmental distractions, and chaotic dynamics characteristic of actual mass casualty incidents. Rater performance in a calm, simulated setting may not perfectly predict rater consistency under the high-pressure conditions of a real emergency. The exclusive reliance on a dichotomous scoring format, while excellent for reducing ambiguity, fails to capture the quality, speed, or fluidity of the psychomotor execution. Future investigations must address these constraints to ensure the instrument's utility extends beyond controlled academic settings into the unpredictable reality of disaster response.

Subsequent research endeavors should prioritize expanding this validation framework to encompass broader demographic samples and more complex simulation paradigms. Future studies must involve larger cohorts of healthcare professionals across diverse clinical settings to establish the observation sheet's generalizability. Incorporating high-fidelity, real-time mass-casualty simulations with standardized patients or advanced manikins will test the instrument's resilience under realistic stress conditions. Researchers should also explore integrating video-based remote assessment protocols to evaluate interrater reliability when evaluators are not physically present in the simulation room. Investigating the correlation between scores obtained from this psychomotor observation sheet and actual patient outcomes in simulated scenarios would provide additional criterion-related validity evidence. Longitudinal studies tracking the psychomotor retention of healthcare professionals over time using this specific instrument could yield valuable insights into the decay rates of triage skills. These progressive research steps will ultimately refine the tool and solidify its position as a gold standard for disaster triage competency assessment.

CONCLUSION

The newly developed START triage psychomotor observation sheet exhibits exceptional psychometric robustness, characterized by comprehensive expert consensus and outstanding interrater consistency. All evaluated items achieved maximum content validity and superior agreement across all behavioral indicators. This rigorous validation confirms the instrument's suitability for objectively measuring clinical execution and procedural adherence among healthcare providers during mass-casualty preparedness exercises. Educational institutions and emergency training coordinators can confidently deploy this standardized tool for competency assessment and simulation-based learning. Future investigations must expand the validation framework to encompass larger, more diverse clinical cohorts and high-fidelity disaster scenarios to establish broader generalizability.

ACKNOWLEDGEMENT

The authors would like to thank the experts, raters, and respondents who contributed to the validation and reliability testing process of this instrument.

CONFLICT OF INTEREST

The authors declare no conflict of interest related to this manuscript.

REFERENCES

- Abdul-Nabi, S. S., & Hitti, E. (2025). A review of mass casualty incident triage tools for hospital-based triage. *Turkish Journal of Emergency Medicine*, 25(4), 251–255. https://doi.org/10.4103/tjem.tjem_77_25
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Armstrong, J. B. P. (2024). Preparing your emergency department for disaster: Optimizing surge capacity during mass casualty events. *Healthcare Management Forum*. <https://doi.org/10.1177/08404704231199403>
- Bourassa, M., Kolb, W. H., Cunningham, S., Neurauter, V., Bourassa, A., Grieshaber, E., Loyd, K., Walden, R. R., & Wassinger, C. (2024). Psychomotor skills assessment in healthcare education: A scoping review. *Physical Therapy Reviews*, 29(1–3), 64–86. <https://doi.org/10.1080/10833196.2024.2348927>
- Carenzo, L., Ingrassia, P. L., Foti, F., Albergoni, E., Colombo, D., Sechi, G. M., Zoli, A., & Sironi, S. (2022). A region-wide all-hazard training program for prehospital mass casualty incident management: A real-world case study. *Disaster Medicine and Public Health Preparedness*, 17, e184. <https://doi.org/10.1017/dmp.2022.84>
- Elshami, S., Yakti, O., Mohamed Ibrahim, M. I., Awaisu, A., Sherbash, M., & Mukhalalati, B. (2025). Instruments for the assessment of disaster management among healthcare professionals: A scoping review. *Frontiers in Public Health*, 13, 1540743. <https://doi.org/10.3389/fpubh.2025.1540743>
- Faizatiwahida, N., Alfrisa, B., Nurjannah, I., & Alim, S. (2025). The effect of Simple Triage and Rapid Treatment (START) training on cognitive, affective, and psychomotor skills in performing START triage among medical staff and healthcare professionals. *Jurnal Berita Ilmu Keperawatan*, 18(2), 173–184. <https://doi.org/10.23917/bik.v18i2.9861>

