MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438

Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

Research Article

EVALUATION OF THE QUALITY AND GUIDELINE ADHERENCE OF YOUTUBE VIDEOS ON NEONATAL FEVER: DEVELOPMENT AND VALIDATION OF THE NEWBORN FEVER SAFETY AND QUALITY SCORE (NFSQS)

Elif KELEŞ GÜLNERMAN[1]*, İbrahim Murat HİRFANOĞLU[2]

[1] Department of Radiology, Northwestern University Feinberg School of Medicine, Machine and Hybrid Intelligence Lab, Chicago, IL, USA

[2] Division of Neonatology, Department of Pediatrics, Gazi University School of Medicine, Ankara, TÜRKIYE

*Correspondence: elifkeles.dr@gmail.com

Received: *15 June 2026*
Revised: *19 June 2026*
Accepted: *03 August 2026*
Published: *24 September 2026*

Copyright: © 2026 by the authors. Published by Aydın Adnan Menderes University, Faculty of Medicine and Faculty of Dentistry. This article is openly accessible under the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Objective: To develop and validate the Newborn Fever Safety and Quality Score (NFSQS), a content-specific instrument for evaluating the quality and guideline adherence of Turkish-language YouTube videos on neonatal fever.

Materials and Methods: An initial pool of 18 candidate items was developed from commonly searched parental concerns and recommendations from the American Academy of Pediatrics (AAP), National Institute for Health and Care Excellence (NICE), and Turkish Neonatal Society (TND) guidelines. A two-round expert panel of eight specialists (four neonatologists and four pediatricians) established content validity and clinical relevance. The final 6-item NFSQS (score range: 0–24) was applied to 15 long-form YouTube videos. Internal consistency, content, concurrent, and construct validity were evaluated using Cronbach's alpha, Content Validity Index (CVI), modified DISCERN (mDISCERN), Global Quality Scale (GQS), JAMA Benchmark Criteria, and independent E-GRADE classification.

Results: The NFSQS achieved an S-CVI/Ave of 0.979 and a Cronbach's alpha of 0.846. The mean total NFSQS score was 11.1 ± 4.9 (range: 4-20). Significant positive correlations were observed with mDISCERN ($r = 0.670$, $p = 0.006$), GQS ($r = 0.575$, $p = 0.025$), and JAMA Benchmark Criteria ($r = 0.573$, $p = 0.026$). NFSQS scores correlated strongly with E-GRADE categories (Spearman's $\varrho=0.718$, $p=0.003$) and differed significantly across categories (Kruskal-Wallis $H=8.05$, $p=0.018$). No video achieved the highest guideline-adherence category.

Conclusion: NFSQS demonstrated excellent content validity, high internal consistency, concurrent validity with established video-quality instruments, and strong construct validity.It may provide a practical framework for identifying clinically safe, guideline-consistent online information on neonatal fever.

Keywords: Neonatal fever; YouTube; Health information quality; Content validity; Digital health; Video assessment

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

INTRODUCTION

Video-sharing platforms have become an important source of health information for patients and caregivers (1). YouTube is one of the most widely used platforms for obtaining medical information and contains a large volume of content related to child health and parenting (1-4). Although online videos may improve access to health information, the accuracy and quality of the information presented are variable because uploaded content is not routinely subjected to scientific review or professional oversight (1,5).

The neonatal period, defined as the first 28 days of life, is associated with increased susceptibility to serious infections because of the immaturity of the neonatal immune system (6) . Unlike older children, fever in a neonate is considered a potentially serious clinical finding that requires prompt medical evaluation (7,8). Neonatal fever may be the initial or only manifestation of severe bacterial infections such as sepsis, meningitis, and urinary tract infections (8-10) .

Current clinical guidelines recommend a cautious approach to the evaluation of febrile neonates. The American Academy of Pediatrics (AAP) and the National Institute for Health and Care Excellence (NICE) define a temperature of ≥38.0°C as clinically significant in young infants and recommend timely medical assessment for possible serious bacterial infection (8,9) . Early recognition and appropriate management are essential to reduce morbidity and potential complications associated with delayed diagnosis (11,12).

Despite the availability of evidence-based guidelines, health information shared on social media platforms may not always reflect current recommendations (1,2,5). In particular, information intended for older infants and children may be generalized to newborns, despite important differences in the evaluation and management of fever during the neonatal period (9,12). As a result, caregivers may encounter inconsistent recommendations regarding fever measurement, home care and monitoring practices medication use, and indications for seeking medical care.

