

Article

Evaluation of Emergency Social Media Language Efficiency Based on Persuasion Theory and Data Envelopment Analysis: A Case Study of the 2025 Beijing Extreme Rainfall Event

Jingqi Gao ^{1,2} , Yutong Zu ^{1,*} , Shigen Fu ² , Jianwu Chen ³ , Shufang Li ² and Hezhuang Lou ⁴ ¹ School of Engineering and Technology, China University of Geosciences (Beijing), Beijing 100083, China; gjq_0421@163.com² China Academy of Safety Science and Technology, Beijing 100012, China; lishufang0616@163.com (S.L.)³ School of Resources and Safety Engineering, University of Science and Technology, Beijing 100083, China⁴ Beijing Drainage Group Co., Ltd., Beijing 10004, China; qililhz0405@126.com

* Correspondence: 3002220025@email.cugb.edu.cn

Abstract

In the context of urban extreme weather events, the efficacy of the “emergency language” employed by governments and public institutions on social media in effectively reaching and guiding the public in a timely manner necessitates a quantifiable evaluation framework. An indicator system was constructed on the basis of Hovland’s persuasion theory. This system comprised five input characteristics (word count/structural clarity, first/second-person perspective, emotional appeal, evidence and framing, and media format) along with three output indicators (reposts, comments, and likes). A data envelopment analysis (DEA) model that is oriented towards output was employed, with disseminators being categorized into four distinct decision-making units: central mainstream media, other government media, local government media, and other media. It is imperative to note that the outputs were subjected to a process of normalization through the implementation of a scale factor. The data were sourced from the Weibo platform within the specified time window, which was from 10:00 on 24 July 2025, to 12:00 on 19 August 2025, with a sample size of 744. The findings revealed substantial disparities in technical efficiency across different disseminator types. A subset of local government media demonstrated a technical efficiency ≈ 1.00 yet low scale efficiency. Posts exhibiting clear structures, actionable points, and accompanying images or videos achieved higher cross-efficiency scores. It is therefore evident that the proposed DEA model provides a benchmark for maximizing dissemination effectiveness under given information characteristics. It is recommended that posting frequencies be maintained at consistent intervals during periods of heightened activity, that a template structure be adopted in accordance with the “fact–action–assistance channel” model, and that the proportion of rich media content be augmented.

Keywords: persuasion theory; emergency language; data envelopment analysis (DEA)

Academic Editor: Rui Araújo

Received: 13 September 2025

Revised: 17 October 2025

Accepted: 24 October 2025

Published: 26 October 2025

Citation: Gao, J.; Zu, Y.; Fu, S.; Chen, J.; Li, S.; Lou, H. Evaluation of Emergency Social Media Language Efficiency Based on Persuasion Theory and Data Envelopment Analysis: A Case Study of the 2025 Beijing Extreme Rainfall Event. *Appl. Sci.* **2025**, *15*, 11435. <https://doi.org/10.3390/app152111435>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An Emergency Events Database (EM-DAT) report titled “2024 Disasters in Numbers” indicates that in 2024, 393 natural disasters worldwide resulted in 16,753 deaths, affected 167.2 million people, and caused total economic losses of USD 241.95 billion. Among these, meteorological disasters (floods, extreme temperatures, storms, etc.) had the most severe impact. Statistics from China’s Ministry of Emergency Management for 2024 also

revealed that various natural disasters affected 94.13 million people nationwide to varying degrees, resulting in 856 deaths or missing persons and requiring the emergency relocation of 3.645 million people. With the increasing frequency of extreme weather events in recent years, information dissemination has become critical in responding to highly hazardous and destructive natural disasters. Emergency language services, provided during crisis situations, serve as a vital means to mitigate societal risks and disaster impacts. Effective emergency language services help guide the public in overcoming panic, alleviating public opinion pressure, and reducing factors contributing to unrest. Consequently, these services play a critical role in emergency response and post-disaster relief operations. Existing research indicated that the United States has established a comprehensive “language assessment-emergency response-post-disaster evaluation” response mechanism [1]. However, emergency language service research in China remains in its infancy, and emergency language services for sudden public incidents require further in-depth study.

Data envelopment analysis (DEA) was proposed by Charnes et al. and is recognized as an efficient analytical tool for evaluating efficiency in multi-input, multi-output scenarios [2]. This method not only identifies improvement directions for each indicator but also quantifies specific improvement values to transform technologically inefficient decision-making units. Currently, the DEA method primarily encompasses three models: DEA-Charnes, Cooper, and Rhodes (DEA-CCR), DEA-Banker, Charnes, and Cooper (DEA-BCC), and Slacks-Based Measure-DEA (SBM-DEA). Among these, DEA-CCR is the inaugural DEA model, incorporating both technical and scale efficiency under constant returns to scale; The DEA-BCC model builds upon DEA-CCR by abandoning the constant returns to scale assumption, dividing technical efficiency into scale efficiency and pure technical efficiency; the SBM-DEA model extends traditional models by considering non-desirable outputs to evaluate the relative efficiency of decision units. Due to its broad applicability and relatively straightforward principles, the DEA method has seen extensive adoption in recent years across fields such as operations research and management science, extending into safety domains. Emrouznejad et al. also confirmed the applicability of DEA in social media performance evaluation. However, research utilizing DEA to explore the effectiveness of emergency communication protocols remains largely unexplored [3].

Hovland’s persuasion theory views persuasion and attitude change as a system, where persuaders (communicators), persuasive content, and persuasive context are considered key factors influencing attitude change [4]. In emergency response, information is not necessarily “effective simply by being disclosed.” Its impact depends on “who releases it, how it is articulated, and in what context it is presented.” Hovland’s persuasion theory provides precisely such an analytical perspective: persuasion is a systematic process driven by the communicator, the persuasive content, and the communication context. Therefore, this study used the 2025 Beijing Extreme Rainfall Event as a case study to analyze information disseminated via social media during the incident’s emergency response phase. Based on persuasion theory, this study constructed an evaluation model for emergency social media language systems by examining communicators and persuasive content. Employing the DEA-BCC model to compare the efficiency of input and output indicators across different decision-making units (communicators), it assesses the effectiveness of current emergency social media language services. Recommendations are proposed to enhance the efficiency of these services, aiming to provide theoretical support for improving the scientific rigor and effectiveness of emergency management language services.

2. Related Research and Theoretical Basis

Emergency language refers to linguistic forms used for communication and coordination during emergencies, characterized by high timeliness and specialized functions.

Emergency language services provide linguistic support and assistance to affected populations during crises, encompassing but not limited to translation, information dissemination, and psychological reassurance. The outbreak of the 2019 COVID pandemic has further increased the global attention to emergency languages and related services and prompted the academic community to carry out research from multiple perspectives.