- Faizatiwahida, N., Irnu, J. A., Ayuningtyas, P. A., Hatta, H., Nurjannah, I., & Alim, S. (2024). Validity and reliability of the START triage knowledge and attitude instrument in medical personnel and health workers. *Jurnal Berita Ilmu Keperawatan*, 17(2), 202–210. <https://doi.org/10.23917/bik.v17i2.5147>
- Franc, J. M., Kirkland, S. W., Wisnesky, U. D., Campbell, S., & Rowe, B. H. (2022). METASTART: A systematic review and meta-analysis of the diagnostic accuracy of the Simple Triage and Rapid Treatment (START) algorithm for disaster triage. *Prehospital and Disaster Medicine*, 37(1), 106–116. <https://doi.org/10.1017/S1049023X2100131X>
- Gabbe, B. J., Veitch, W., Mather, A., Curtis, K., Holland, A. J., Gomez, D., Civil, I., Nathens, A., Fitzgerald, M., Martin, K., Teague, W. J., & Joseph, A. (2022). Review of the requirements for effective mass casualty preparedness for trauma systems: A disaster waiting to happen? *British Journal of Anaesthesia*, 128(2), e158–e167. <https://doi.org/10.1016/j.bja.2021.10.038>
- Hosseinzadeh, M., Nematollahi, A. V., Afra, Y., Amini, P., & Navaei, A. R. (2024). Effectiveness of distance training based on the Simple Triage and Rapid Treatment (START) triage system in pre-hospital emergency. *Journal of Medical Education and Curricular Development*, 11. <https://doi.org/10.1177/23821205241278658>
- Huabbangyang, T., Thepmanee, D., Buaprasert, P., Chansomboon, P., Sri-on, J., & Rojsaengroeng, R. (2024). Content validity assessment of a newly developed emergency medical dispatch and triage protocol in Thailand. *Journal of Clinical Medicine*, 14(19), 7125. <https://doi.org/10.3390/jcm14197125>
- Javid, S. A., Jamil, M., Noor, B., Umer, M. R., Asghar, A., Hassan, M. R., Akbar, A., Hussain, S., & Khan, W. (2025). Emergency preparedness and response: Strategies for mass casualty incidents. *Cureus*, 17(7), e89158. <https://doi.org/10.7759/cureus.89158>
- Lin, Y., Chen, K., Wang, J., & Lai, P. (2022). Simple Triage and Rapid Treatment protocol for emergency department mass casualty incident victim triage. *The American Journal of Emergency Medicine*, 53, 99–103. <https://doi.org/10.1016/j.ajem.2021.12.037>
- Mokkink, L. B., Herbelet, S., Tuinman, P. R., & Terwee, C. B. (2025). Content validity: Judging the relevance, comprehensiveness, and comprehensibility of an outcome measurement instrument—A COSMIN perspective. *Journal of Clinical Epidemiology*, 185, 111879. <https://doi.org/10.1016/j.jclinepi.2025.111879>
- Moura, B. R. S., Oliveira, G. N., Medeiros, G., Vieira, S., & Nogueira, S. (2022). Rapid triage performed by nurses: Signs and symptoms associated with identifying critically ill patients in the emergency department. *International Journal of Nursing Practice*, 28(1), e13001. <https://doi.org/10.1111/ijn.13001>
- Pandawa, R. M., Saleh, F. M., & Muhammad, R. (2025). Development of disaster nursing competency assessment instruments. *International Journal of Public Health Science*, 14(1), 94–106. <https://doi.org/10.11591/ijphs.v14i1.24697>
- Setyawati, A.-D., Lu, Y.-Y., Wu, S.-F. V., & Liang, S.-Y. (2025). Effects of disaster triage training program on knowledge, skills, and resilience among nurses: A randomized controlled trial. *Disaster Medicine and Public Health Preparedness*, 19, e298. <https://doi.org/10.1017/dmp.2025.10233>
- Shabani, M., Heidarimoghadam, R., Faradmal, J., & Babamiri, M. (2025). Identifying the cognitive, psychomotor, and sensory-perceptual abilities required for the role of pre-hospital emergency medical personnel. *WORK*. <https://doi.org/10.1177/10519815241295907>
- Wisnesky, U. D., Kirkland, S. W., Rowe, B. H., Campbell, S., & Franc, J. M. (2022). A qualitative assessment of studies evaluating the classification accuracy of personnel using START in disaster triage: A scoping review. *Frontiers in Public Health*, 10, 676704. <https://doi.org/10.3389/fpubh.2022.676704>
- Zhao, X., Feng, G. C., Ao, S. H., & Liu, P. L. (2022). Interrater reliability estimators tested against true interrater reliabilities. *BMC Medical Research Methodology*, 22, Article 232. <https://doi.org/10.1186/s12874-022-01707-5>