Several instruments have been developed to assess the quality of online health information, including the DISCERN score, Global Quality Scale (GQS), and JAMA Benchmark Criteria (13-15). These tools evaluate characteristics such as reliability, transparency, authorship, and overall quality (13-15). However, they do not specifically assess whether content conveys clinically important safety messages regarding neonatal fever, such as appropriate temperature measurement, recognition of warning signs, timely healthcare seeking, or avoidance of potentially inappropriate home interventions and medication use before medical evaluation. Consequently, videos may achieve acceptable general quality ratings while still omitting or misrepresenting critical recommendations for neonatal care (16).

Given that neonatal fever may represent a medical emergency requiring prompt evaluation, inaccurate or incomplete online information has the potential to influence parental decision-making and delay appropriate medical assessment. Therefore, there is a need for condition-specific evaluation tools that assess not only the general quality of online content but also its consistency with clinically important and guideline-based recommendations. The aim of this study was to develop and validate the Newborn Fever Safety and Quality Score (NFSQS), a content-specific instrument designed to evaluate the quality, clinical safety, and guideline adherence of YouTube videos related to neonatal fever.

MATERIALS AND METHODS

Study Design and Video Selection

This cross-sectional study was conducted to develop and validate the Newborn Fever Safety and Quality Score (NFSQS) for the evaluation of YouTube videos related to neonatal fever. A structured YouTube search was performed using the Turkish keywords “Bebeklerde Ateş”, “Yenidoğan Bebeklerde Ateş”, and “Bebeklerde Ateş Nasıl Düşürülür”. To minimize the influence of previous search activity and personalized recommendations, all searches were conducted in a web browser operating in incognito mode with cleared browsing data and disabled location services. Videos were screened for eligibility and selected according to predefined inclusion and exclusion criteria. Only Turkish-language videos intended for parental education regarding neonatal fever were considered for analysis.

A total of 122 videos were initially identified through the predefined YouTube search strategy. Of these, 42 videos with a duration of less than one minute were excluded. An additional 36 videos were excluded because they primarily focused on fever management in older infants and children or on general fever-reduction practices rather than neonatal fever. Furthermore, 29 videos were excluded because neonatal fever was not specifically addressed. Following application of the eligibility criteria,

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

15 Turkish-language long-form videos specifically addressing neonatal fever remained and were included in the final analysis (Figure 1).

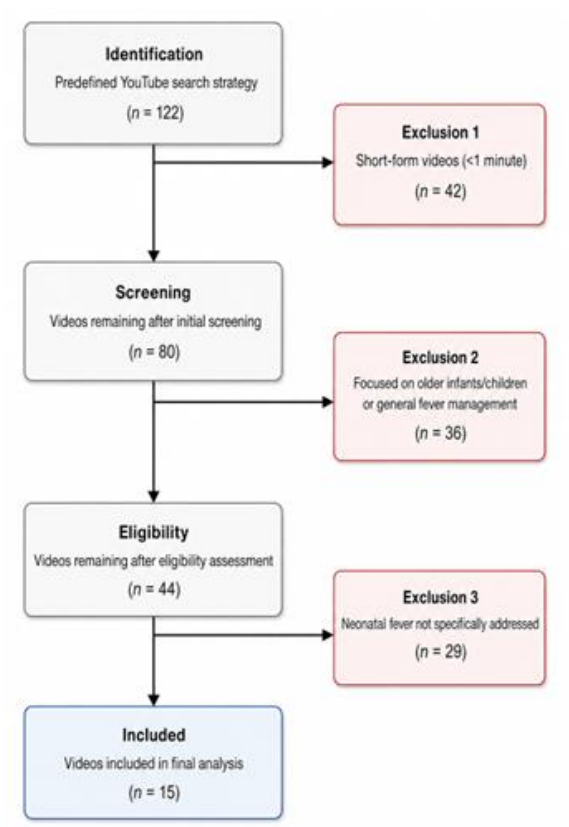


Figure 1. Flow diagram of the video selection process.

Clinical Content Item Development

An initial pool of 18 candidate items was developed based on commonly searched parental questions regarding neonatal fever and recommendations from the American Academy of Pediatrics (AAP, 2021), National Institute for Health and Care Excellence (NICE NG145), and Turkish Neonatal Society (TND) guidelines (8,9,17).