The importance of emergency language service was particularly prominent during COVID-19. For example, Han et al. highlighted the importance of social media engagement during the COVID-19 public health emergency [5]. Moreover, Chang et al. analyzed telemedicine service usage and emergency department visits among non-elderly patients with limited English proficiency during COVID-19 [6]. Xie et al. employed the SERVQUAL model to investigate factors influencing the quality of emergency medical language services during the pandemic [7]. Through questionnaire collection and analysis, they concluded that such services require improvements and upgrades in organizational structure, talent development, and service channel expansion. Jason et al. focused on communication barriers faced by deaf sign language users in U.S. emergency medical services, evaluating the effectiveness of proposed training methods and visual communication tools for emergency medical practitioners [8]. Similarly, Jeconiah et al. highlighted the importance of cross-language communication in emergency language services during COVID-19 [9]. Zheng et al. investigated the application of emergency language services in Chinese border towns using Mengding, Yunnan, as a case study, providing reference for emergency language policies and planning in China's border provinces [10].

With the progress of technology, the application of emerging technologies in emergency language service has become a research hotspot. Guo et al. searched the Web of Science Core Collection using “emergency” and “language” as keywords, employing a literature review analysis to explore future trends in emergency language services [11]. They analyzed existing research keywords and noted a current research bias toward emerging technologies such as natural language processing, language modeling, and machine learning. Hailay et al. analyzed challenges faced by in-service language teachers transitioning to online learning platforms during COVID-19, offering recommendations for enhancing language teachers' digital literacy and effective online learning in the post-pandemic era [12]. Enze et al. explored the application of natural language processing methods in automated electronic health record documentation, investigating and developing a large language model framework to address automatic transcription and structured data extraction challenges [13]. Noack et al. engaged users in developing digital communication tools to overcome language barriers in emergency medical services [14]. In summary, current research on emergency language services primarily focuses on emergency medical contexts, with limited studies employing quantitative methods to examine such services during sudden events or workplace safety incidents.

In general, the current research on emergency language service mainly focuses on the emergency medical background and lacks quantitative methods to analyze the service efficiency in emergencies or workplace safety incidents. This limitation has prompted researchers to turn to other methods, such as data envelopment analysis (DEA), to provide more objective evaluation. For example, Leila et al. employed DEA to identify and analyze key factors and sub-factors influencing employees' perceptions of safety performance [15]. Suh employed DEA to calculate and evaluate productivity and safety effectiveness while analyzing input–output patterns [16]. In the construction industry, Qi et al. treated annual safety performance across regions as decision-making units (DMUs), using DEA to assess safety performance in three Chinese regions and noting that the construction industry performs better in preventing workplace fatalities than in preventing non-fatal accidents [17]. Nahangi et al. employed a DEA-CCR model to analyze safety efficiency across

112 construction sites, identifying variables with the greatest impact on safety management efficiency [18]. Compared with regression analysis and structural equation modeling, in which a function form needs to be preset and the classification model only outputs high and low labels, DEA is hypothesis-free and processes multiple inputs and outputs at the same time, and the continuous efficiency value and improvement amount are more in line with the comparative logic of “equal language input–equal interactive output”.

In summary, the existing studies have not only clarified the core values and challenges of emergency language service in practice but also revealed the shortcomings of the current evaluation methods in objectivity and systematicness. The DEA method provides a potential solution for quantitatively evaluating the efficiency of emergency language service, with its unique advantages of no preset function form, simultaneous processing of multiple indicators, and provision of a specific improvement direction. Therefore, the construction of a DEA effectiveness evaluation model for emergency language service in this paper is not only an important supplement to the existing research methodology but also has significant theoretical and practical significance.

3. Method and Materials

3.1. Data Collection and Pro-Processing

Weibo is currently the only major Chinese platform that simultaneously provides public retweet links and user interaction data. Given WeChat’s data isolation and the low usage of Twitter/X in mainland China, this study selected Weibo as the data acquisition platform. The research employed Weibo’s official advanced search API (application programming interfaces) + polling crawling strategy, using keywords such as “Beijing heavy rain,” “emergency response,” “early warning,” “Miyun,” and “Huairou” (the hardest-hit areas). It collected Weibo posts published between 10:00 on 24 July 2025 (when the Beijing Meteorological Observatory issued an orange rainstorm warning) to 12:00 on 19 August 2025, during the emergency response phase as foundational data for emergency social media language services. Data preprocessing involved deduplication, ad filtering (regular expressions + manual sampling), retaining posts ≥ 10 Chinese characters, and limiting to government/official media verified accounts. This yielded 787 verified official Weibo posts related to the 2025 Beijing extreme rainfall event.

Text content underwent preprocessing using HIT’s stopword list to remove meaningless vocabulary before thematic or analytical research was conducted. For comment data, retweeted posts, # topics, retweets, follows, pure @ mentions, “Hello, your Super Language feature is now activated,” null values, pure numeric values, graphic symbols (microphone, heart, rose, candle, applause, etc.), and emojis were excluded as meaningless comments. This yielded 744 valid tweet data points, totaling approximately 165,000 characters.

Hovland’s persuasion theory points out that the communicator is the main body of persuasion activities and states that the higher the credibility of the persuader (communicator) is, the greater the persuasive effect. The credibility of the communicator depends on 2 key factors: the first is the level of trust the audience has in the communicator and the second is the level of the audience’s perception of his or her professional competence. The communicator’s credibility can influence the audience’s understanding and judgment of the message content, which in turn acts on the change of the audience’s attitude. With reference to the collected data, the communicators are divided into four categories of accounts: central mainstream media, other government media, local government media, and other media. In order to ensure the reliability of feature interpretation, a randomly selected 10~20% sample was independently reviewed by two people, and consistency indicators were reported. Inter-rater agreement was Cohen’s $\kappa = 0.82$ (95% CI, 0.78–0.86), with an overall percent agreement of 88%; category-specific κ values were all > 0.75 . Disagreements

were adjudicated by a third reviewer, yielding a post-adjudication increase of $\Delta\kappa = 0.03$, indicating reliable annotations and controlled measurement error.

3.2. *Establishing an Evaluation System for the Effectiveness of Emergency Social Media Language Services*

3.2.1. Selection of Input Indicators for Emergency Language Service Evaluation

Persuasive content is influenced by emotional factors and presentation methods and can be categorized into the following five aspects.

(1) Narrative Immersion. Immersion serves as a crucial psychological mechanism underpinning persuasive outcomes. It represents a psychological state or process integrating attention, imagery, and sensation, reflecting the audience's level of focus on the narrative text. This encompasses emotional engagement with the plot, cognitive attention, mental imagery, and a corresponding reduction in awareness of the real world. During narrative interventions, variations in audience immersion levels yield differing persuasive effects. In highly immersive contexts, audiences may be more inclined to transcend existing cognitive frameworks and past social experiences, thereby facilitating deeper contemplation of persuasive claims presented within the narrative. This, in turn, influences their beliefs, attitudes, and behavioral intentions. Conversely, in low-immersion scenarios, audiences are more likely to harbor skepticism toward the claims presented in the narrative.

Research indicates that narrative information can evoke higher levels of immersion among audiences and directly influence their risk perception [19,20]. Therefore, the presence or absence of narrative immersion is selected as an evaluation criterion for engagement.