The candidate items were designed to evaluate clinically relevant aspects of neonatal fever recognition, temperature measurement, warning signs, healthcare-seeking behavior, home care and monitoring practices, medication safety, and adherence to guideline-based recommendations. Items Q1–Q6 addressed core recommendations directly related to neonatal fever assessment and management, whereas items Q7–Q18 evaluated additional topics including environmental factors, clothing practices, and other aspects of infant care.

All candidate items were reviewed by the expert panel to determine their relevance and suitability for inclusion in a neonatal fever-specific video evaluation instrument.

Expert Panel Review and Content Validity Assessment

A structured two-round expert panel process was conducted to establish content validity and refine the final version of the Newborn Fever Safety and Quality Score (NFSQS). The expert panel consisted of eight physicians, including four neonatologists and four pediatricians with clinical experience in the assessment and management of febrile neonates. All panel members independently reviewed the candidate items. During the first round, item relevance was evaluated using a four-point rating scale and Item-Level Content Validity Index (I-CVI) values were calculated.

Round 1: Content Validity Assessment

During the first round of expert review, panel members evaluated the relevance of each candidate item using a four-point scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant) (18). Item-level Content Validity Index (I-CVI) values were calculated according to the established methods(18). For an eight-member expert panel, an I-CVI value of ≥0.78 was considered acceptable for item retention.

Seven items (Q11, Q13, Q14, Q15, Q16, Q17, and Q18) did not meet the predefined retention criterion and were excluded from further evaluation. These items primarily addressed laboratory investigations and hospital-based diagnostic procedures that were considered less applicable to parental education and decision-making in the home setting.

Round 2: Clinical Relevance Assessment

The remaining 11 items underwent a second round of expert review focusing on their clinical relevance to neonatal fever education and parental decision-making. Following independent expert evaluation, five additional items (Q7, Q8, Q9, Q10, and Q12) were excluded because they were considered to have limited clinical relevance to parental education and decision-making regarding neonatal fever. At the completion of the expert review process, six items were retained for inclusion in the final version of the Newborn Fever Safety and Quality Score (NFSQS).

Validation Cohort and Reference Instruments

The final version of the NFSQS was applied to a validation cohort consisting of 15 Turkish-language YouTube videos

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

related to neonatal fever. Only long-form Turkish-language educational videos with a duration exceeding one minute were included. All videos were independently evaluated by two board-certified neonatologists experienced in neonatal infectious diseases and newborn care. Each reviewer assessed the videos using the NFSQS, modified DISCERN (mDISCERN), Global Quality Scale (GQS), JAMA Benchmark Criteria, and E-GRADE classification. Disagreements were resolved through consensus discussion.

To evaluate concurrent validity, NFSQS scores were compared with three widely used instruments for assessing the quality of online health information: the modified DISCERN score (mDISCERN), the Global Quality Scale (GQS), and the JAMA Benchmark Criteria (13-15). The modified DISCERN score evaluates the reliability and quality of health information presented in online resources (13). The Global Quality Scale assesses the overall quality, flow, and usefulness of educational content, whereas the JAMA Benchmark Criteria evaluate transparency through authorship, attribution, disclosure, and currency (14,15).

Videos were also independently categorized into E-GRADE levels according to their adherence to guideline-based neonatal fever management recommendations (16). The classification framework was adapted from the Evidence-GRADE (E-GRADE) methodology described by Kang et al. and modified for neonatal fever-specific educational content. Videos were evaluated according to the accuracy, completeness, and guideline concordance of recommendations regarding fever definition, warning signs, home care and monitoring practices, medication safety, and healthcare-seeking behavior.

Grade A represented comprehensive, guideline-consistent recommendations without clinically relevant inaccuracies(16). Grade B represented generally accurate content with minor omissions but no clinically significant inaccuracies(16). Grade C was assigned to videos containing substantial omissions or partially inaccurate recommendations (16). Grade D represented content containing recommendations inconsistent with current neonatal fever management guidelines or potentially unsafe advice(16). Inter-rater reliability was assessed using Cohen's kappa coefficient, weighted kappa coefficient, and exact agreement.

Statistical analysis

Statistical analyses were performed using Python version 3.10 (Python Software Foundation, Wilmington, DE, USA) with the pandas (v1.5.3), NumPy (v1.23.5), and SciPy (v1.10.1) libraries. Content validity was assessed using the Item-Level Content Validity Index (I-CVI) and Scale-Level Content Validity Index (S-CVI/Ave) (18). Internal consistency of the final instrument was evaluated using Cronbach's alpha coefficient. An I-CVI value of ≥0.78 and an S-CVI/Ave value of ≥0.90 were considered indicative of acceptable content validity.

Concurrent validity was evaluated using Pearson correlation coefficients between NFSQS total scores and the modified DISCERN score (mDISCERN), Global Quality Scale (GQS), and JAMA Benchmark Criteria. Associations between NFSQS scores and E-GRADE categories were evaluated using Spearman rank correlation analysis. Differences in NFSQS scores across E-GRADE categories were assessed using the Kruskal-Wallis test. Continuous variables are presented as mean ± standard deviation (SD) or median (minimum–maximum), as appropriate. A two-sided p value of <0.05 was considered statistically significant.

RESULTS

Expert Panel Evaluation

The expert panel evaluated the initial pool of 18 candidate items. Six items were retained in the final instrument, whereas 12 items were excluded following content



Figure 2. Corre Heatmap illustrating Pearson correlation coefficients among the six NFSQS domains and the total NFSQS score. Darker colors indicate stronger positive correlations. The strongest correlations with the total score were observed for Etiology, Warning Signs, Home, Care and Monitoring whereas Fever Definition demonstrated the weakest correlation.lation matrix of NFSQS domains and total score.

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

Table 2. NFSQS Scoring Subdomains

Domain	Score 0	Score 1 (Incomplete + Incorrect)	Score 2 (Comprehensive but inaccurate)	Score 3 (Incomplete + Correct)	Score 4 (Comprehensive + Correct)
Fever Definition and Measurement	No fever definition provided	Incomplete or incorrect fever definition	Comprehensive but inaccurate description of fever measurement methods	Correct fever definition with incomplete explanation of measurement methods	Comprehensive and guideline-consistent description of neonatal fever definition and temperature measurement
Etiology	No discussion of causes	Inaccurate explanation of causes of neonatal fever	Comprehensive but partially inaccurate discussion of causes	Correct discussion of causes with limited detail	Comprehensive and guideline-consistent discussion of potential causes, including serious bacterial infection
Warning Signs	No warning signs described	Important warning signs omitted or incorrectly described	Comprehensive but partially inaccurate discussion of warning signs	Correct discussion of warning signs with limited detail	Comprehensive description of clinically important warning signs requiring medical evaluation
Healthcare Seeking Behavior	No recommendation regarding medical evaluation	Inappropriate recommendations regarding healthcare seeking	Comprehensive but partially inaccurate recommendations	Correct recommendation with limited detail	Comprehensive and guideline-consistent recommendations regarding medical evaluation and emergency care
Home Care and Monitoring	No home care or monitoring advice provided	Inappropriate or potentially unsafe home care or monitoring recommendations	Detailed but partially inaccurate home care or monitoring recommendations	Correct home care and monitoring recommendations with limited detail	Comprehensive and guideline-consistent non-pharmacological home care and monitoring recommendations (e.g., monitoring, appropriate temperature measurement, and avoiding delays in seeking medical care)
Medication Safety	No information regarding medication administration before medical evaluation	Incorrect or potentially unsafe advice regarding medication use before medical evaluation	Detailed but inaccurate recommendations regarding medication use before medical evaluation	Correct recommendation to avoid medication administration before medical evaluation, but with limited explanation	Comprehensive, guideline-consistent recommendation advising against medication administration before medical evaluation, including appropriate rationale and safety considerations

NFSQS: Newborn Fever Safety and Quality Score.

validity and clinical relevance assessments. The retained items achieved I-CVI values ranging from 0.875 to 1.000, whereas excluded items demonstrated I-CVI values ranging from 0.250 to 0.750 (Table 1). The final six-item NFSQS achieved an S-CVI/Ave of 0.979, indicating excellent content validity of the scale (Table 2).