(2) Narrative Perspective. Narrative perspectives primarily include first-person, third-person, and multiple perspectives. The choice of narrative perspective is generally considered to influence the audience's reading experience. When reading first-person narratives, audiences tend to adopt an internal perspective, perceiving life situations through the character's subjective experiences. Conversely, third-person narratives induce an external perspective, leading audiences to observe characters more objectively [21]. Research indicates that first-person perspectives are more effective than are third-person ones. Studying human papillomavirus, Nan et al. discovered that first-person narratives evoked higher risk perception than do third-person narratives [22]. Chen et al. found that first-person narratives fostered greater identification than did third-person narratives in diabetes type II research, thereby enhancing persuasive effectiveness [23]. Consequently, narrative perspective was selected as an engagement indicator.

(3) Information Framing. Information framing encompasses gain frames and loss frames. Frames with different characteristics elicit varying responses from audiences and lead to differences in their behavioral intentions. A gain frame refers to statements highlighting the benefits or gains resulting from taking action. Brusse et al. combined information framing theory with entertainment education strategies to create two animated videos: one presenting a gain frame for complying with the "no drunk driving" norm and another presenting a loss frame for non-compliance [24]. They found that audiences exposed to the gain-frame video exhibited relatively less resistance and a higher intention to avoid drunk driving compared to those exposed to the loss-frame video. Other research indicates that information framing can serve as a moderator influencing the relationship between message type and outcome variables. For instance, Liu et al. found that under gain framing, college students reading narrative versus non-narrative information showed no difference in empathy levels [25]. However, under loss framing, students reading narrative information exhibited higher empathy levels compared to those exposed to non-narrative information. Therefore, the selection of information framing type was adopted as an input indicator.

(4) Narrative Technique. Hovland’s research on the effectiveness of fear appeals revealed that fear-based messages in persuasive content can heighten audience attention and promote attitude change [26]. Therefore, the inclusion of fear-inducing narrative techniques served as an engagement indicator.

(5) Media Type. Media type primarily refers to the medium through which narrative information is conveyed or presented to audiences during communication, including text, video, audio, etc. Comparisons of persuasive effects across different media types indicate that audio/video narratives demonstrate significantly greater impact, whereas text-based narratives yield less pronounced effects. Consequently, narrative information delivered via audio and video exerts greater persuasive power on audiences than does that conveyed through text [27]. Therefore, the choice of whether to incorporate video as a medium type in official media’s online emergency text selection was adopted as an input indicator.

To accommodate the DEA constraint of minimizing input, narrative immersion (X1), third-person perspective (X2), and gain information framework (X3) were treated as positive features and converted into deficiency scores ($X_k^{\text{in}} = \max(X_k) - X_k$) during modeling. Fear emotion (X4) and pure text (X5) inherently require minimal input and were directly incorporated into the model as inputs.

The characteristics of the input indicators are summarized in Table 1.

Table 1. Characteristics of input indicators for the evaluation of emergency language services.

Input Indicator	Characteristic
Narrative immersion	Does it contain words related to characters, events, settings, time, or emotions?
Narrative perspective	Does it include personal pronouns such as “I”, “we”, “you”, or “you” (plural)?
Information framing	Are there sentences that describe benefits or gains resulting from taking action?
Narrative technique	Does it contain words related to fear?
Media type	Does it involve words related to video or audio?

3.2.2. Selection of Output Indicators for Emergency Language Service Evaluation

The output of official media’s emergency social media language effectiveness is directly reflected in user responses. When information disseminated by official media via Weibo reaches audiences through various channels, recipients first form preliminary perceptions by browsing the content. Based on their emotional judgments, they decide whether to like the post to express their stance. If audiences endorse the information and perceive it as meeting their needs, they may take further actions such as reposting or commenting. The three evaluation indicators, likes, shares, and comments, most directly reveal the effectiveness of information disseminated by official media. Therefore, the number of likes, shares, and comments were selected as output evaluation indicators for the effectiveness of emergency language services.

In summary, an emergency social media language service evaluation system grounded in persuasion theory was constructed, as illustrated in Figure 1. The primary indicators comprised three categories: communicator characteristics, input metrics, and output metrics. The secondary indicators included four categories (e.g., central mainstream media), five categories (e.g., narrative immersion), and three categories (e.g., comment count), totaling 12 indicators.

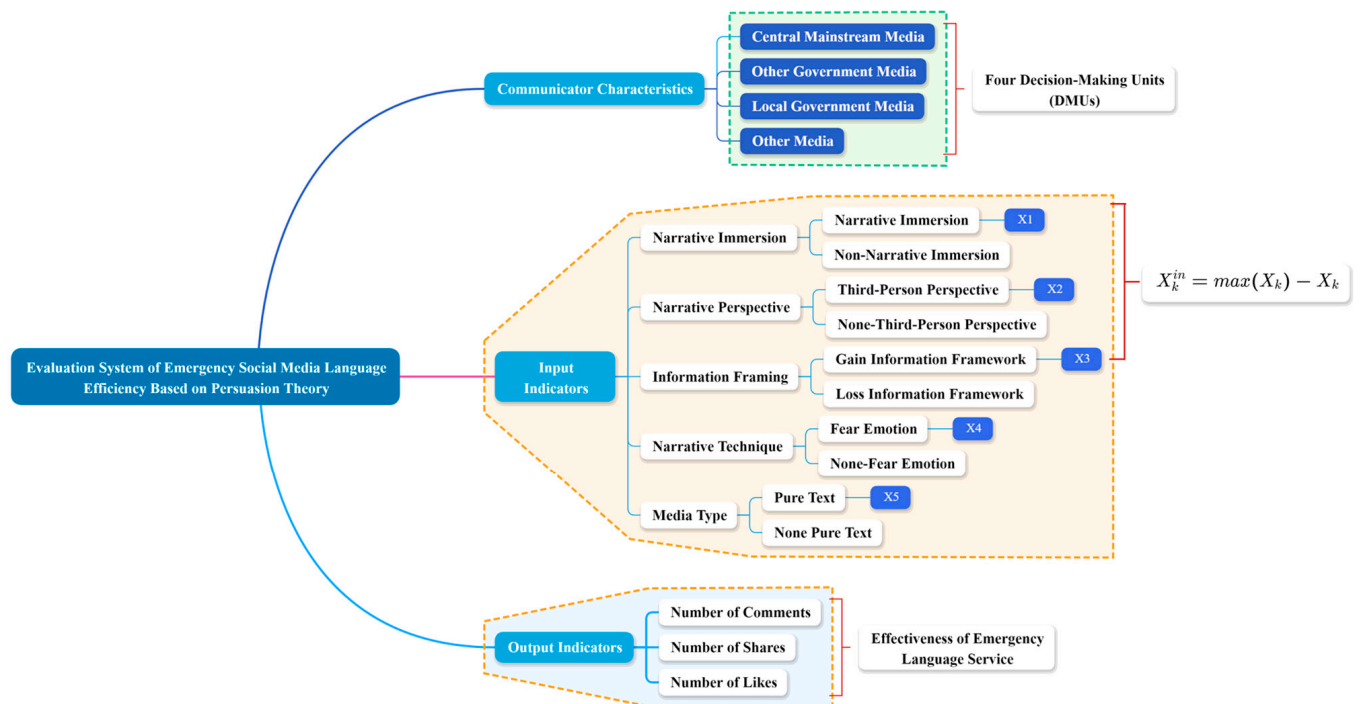


Figure 1. Index system for evaluating the effectiveness of emergency social media language services.

3.3. Procedure for Evaluating Effectiveness

The procedure for evaluating the effectiveness of emergency social media language services is illustrated in Figure 2.

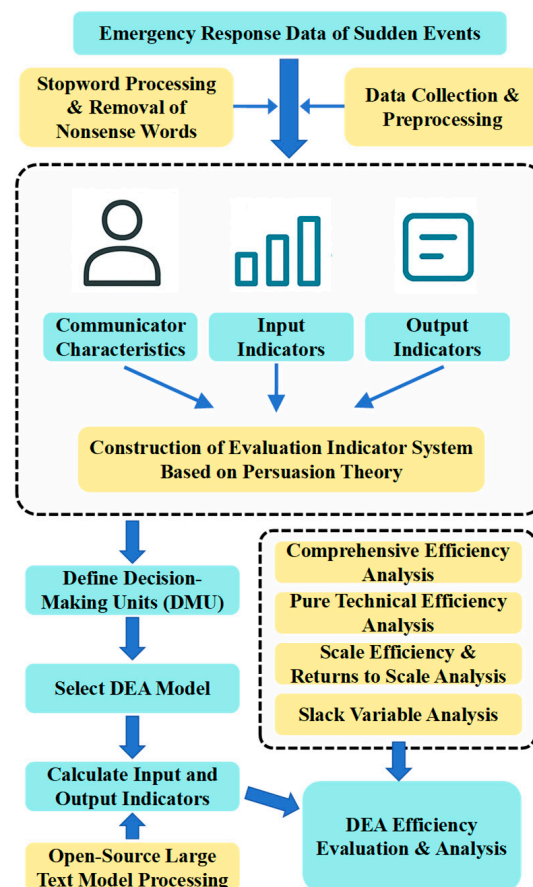


Figure 2. Evaluation process of the effectiveness of emergency social media language services.

The specific evaluation steps are as follows:

- (1) Identify data sources for emergency response phases of public safety incidents and collect data from official reports, social media, and other platforms using web crawlers, application programming interfaces (APIs), and other methods.
- (2) Conduct data preprocessing; specifically, clean data by removing duplicates and irrelevant information, verify and correct data with formatting or spelling errors, and apply a stopword list for text filtering to enhance data quality.
- (3) Establish an effectiveness evaluation framework by defining input and output indicators and adopting manual coding and other verification procedures to examine indicator reliability.
- (4) Conduct empirical analysis to identify decision units, select appropriate DEA models, and statistically analyze input and output metrics.
- (5) Conduct DEA efficiency analysis using linear programming solvers (e.g., Lingo, MATLAB 2024a) to solve the DEA model. Calculate the efficiency value θ for each decision unit, where efficiency value $\theta = 1$ indicates DEA efficiency, and θ value < 1 indicates efficiency loss. Additionally, compute and analyze the slack variables for each decision unit.
- (6) Analyze DEA efficiency results to determine whether decision units are efficient. Investigate causes of efficiency losses in inefficient units and propose targeted improvement recommendations based on DEA findings.

Through these steps, a systematic evaluation of the effectiveness of emergency social media language services was achieved.

3.4. Robustness Checks for Efficiency Conclusions

Given the small sample size and potential scenarios such as multiple DMUs being equally efficient or certain inputs being zero, this study employed a multi-pronged strategy, including a grouped comparative DEA re-analysis, cross-efficiency, and super-efficiency evaluation under variable returns to scale (DEA-BCC) output orientation to enhance ranking resolution and conclusion robustness.

(1) A grouped comparative DEA re-analysis: To control for the potential confounding effect of institutional hierarchy dominating the efficiency assessment, this study conducted a grouped comparative DEA re-analysis alongside the global model. The sample was divided into two homogeneous subsets—a “high-level” group (central official media) and a “low-level” group (local government, other government and other media)—with independent efficiency frontiers established for each. Using the efficiency score results from the full sample as the benchmark, DEA-BCC models were independently run for the high-level group and the low-level group. If certain local media exhibited high efficiency within the “low-level” group but their efficiency scores diluted and lowered in the full sample due to the inclusion of central official media, this would strongly demonstrate the deterministic influence of institutional hierarchy on efficiency assessment. Conversely, if linguistic features remained significantly correlated with efficiency rankings after the hierarchy effect (i.e., in within-group comparisons) was controlled for, the conclusion that linguistic features drive efficiency would gain stronger support. This approach aimed to isolate the innate advantage of institutional hierarchy, thereby revealing the genuine impact of acquired operational strategies, such as linguistic features, on dissemination efficiency more clearly.

(2) Cross-efficiency: The optimal weight vector for each DMU was first determined, and then other DMUs were evaluated using this weight to obtain the cross-evaluation matrix (CE), where

$$CE(i \rightarrow j) = \frac{u_i y_i + u_{0,i}}{v_i x_j} \in (0, 1]$$

Robust ranking was performed based on column means, reflecting the transferability of weight patterns.

(3) Super-efficiency: To distinguish fully efficient DMUs, an output-oriented model employing self-exclusion re-evaluation yielded super-efficiency values $\phi > 1$. Given the small sample size and presence of zero inputs in this study, zero-value perturbation and relaxed convexity constraints were applied to prevent potential infeasibility under BCC-OO's convexity equality constraints. If infeasibility persisted, the method fell back to CCR-OO as a fallback, used solely for super-efficiency calculations without affecting cross-efficiency. This approach ensured feasibility and comparability under small sample sizes and extreme outlier conditions.

To clearly define all abbreviations and terms used throughout the paper, the list of abbreviations is shown in Table 2.

Table 2. List of abbreviations.

Abbreviation	Full Form	Definition/Explanation
DEA	data envelopment analysis	A non-parametric method used to evaluate the relative efficiency of peer decision-making units.
DEA-BCC	DEA model by Banker, Charnes, and Cooper	A DEA model that assumes variable returns to scale (VRS). Used in this study to calculate pure technical efficiency.
DMU	decision-making unit	The entity being evaluated for efficiency. In this paper, it refers to the four types of official social media accounts.
CE	cross-efficiency value	A DEA extension that allows efficient DMUs (on the frontier) to have scores greater than 1.0. This enables meaningful ranking among the already efficient units and identifies “how much” more efficient they are than the frontier.
ϕ	super-efficiency value	An extension of DEA that forms an efficiency matrix by evaluating every DMU using the optimal weights of every other DMU. The mean cross-efficiency for a DMU represents its peer-appraisal score, indicating the robustness and transferability of its performance beyond its own “self-appraisal” weights.
PTE	pure technical efficiency	Efficiency score from the BCC model, excluding the impact of scale size.
SE	scale efficiency	An SE of 1 indicates optimal scale, while $SE < 1$ indicates scale inefficiency (operating at an inappropriate size).
TE	comprehensive technical efficiency	$TE = PTE \times SE$.
θ	efficiency value	Standard notation for the efficiency score in DEA. A value of 1 indicates full efficiency.