Internal Consistency and Score Distribution

The validation cohort consisted of 15 Turkish-language YouTube videos related to neonatal fever. Inter-rater agreement across all video assessment instruments was excellent, with an overall Cohen's kappa coefficient of 0.987, weighted kappa coefficient of 0.994, and exact agreement of 99.2%, demonstrating near-perfect consistency between reviewers. The mean NFSQS total score was 11.1 ± 4.9 (range: 4-20). Considerable variability in NFSQS scores was observed across videos (Table 3). Internal consistency of the six-item instrument was good, with a Cronbach's alpha coefficient of 0.846 (Figure 2). The item-total correlation analysis demonstrated positive correlations between all domains and the total NFSQS score. The strongest correlations were observed for N_Etiology(r = 0.86), N_Warning Signs(r = 0.83), N_Home Care and Monitoring (r = 0.88), N_Healthcare Seking Behavior (r = 0.72), and N_Medication Safety (r = 0.79). N_Fever Definition demonstrated the lowest correlation with the total score (r = 0.41) (Figure 2).

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

Table 3. Characteristics and Quality Scores of Evaluated YouTube Videos.

Video ID	Author Type	NFSQS	mDISCERN	GQS	JAMA	E-GRADE
V1	Medical Expert	11	3	3	2	C
V2	Allied Health Professional	4	3	2	1	D
V3	Medical Expert	12	3	3	2	C
V4	Medical Expert	13	3	2	2	C
V5	Medical Expert	6	1	3	1	D
V6	Medical Expert	13	3	3	2	C
V7	Allied Health Professional	8	2	3	2	C
V8	Medical Expert	7	2	3	2	D
V9	Medical Expert	6	3	3	2	C
V10	Allied Health Professional	10	3	3	2	D
V11	Medical Expert	20	4	4	2	B
V12	Allied Health Professional	13	2	3	2	D
V13	Independent Content Creator	7	3	3	1	D
V14	Medical Expert	18	4	4	2	B
V15	Medical Expert	18	4	3	2	B

NFSQS: Newborn Fever Safety and Quality Score; GQS: Global Quality Scale; JAMA: JAMA Benchmark Criteria. E-GRADE categories were assigned independently according to adherence to guideline-based neonatal fever management recommendations

Concurrent Validity

The NFSQS demonstrated significant positive correlations with all reference instruments, including the modified DISCERN score (r = 0.670, p = 0.006), Global Quality Scale (GQS) (r = 0.575, p = 0.025), and JAMA Benchmark Criteria (r = 0.573, p = 0.026), supporting the concurrent validity of the instrument (Figure 3).

Classification of Videos According to E-GRADE Categories

Videos were independently classified according to E-GRADE categories based on their adherence to guideline-based neonatal fever management recommendations. No videos met the criteria for Grade A classification. Three videos were classified as Grade B, six as Grade C, and six as Grade D. A strong positive correlation was observed between NFSQS total scores and E-GRADE categories (Spearman's $\rho = 0.718$, $p = 0.003$). Higher NFSQS scores were consistently associated with higher levels of guideline adherence. Furthermore, NFSQS scores differed significantly across E-GRADE categories (Kruskal-Wallis $H = 8.05$, $p = 0.018$), indicating that the instrument successfully discriminated between videos with varying levels of adherence to evidence-based neonatal fever management recommendations (Figure 4).

DISCUSSION

In this study, we developed and validated the Newborn Fever Safety and Quality Score (NFSQS), a content-specific instrument for assessing the quality, clinical safety, and guideline adherence of YouTube videos related to neonatal fever. The NFSQS demonstrated excellent content validity, good internal consistency, strong inter-

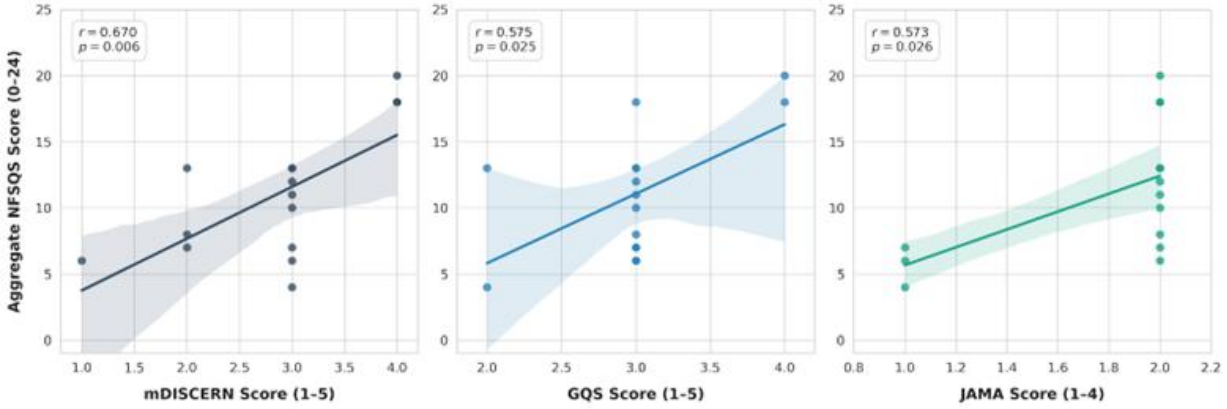


Figure 3. Correlations between NFSQS scores and established video quality assessment instruments. Scatter plots with fitted linear regression lines and 95% confidence intervals demonstrating the relationships between NFSQS total scores and reference instrument scores across the evaluated YouTube videos (n = 15). (A) Correlation between NFSQS and modified DISCERN scores (r = 0.670, p = 0.006). (B) Correlation between NFSQS and Global Quality Scale (GQS) scores (r = 0.575, p = 0.025). (C) Correlation between NFSQS and JAMA Benchmark Criteria scores (r = 0.573, p = 0.026). NFSQS: Newborn Fever Safety and Quality Score; GQS: Global Quality Scale.

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

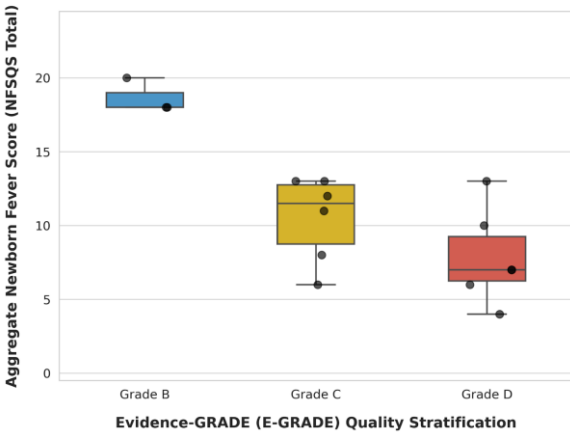


Figure 4. NFSQS scores according to independently assigned E-GRADE categories**.** Boxplots display median values, interquartile ranges, and individual video scores. Higher E-GRADE categories were associated with significantly higher NFSQS scores (Kruskal-Wallis H = 8.05, p = 0.018). A strong positive correlation was observed between NFSQS scores and E-GRADE levels (Spearman's ϱ = 0.718, p = 0.003).

rater reliability, and significant correlations with established video quality assessment instruments. Furthermore, its association with independently assigned E-GRADE categories supports its construct validity. These findings suggest that the NFSQS is a reliable and reproducible tool for evaluating neonatal fever–related online educational content.

The increasing use of social media platforms as sources of health information has raised concerns regarding the quality and accuracy of online medical content. Neonatal fever represents a particularly important topic because management recommendations for newborns differ substantially from those for older infants and children. Consequently, inaccurate or incomplete information may have important implications for parental decision-making and healthcare-seeking behavior.

Several instruments have previously been used to evaluate online health information, including the DISCERN score, JAMA Benchmark Criteria, and Global Quality Scale (13-15). These instruments provide valuable information regarding reliability, transparency, and overall content quality. However, they were not specifically designed to assess adherence to guideline-based recommendations for neonatal fever management. Similar observations have been reported in previous studies evaluating online medical information. Kang et al. reported discrepancies between the quality of online medical content and the level of supporting evidence, suggesting that domain-specific assessment tools may provide additional information beyond general quality measures (16,19). Previous studies have similarly demonstrated that source quality indicators do not necessarily reflect clinical accuracy in health-related video content (20).

Furthermore, NFSQS scores demonstrated a strong positive correlation with independently assigned E-GRADE categories, providing additional support for the construct validity of the instrument. Videos with higher levels of guideline adherence consistently achieved higher NFSQS scores. In the present study, videos with lower NFSQS scores more frequently contained recommendations that were inconsistent with current neonatal fever management guidelines, including home medication use before medical evaluation and the use of traditional home remedies. These findings suggest that general quality indicators alone may not fully capture clinically relevant aspects of neonatal fever information and support the need for condition-specific evaluation tools.

The NFSQS demonstrated significant correlations with mDISCERN, GQS, and JAMA Benchmark Criteria, indicating that the instrument measures constructs related to overall content quality while simultaneously focusing on neonatal fever-specific recommendations. In addition, the Cronbach's alpha coefficient of 0.846 indicated good internal consistency among the six domains included in the final instrument. The S-CVI/Ave value of 0.979 further supports the content validity of the scale.