4. Results

An empirical analysis based on the 2025 Beijing Extreme Rainfall Event was conducted. First, the DMUs were identified as four categories of communication entities: central mainstream media, other government media, local government media, and other media. Since the evaluation system's output indicators did not involve undesirable outputs, the DEA-BCC model was selected for efficiency analysis.

Next, based on the input indicator characteristics in Table 1, the open-source text large model (<https://chatglm.cn/main/alltoolsdetail?lang=zh>; based on ChatGLM-4) (ac-

cessed on 20 August 2025) was employed to statistically analyze the following narrative engagement metrics across the four communication entities: narrative immersion count X1, third-person perspective (narrative perspective) count X2, gain information framework count X3, fear emotion (narrative technique) count X4, and the quantity of pure text (media type) X5. This study used the role–task–constraint–format–example five-tuple prompt method. With narrative immersion serving as an example, the prompt is as follows:

“Role: You are a crisis communication analyst.

Task: Reading the text of Weibo below to determine whether narrative immersion (X1) occurs. The criteria are the presence of words related to characters, events, settings, time, or emotions in the text. If present, narrative immersion is deemed to occur; otherwise, it does not. Mark “1” for posts requiring narrative immersion; mark “0” otherwise.

Constraint: Only based on the text itself, without associating with external knowledge.

Format: The output must be JSON, and the fields: {“X1”: 0 or 1, “matching words”: [criteria word list], “reason”: one sentence reason}.

Example:

Sample input:

[Huairou traffic police ensured the smooth flow of disaster relief vehicles] in the afternoon of 29 July, the traffic detachment of Huairou branch of Beijing Municipal Public Security Bureau organized police to conduct temporary traffic management at the entrance of Hefang, G111 National Highway in Huairou District to prevent social vehicles from rushing into high-risk mountainous areas and persuade them to leave.

Sample output:

{“X1”: “0”, “matching words”: [], “reason”: “the text only contains institutional appellations and objective statements, without specific characters, time, environment and emotional words, so narrative substitution is not carried out.”}

When actually called, each indicator was called independently once to eliminate interference between indicators.

In order to ensure the extraction reliability of five input indicators, a verification process was conducted. First, this study used manual coding to further test the reliability. From 744 data items, 300 pieces of data were randomly selected according to DMUs hierarchy and were independently coded by one staff member of emergency education and two graduate students of communication. The five input indicators were marked one by one using the binary 0/1 method, allowing the “unable to judge” mark, which was unified after discussion. The calculated interval metric Krippendorff’s $\alpha = 0.81$ (95% CI: 0.77~0.85), which is higher than the threshold of 0.80, indicating that it is acceptable. Second, to verify whether the open-source model based on ChatGLM-4 had overfitting, this study calculated the corresponding reference value by using three traditional pipelines: emotional dictionary, part-of-speech tagging, and imperative rules. The results were significantly correlated with the LMM model output of 0.68~0.92, indicating that the model not only conforms to the classical method but also can capture the implicit expression. In addition, ChatGLM3-6B was used to reproduce all 744 extracted data items and rerun DEA-BCC, and Spearman ρ was 0.93 compared with the original model, which proved that the input index system is robust across models.

Additionally, three output metrics, i.e., Y1 (shares), Y2 (comments), and Y3 (likes), were separately tallied. The results are presented in Table 3.

Second, using MATLAB R2021b software to conduct DEA analysis on the data yielded the comprehensive technical efficiency, pure technical efficiency, scale efficiency, and returns to scale for the four DMUs, with results presented in Table 4.

Table 3. The input and output situations of emergency social media language services.

DMU	Input Indicator					Output Indicator		
	X1	X2	X3	X4	X5	Y1	Y2	Y3
Central mainstream media	34	263	3	1	2	51,897	72,259	291,777
Other government media	13	96	2	2	4	7468	7946	69,760
Local government media	59	167	3	5	1	12,169	6985	99,760
Other media	14	102	3	3	2	6483	5910	71,468
Mean $\pm$ SD	30.00 $\pm$ 21.618	157.00 $\pm$ 77.636	2.75 $\pm$ 0.500	2.75 $\pm$ 1.708	2.75 $\pm$ 1.258	19,504.25 $\pm$ 21,737.22	23,275 $\pm$ 32,666.588	133,191.25 $\pm$ 106,615.152

Table 4. Effectiveness analysis results.

DMU	Comprehensive Technical Efficiency (TE)	Pure Technical Efficiency (PTE)	Scale Efficiency (SE)	Variable Returns to Scale (Return Coefficient)
Central mainstream media	1.000	1.000	1.000	Unchanged (1.000)
Other government media	0.655	1.000	0.655	Incremental (0.239)
Local government media	0.684	1.000	0.684	Incremental (0.342)
Other media	0.632	1.000	0.632	Incremental (0.245)
Mean	0.743	1.000	0.743	-

Note: TE = PTE $\times$ SE.

Table 4 shows that among the four DMUs, only the central mainstream media achieved a TE of 1, indicating it attained the optimal effect in emergency social media language services and realized efficient resource utilization. The average TE across the four DMUs was 0.743. Among them, three units had a TE below the average, while the other media accounts exhibited the lowest TE at 0.632. This indicates that communication entities generally underperform in utilizing emergency social media language services. Further investigation into the causes and implementation of corresponding measures are needed to enhance service effectiveness and improve overall efficiency. Notably, while the TE values for local government media accounts, other government accounts, and other media accounts were ineffective, their PTE reached 1. This indicates that the primary reason for the ineffective TE in these three communication entities was low SE.

The scale returns and scale coefficient results in Table 4 reveal that central mainstream media accounts achieved a scale efficiency of 1, with scale returns remaining constant. This indicates that output reached maximum scale with the same input level. Further increasing input metrics may not significantly enhance the effectiveness of their emergency social media language services. Optimizing resource allocation or improving input management should be considered to maintain a balanced input-output return ratio. Local government media, other government media, and other media accounts all exhibited increasing returns to scale, indicating a need for sustained input increases to achieve optimal input-output ratios.

To address the confounding effects introduced by the type of communicator (institutional hierarchy of media accounts), a grouped (DEA) was first conducted. The results are presented in Table 5.

Table 5. Grouped DEA re-analysis results.