The relatively lower item-total correlation observed for the Fever Definition and Measurement domain suggests that this content area may represent a specific weakness in neonatal fever education videos. Although many videos addressed broader topics such as warning signs, home care, healthcare-seeking behavior, and medication safety, they were less consistent in clearly defining fever and explaining appropriate temperature assessment. This is clinically important because neonatal fever is typically defined as a rectal temperature of ≥38.0°C**,** and failure to communicate this threshold or the recommended method of measurement may lead caregivers to underestimate the seriousness of fever in early infancy. Inaccurate or incomplete guidance may delay timely medical evaluation, particularly during a period in which fever can be the only early sign of serious bacterial or viral infection. Therefore, future videos should place greater emphasis on standardized fever definitions, accurate temperature measurement techniques, and clear instructions about when urgent medical care is warranted.

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

The E-GRADE classification system provided a practical framework for categorizing videos according to their consistency with guideline-based recommendations. The identification of videos with very low NFSQS scores despite acceptable presentation quality highlights the importance of evaluating both content quality and clinical accuracy when assessing online health information. This finding further supports the value of condition-specific assessment tools in digital health research. The scarcity of neonatal fever-specific educational videos identified during the screening process underscores the need for high-quality, evidence-based digital resources targeted specifically toward parents of newborns.

Notably, no video achieved Grade A classification, indicating the absence of fully guideline-concordant neonatal fever educational content within the evaluated sample. Given that neonatal fever may represent a medical emergency requiring prompt assessment, the lack of comprehensive and clinically accurate educational videos raises important concerns regarding parental access to reliable online information. This finding highlights a critical gap in digital health education and underscores the need for greater involvement of healthcare professionals and professional organizations in the development of evidence-based neonatal educational content.

Only 15 of 122 videos identified through the predefined search strategy met the eligibility criteria for neonatal fever-specific educational content. This finding suggests that caregivers searching for neonatal fever information on YouTube may frequently encounter videos focused on older infants, children, or general fever management rather than neonatal fever specifically. The limited availability of neonatal fever-specific content highlights an important gap in digital health education and underscores the need for high-quality, evidence-based resources targeted toward parents of newborns. Although internal consistency was high, reliability estimates should be interpreted cautiously because the validation cohort was relatively small.

Limitations

This study has several limitations. First, the video sample represents a single time point, and YouTube search results may change over time, potentially affecting reproducibility. Second, short-form video content ("Shorts") was not included in the analysis; therefore, the findings cannot be generalized to this format. Third, although only 15 videos were included, all eligible Turkish-language long-form videos specifically addressing neonatal fever identified through the predefined search strategy were incorporated into the analysis. The relatively small number of eligible videos highlights the limited availability of neonatal fever-specific educational content on Turkish-language YouTube. Finally, only Turkish-language videos were assessed; therefore, additional studies are required to determine the applicability of the NFSQS across across different languages and cultural settings.

CONCLUSION

The Newborn Fever Safety and Quality Score (NFSQS) is a valid and reliable instrument for evaluating YouTube videos related to neonatal fever. The scale demonstrated excellent content validity, high internal consistency, significant concurrent validity with established video quality instruments, and strong construct validity through its association with independent guideline-adherence classifications.

Unlike general quality assessment instruments, the NFSQS focuses specifically on recommendations related to neonatal fever recognition, clinical safety, and healthcare-seeking behavior. The instrument may provide a practical and standardized approach for assessing the quality and guideline adherence of online neonatal fever information.

Importantly, no video in the validation cohort achieved Grade A classification, highlighting important gaps in the availability of comprehensive, evidence-based educational content on neonatal fever for parents and caregivers. Further studies involving larger and more diverse video samples, cross-cultural adaptation, and validation in different languages are warranted to confirm the applicability and generalizability of the NFSQS across diverse digital health environments.

Acknowledgments

The authors thank the members of the expert panel for their participation in the content validity and expert review process.

Authorship contributions

EK and IMH conceived and designed the study. They developed the Newborn Fever Safety and Quality Score (NFSQS), conducted the expert validation process, independently evaluated the YouTube videos, analyzed and interpreted the data, and drafted and revised the manuscript. Both authors approved the final version of the manuscript.