DMU Group	DMU Type	Comprehensive Technical Efficiency (TE)	Pure Technical Efficiency (PTE)	Scale Efficiency (SE)	Variable Returns to Scale (Return Coefficient)
High-level group	Central mainstream media	1.000	1.000	1.000	Unchanged (1.000)
Group mean		1.000	1.000	1.000	-
Low-level group	Other government media	1.000	1.000	1.000	Incremental (1.000)
	Local government media	0.980	1.000	0.980	Incremental (>1.000)
	Other media	0.945	1.000	0.945	Incremental (>1.000)
Group mean		0.975	1.000	0.975	-

Note: $TE = PTE \times SE$.

In the comparison of efficiency scores and rankings of the same media units in the full-sample model and the sub-sample models, as shown in Table 5, it is evident that in the full-sample Model 1, the top four positions in efficiency rankings are all occupied by central official media. This indicates that the central mainstream media achieved a perfect score ($TE = 1.000$) in both the pooled and high-level group analysis, confirming that institutional hierarchy itself is a powerful determinant of efficiency. Central mainstream media was the only type in the high-level group. Its efficiency remained 1.000, but there were no longer other media types whose scores were suppressed by its presence. In the low-level group, the efficiency scores for other government media, local government media, and other media increased significantly compared to the pooled model (where the TEs were 0.655, 0.684, and 0.632, respectively). For instance, other government media had perfect efficiency ($TE = 1.000$), becoming the benchmark within its group, and local government media had a high score of 0.980.

This dramatic improvement in efficiency scores for low-level media after for institutional hierarchy was controlled for clearly demonstrates that their dissemination efficiency (potentially driven by factors like linguistic features) is indeed outstanding when evaluated on a level playing field. The “language features drive efficiency” hypothesis is supported by robust, controlled evidence.

The results from the cross-efficiency evaluation matrix (Figure 3) and average cross-efficiency ranking (Figure 4) are provided below. This paper only reports the relative efficiency and rankings within the sample, without making statistical inferences about the population.

Figure 3 presents a cross-efficiency heatmap for four subject categories (behavior evaluators, listed as the evaluated). The robust ranking based on column means, as shown in Figure 4, is as follows: central mainstream media (0.875) > local government media (0.815) > other media (0.395) > other government media (0.260). This indicates that in terms of peer evaluation, the weighting patterns of central mainstream media and local government media exhibited stronger transferability, while other government media and other media received lower recognition under peer assessment.

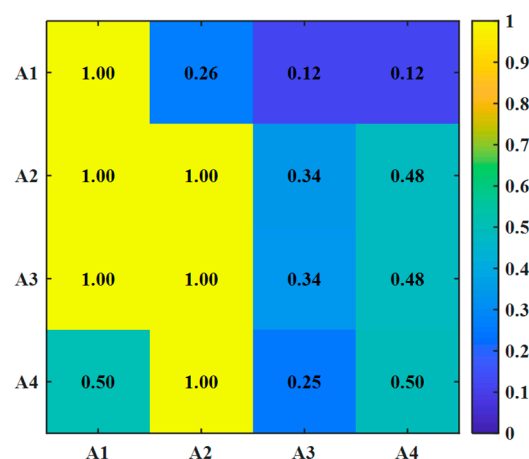


Figure 3. Cross-efficiency heatmap (BCC-OO). A1: central mainstream media; A2: other government media; A3: local government media; A4: other media.

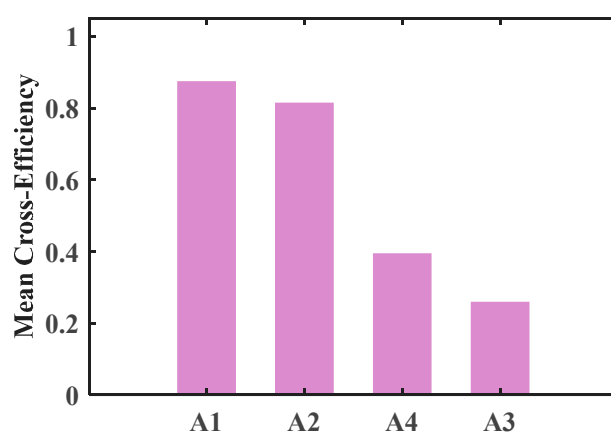


Figure 4. Average cross-efficiency ranking. A1: central mainstream media; A2: other government media; A3: local government media; A4: other media.

The super-efficiency results are shown in Figure 5.

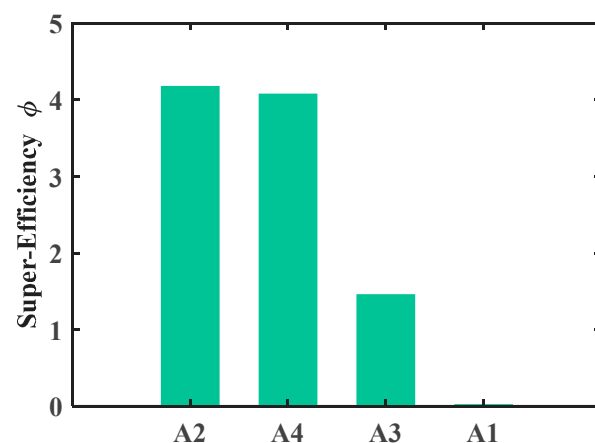


Figure 5. Super-efficiency ranking. A1: central mainstream media; A2: other government media; A3: local government media; A4: other media.

Other government media $\phi = 4.18$, other media $\phi = 4.08$, local government media $\phi = 1.46$, and central mainstream media $\phi \approx 0.03$. The results indicate that other government media and other media achieved frontier status under their “optimal weights” but exhibited low mutual recognition, constituting “exclusive frontiers.” Overall, cross-efficiency ranking better reflects generalized effectiveness.

Simultaneously, the DEA-BCC model was employed to calculate and analyze the slack variables of each DMU. The improvement levels and average values of DUM's input and output indicators were computed using the following formula, with the results presented in Table 6.

$$L = \begin{cases} I_{e,i}/I_{a,i} \\ O_{s,i}/O_{a,i} \end{cases} \times 100\%$$

where, L represents the improvement percentage level (%), $I_{e,i}$ represents the excess input value of the i -th DUM, $I_{a,i}$ represents the actual input value of the i -th DUM, $O_{s,i}$ represents the output deficiency value of the i -th DUM, and $O_{a,i}$ represents the actual output value of the i -th DUM.

Table 6. Analysis of slack variable results.

DMU	Input Indicator					Output Indicator		
	s_{X1}^-	s_{X2}^-	s_{X3}^-	s_{X4}^-	s_{X5}^-	s_{Y1}^+	s_{Y2}^+	s_{Y3}^+
Central mainstream media	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Other government media	0.386	0.000	0.593	1.071	2.142	4939.882	9330.166	0.000
Local government media	28.720	24.275	1.026	3.077	0.000	5574.841	17,720.71	0.000
Other media	0.514	0.000	1.160	1.650	0.773	6228.676	11,789.155	0.000
Improvement percentage level (%)	13.85	3.63	25.63	42.50	23.05	52.00	142.65	0

Note: $s_{X1}^- \sim s_{X5}^-$ mean slack variables of input indicators, and $s_{Y1}^+ \sim s_{Y3}^+$ mean slack variables of output indicators.

Table 6 indicates that among the input indicators, narrative technique (fear emotion) showed the largest improvement margin at 42.50%, followed by gain information framework, with an average improvement percentage of 25.63%. Among DMUs, local government media and other government media accounts exhibited notably high aggregate input values of 57.098 and 4.191 respectively. This indicates these two communication entities require substantial adjustments to enhance their DEA effectiveness, thereby improving both public opinion dissemination and emergency social media language service capabilities.

In summary, during the emergency response phase, central mainstream media demonstrated the highest effectiveness in emergency social media language services. Local government media, other government media, and other media all require further optimization of their communication investments. Significant disparities exist between other government media and other media compared to central mainstream media, necessitating further exploration of methods to enhance their emergency language service effectiveness.

5. Discussion

This study constructed an evaluation framework for the effectiveness of emergency social media language services during the emergency response phase based on Hovland's persuasion theory. Using the DEA method, it empirically assessed the service efficacy of four official communication entities during the 2025 Beijing Extreme Rainfall Event. The findings clearly reveal significant disparities in emergency language service effectiveness among different communication entities: central mainstream media demonstrated outstanding performance, while local government media, other government media, and other media had considerable room for improvement. These findings provide crucial empirical evidence for understanding the operational mechanisms and optimization pathways of emergency social media language services.

5.1. Main Findings

This study identified central mainstream media as having distinct advantages. Their superior effectiveness can be explained across multiple dimensions of Hovland's the-

ory. As communicators, central media typically possess the highest credibility (authority, trustworthiness) and attractiveness (broad audience base, brand influence). In terms of message organization, their information dissemination is often more standardized, systematic, and comprehensive, aligning with the crisis communication principle of “quickly stating facts, cautiously explaining causes, emphasizing attitudes, and sincerely outlining measures.” [28]. Their nationwide audience reach (audience characteristics) facilitates effective coverage by reaching both key decision-makers and the broader public. These factors collectively enable the conversion of inputs (such as manpower and information gathering capabilities) into efficient outputs (including information coverage, public responsiveness, and action guidance effectiveness).

This paper indicates that local government media and other government media accounts exhibit relatively lower effectiveness. The potential reasons for this include the following: The first reason involves credibility gaps and resource constraints; i.e., media at local levels or within specialized departments may face perceived lower authority and recognition among the public, coupled with often limited human and technical resources. The second reason is challenges in information positioning and timeliness. Released content may prioritize localized or department-specific topics or suffer from fragmentation and delayed updates, making it difficult to swiftly capture and sustain audience attention amid emergency information overload. The third reason pertains to audience targeting and reach. Local government and other government media may have relatively limited core audiences, resulting in insufficient information penetration and coverage during cross-regional, nationwide events. Finally, from the perspective of information organization and persuasion strategies, both types of media may fail to maximize the impact of their messages and have room for improvement in persuasion techniques such as information presentation methods, language accessibility, emotional resonance, and calls to action.

To provide a theoretical lens for interpreting these findings and the subsequent cross-efficiency results, we offer the following mechanistic reflections. By mapping narrative features onto cognitive, emotional, and social-diffusion pathways, we arrived at the following conclusions: clear, structured, and bullet-pointed expression reduces intrinsic cognitive load in high-stress situations, with an expected manifestation of higher action-oriented interactions and shorter response latencies; rich media and gain-framed messages enhance attention retention and intention formation and can be anticipated to yield higher 24/72 h interaction retention rates and deeper forwarding chains; templated expressions and relay coordination provide consistent reference points and low-cost replication paths and can be expected to increase the proportion of forwards containing “@/relay tags” and enhance cross-account temporal coupling. The empirical results of this study can be viewed as correlational evidence supporting these mechanisms but cannot support causal claims.

Additionally, this study found that central mainstream media (0.875) and local government media (0.815) maintained high efficiency in peer evaluations, indicating that their content–structure combination of low X4 (low expressive deficiency) + low X5 (low text-only content) + low X3 (low gain deficiency) possesses strong transferability. They performed well even when evaluated by other entities’ scoring standards. In contrast, other government media (0.260) and other media (0.395) exhibited relatively low mutual evaluation averages. This indicates that while they can achieve frontier status ($\varphi > 1$) under their own weighting schemes, these optimal weights are not universally applicable to others, representing a “proprietary frontier.” Therefore, to enhance generalized effectiveness, optimization should focus on reducing X4 (enhancing narrative and third-person expression), reducing X5 (minimizing plain text, increasing rich media), and strengthening X3 (reinforcing gain-information frameworks) to narrow the gap with the transferable frontier.

5.2. Research Significance

This study successfully applied Hovland's classic persuasion theory to the dynamic and complex field of evaluating emergency social media language services, validating the framework's applicability and explanatory power in interpreting differences in crisis communication effectiveness. By identifying key factors influencing effectiveness (communicator attributes, information characteristics, audience needs) and constructing a quantifiable input-output indicator system, this research enriches the empirical foundation for interdisciplinary studies at the intersection of emergency language services, crisis communication, and persuasion theory.

The findings highlight the urgency of enhancing the effectiveness of other government media and local government media, offering recommendations for optimizing resource allocation and media capacity building. Furthermore, by leveraging big data technology to evaluate emergency language communication outcomes, this study provides data-driven insights for refining information dissemination strategies on mainstream social platforms such as Weibo, which includes topic selection, release timing, and interactive engagement approaches.

5.3. Limitations and Future Directions

The conclusions primarily apply to the specific scenario of "social platform–natural disaster–official account." Differences may exist across other disaster types (e.g., earthquakes, epidemics), platforms (e.g., short videos, instant messaging), and governance systems. Future research should explore cross-platform, cross-disaster, and cross-regional validation studies. The DEA method employed in this study focuses on the relative efficiency in input–output relationships, potentially undercapturing qualitative or lagging indicators such as information quality, emotional resonance, and long-term persuasive effects. Future research may integrate content analysis, affective computing, and survey methods to construct a more comprehensive multidimensional evaluation framework.

6. Conclusions

Through theory-driven evaluation framework construction and empirical analysis, this study reveals significant disparities in the effectiveness of emergency social media language services across different communication entities and their underlying causes. The research not only validates the applicability of Hovland's persuasion theory in emergency communication but also provides concrete directions and strategic recommendations for enhancing the service efficacy of critical yet relatively underperforming communication entities (such as local government and other government media). Overcoming existing limitations and deepening research on information quality, multi-platform ecosystems, audience feedback, and effectiveness across different event types will be crucial future directions for advancing the scientific and precise development of emergency language services. This holds significant practical implications for enhancing the modernization of China's emergency management system and capabilities.