MMDJ
MEANDROS MEDICAL AND DENTAL JOURNAL
September 2026 27(3):430-438
Meandros Medical and Dental Journal
doi: 10.69601/meandrosmdj.1967737

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request. The evaluated YouTube videos are publicly accessible online. https://www.youtube.com/watch?v=KiSc8jQ9RlY&list=PLA80WzQfdSJA-yMJKwm-ImdEcUcewpolm

Declaration of competing interest

The authors declare no competing interests.

Ethics

This study analyzed publicly available YouTube video content and did not involve human participants, patient data, or animal subjects. No personal identifying information was collected or recorded. In accordance with national regulations and institutional policies, studies using publicly available online content without human participant involvement do not require formal ethics committee approval. information was considered a limitation of the study. Ethical approval for the study was granted by the local committee (reference 1775; September 18, 2024), and the work was undertaken in alignment with the Declaration of Helsinki.

Funding

None

REFERENCES

1. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC medical education 2022; 22: 382.
2. Drozd B, Couvillon E, Suarez A. Medical YouTube Videos and Methods of Evaluation: Literature Review. JMIR Medical Education 2018; 4.
3. Gabarron E, Fernandez-Luque L, Armayones M, Lau AYS. Identifying Measures Used for Assessing Quality of YouTube Videos with Patient Health Information: A Review of Current Literature. Interactive Journal of Medical Research 2013; 2.
4. Xu D, Li Y, Li S, Hu H, Wei S, Ma Q, et al. Quality and guideline concordance of neonatal jaundice information on short-video platforms: implications for parental education and newborn safety. BMC Pediatrics 2026; 26.
5. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics Journal 2014; 21: 173-94.
6. Cuenca AG, Wynn JL, Moldawer LL, Levy O. Role of Innate Immunity in Neonatal Infection. Am J Perinatol 2013; 30: 105-12.
7. Pantell RH, Newman TB, Bernzweig J, Bergman DA, Takayama JI, Segal M, et al. Management and Outcomes of Care of Fever in Early Infancy. JAMA 2004; 291: 1203-12.
8. Pantell RH, Roberts KB, Adams WG, Dreyer BP, Kuppermann N, O'Leary ST, et al. Evaluation and Management of Well-Appearing Febrile Infants 8 to 60 Days Old. Pediatrics 2021; 148.
9. Paul SP, Khattak H, Kini PK, Heaton PA, Goel N. NICE guideline review: neonatal infection: antibiotics for prevention and treatment (NG195). Archives of disease in childhood - Education & practice edition 2022; 107: 292-7.
10. Shah BA, Padbury JF. Neonatal sepsis. Virulence 2014; 5: 170-8.
11. Celik IH, Hanna M, Canpolat FE, Mohan P. Diagnosis of neonatal sepsis: the past, present and future. Pediatric Research 2022; 91: 337-50.
12. Leazer RC. Evaluation and Management of Young Febrile Infants: An Overview of the New AAP Guideline. Pediatrics In Review 2023; 44: 127-38.
13. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. Journal of Epidemiology and Community Health 1999; 53: 105-11.
14. Bernard A, Langille M, Hughes S, Rose C, Leddin D, van Zanten SV. A Systematic Review of Patient Inflammatory Bowel Disease Information Resources on the World Wide Web. Official journal of the American College of Gastroenterology | ACG 2007; 102: 2070-7.
15. Silberg WM, Lundberg GD, Musacchio RA. Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet: Caveant Lector et Viewor—Let the Reader and Viewer Beware. JAMA 1997; 277: 1244-5.
16. Kang E, Lee H, Choi J, Ju H. The Quality of Evidence of and Engagement With Video Medical Claims. JAMA Network Open 2026; 9.
17. Satar M, Arısoy AE, Çelik İH. Turkish Neonatal Society guideline on neonatal infections-diagnosis and treatment. Turkish Archives of Pediatrics/Türk Pediatri Arşivi 2018; 53: S88.
18. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health 2007; 30: 459-67.
19. Guler MA, Aydın EO. Development and validation of a tool for evaluating YouTube-based medical videos. Irish Journal of Medical Science (1971 -) 2021; 191: 1985-90.
20. McNamee LA, Shakartzi H, Wasser TE, Li Y, Kim C-S. Assessing the accuracy and quality of YouTube videos on early pregnancy loss. Heliyon 2022; 8.