Author Contributions: Conceptualization, J.G.; Methodology, J.G.; Software, Y.Z.; Investigation, J.G., S.L. and H.L.; Resources, S.F.; Data curation, H.L.; Writing—original draft, J.G.; Writing—review & editing, Y.Z.; Visualization, Y.Z.; Supervision, S.F., J.C. and S.L.; Project administration, J.C.; Funding acquisition, J.G. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work received basic research funding from the China Academy of Safety Science and Technology (2025JBKY22) and the National Language Commission Scientific Research Project (WT145-21).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: Author Hezhuang Lou was employed by the company Beijing Drainage Group CO., LTD. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

- Shao, M.; Song, Y.; Teng, C.; Zhang, Z. Algorithms and simulation of multi-level and multi-coverage on cross-regional emergency facilities. *Wirel. Pers. Commun.* **2018**, *102*, 3663–3676. [\[CrossRef\]](#)
- Charnes, A.; Cooper, W.W. Preface to topics in data envelopment analysis. *Ann. Oper. Res.* **1984**, *2*, 59–94. [\[CrossRef\]](#)
- Emrouznejad, A.; Yang, G.L. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Econ. Plan. Sci.* **2018**, *61*, 4–8. [\[CrossRef\]](#)
- Burgoon, J.K.; Burgoon, M.; Miller, G.R.; Sunnafrank, M. Learning theory approaches to persuasion. *Hum. Commun. Res.* **1981**, *7*, 161–179. [\[CrossRef\]](#)
- Han, X.; Wang, J.; Zhang, M.; Wang, X. Using social media to mine and analyze public opinion related to COVID-19 in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2788. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chang, E.; Davis, T.L.; Berkman, N.D. Differences in telemedicine, emergency department, and hospital utilization among nonelderly adults with limited English proficiency post-COVID-19 pandemic: A cross-sectional analysis. *J. Gen. Intern. Med.* **2023**, *38*, 3490–3498. [\[CrossRef\]](#)
- Xie, J.; Ma, S. Research on the service quality of emergency medical language services during major unexpected public health events. *Front. Public Health* **2023**, *11*, 1169222. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rotoli, J.M.; Hancock, S.; Park, C.; Demers-Mcletchie, S.; Panko, T.L.; Halle, T.; Wills, J.; Scarpino, J.; Merrill, J.; Cushman, J.; et al. Emergency medical services communication barriers and the deaf American sign language user. *Prehospital Emerg. Care* **2022**, *26*, 437–445. [\[CrossRef\]](#)
- Dreisbach, J.L.; Mendoza-Dreisbach, S. The integration of emergency language services in COVID-19 response: A call for the linguistic turn in public health. *J. Public Health* **2021**, *43*, e248–e249. [\[CrossRef\]](#)
- Tang, Z.; Yang, H.; Yang, D. Emergency Language Services in the China-Myanmar Borderland: A Case of Multilingual Translation Center at Mengding. *Asian Soc. Sci.* **2024**, *20*, 1–12. [\[CrossRef\]](#)
- Guo, X.; Xiao, D.; Guo, Y. From crisis to opportunity: Advancements in emergency language services. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1–18. [\[CrossRef\]](#)
- Gebremariam, H.T.; Mulugeta, Z.A. In-service language teachers' engagement with online learning platforms after the emergence of COVID-19. *Ampersand* **2025**, *14*, 100215. [\[CrossRef\]](#)
- Bai, E.; Luo, X.; Zhang, Z.; Adelgais, K.; Ali, H.; Finkelstein, J.; Kutzin, J. Assessment and Integration of Large Language Models for Automated Electronic Health Record Documentation in Emergency Medical Services. *J. Med. Syst.* **2025**, *49*, 65. [\[CrossRef\]](#)
- Noack, E.M.; Schulze, J.; Müller, F. Designing an app to overcome language barriers in the delivery of emergency medical services: Participatory development process. *JMIR mHealth uHealth* **2021**, *9*, e21586. [\[CrossRef\]](#) [\[PubMed\]](#)
- Omidi, L.; Salehi, V.; Zakerian, S.A.; Saraji, J.N. Performance optimization of human factors and safety performance using an integrated DEA-TOPSIS approach: A case study in the process industry. *Soc. Sci. Humanit. Open* **2025**, *12*, 101766. [\[CrossRef\]](#)
- Suh, Y. Developing Productivity–Safety Effectiveness Index Using Data Envelopment Analysis (DEA). *Appl. Sci.* **2025**, *15*, 1989. [\[CrossRef\]](#)
- Qi, H.; Zhou, Z.; Li, N.; Zhang, C. Construction safety performance evaluation based on data envelopment analysis (DEA) from a hybrid perspective of cross-sectional and longitudinal. *Saf. Sci.* **2022**, *146*, 105532. [\[CrossRef\]](#)
- Nahangi, M.; Chen, Y.; McCabe, B. Safety-based efficiency evaluation of construction sites using data envelopment analysis (DEA). *Saf. Sci.* **2019**, *113*, 382–388. [\[CrossRef\]](#)
- Ooms, J.A.; Jansen, C.J.M.; Hoeks, J.C.J. The story against smoking: An exploratory study into the processing and perceived effectiveness of narrative visual smoking warnings. *Health Educ. J.* **2020**, *79*, 166–179. [\[CrossRef\]](#)
- Murphy, S.T.; Frank, L.B.; Chatterjee, J.S.; Baezconde-Garbanati, L. Narrative versus nonnarrative: The role of identification, transportation, and emotion in reducing health disparities. *J. Commun.* **2013**, *63*, 116–137. [\[CrossRef\]](#)
- Klauck, T.; Köppe, T.; Onea, E. Internally focalized narration from a linguistic point of view. *Sci. Study Lit.* **2012**, *2*, 218–242. [\[CrossRef\]](#)
- Nan, X.; Dahlstrom, M.F.; Richards, A.; Rangarajan, S. Influence of evidence type and narrative type on HPV risk perception and intention to obtain the HPV vaccine. *Health Commun.* **2015**, *30*, 301–308. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, M.; Bell, R.A.; Taylor, L.D. Persuasive effects of point of view, protagonist competence, and similarity in a health narrative about type 2 diabetes. *J. Health Commun.* **2017**, *22*, 702–712. [\[CrossRef\]](#)

24. Brusse, E.D.A.; Fransen, M.L.; Smit, E.G. Framing in entertainment-education: Effects on processes of narrative persuasion. *Health Commun.* **2017**, *32*, 1501–1509. [[CrossRef](#)]
25. Liu, S.; Yang, J.Z.; Chu, H. Now or future? Analyzing the effects of message frame and format in motivating Chinese females to get HPV vaccines for their children. *Patient Educ. Couns.* **2019**, *102*, 61–67. [[CrossRef](#)]
26. Hovland, C.I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *Am. Psychol.* **1959**, *14*, 8. [[CrossRef](#)]
27. Shen, F.; Sheer, V.C.; Li, R. Impact of narratives on persuasion in health communication: A meta-analysis. *J. Advert.* **2015**, *44*, 105–113. [[CrossRef](#)]
28. Reynolds, B. Crisis and Emergency Risk Communication. *Appl. Biosaf.* **2005**, *10*, 47–56. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.