

# Research on Gaokao Public Opinion Based on Weibo Big Data: Spatiotemporal Variations in Attention and Sentiment Components and Analysis of Macro Factor Influences

**Authors:** Dong Jincheng, Ai Yuheng, Zhu Tingshao, Tingshao Zhu

**Date:** 2025-01-10T00:00:00+00:00

## Abstract

Based on Weibo big data text content, this study analyzes changes in public opinion intensity and emotional composition regarding the Gaokao (college entrance examination) across provincial-level administrative regions in mainland China from 2019 to 2024, and examines the relationship between these changes and macroeconomic indicators. The study employs the number of text entries as an intensity metric, combined with Baidu Index analysis. Results indicate that Gaokao-related public opinion intensity varies across different years and provinces, and is closely correlated with local economic indicators such as per capita regional GDP and local fiscal education expenditure. Furthermore, emotional composition analysis conducted using the LinCui Analysis System and Dalian University of Technology Emotion Lexicon reveals that Gaokao-related texts exhibit distinct compositional characteristics across different time periods and regions, which are associated with certain real-world factors. This study provides a novel big data perspective for understanding the patterns of change in Gaokao public opinion intensity and emotional composition, as well as their relationship with socioeconomic factors.

## Full Text

### Public Opinion Research on the College Entrance Examination Based on Weibo Big Data: An Analysis of Temporal-Spatial Variations in Attention and Sentiment Components and the Influence of Macro-Level Factors

Dong Jincheng<sup>1,2</sup>, Ai Yuheng<sup>3,4</sup>, Zhu Tingshao<sup>1,2</sup>

<sup>1</sup>Institute of Psychology, Chinese Academy of Sciences, Beijing 100101

<sup>2</sup>Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049

<sup>3</sup>Institute of Automation, Chinese Academy of Sciences, Beijing 100190

<sup>4</sup>School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049

## Abstract

This study analyzes temporal-spatial variations in public attention and sentiment toward the college entrance examination (Gaokao) across provincial-level administrative regions in mainland China from 2019 to 2024, using textual content from Weibo big data, and examines the relationship between these variations and macroeconomic indicators. Employing text volume as an attention metric and integrating Baidu Index data, the results demonstrate significant inter-annual and inter-provincial differences in Gaokao-related public attention, which correlate closely with local economic indicators such as per capita gross regional product and local fiscal expenditure on education. Sentiment component analysis conducted via the Lincui Analysis System and the Dalian University of Technology Sentiment Lexicon reveals that Gaokao-related texts exhibit distinct temporal and regional characteristics, associated with various real-world factors. This study provides a novel big-data perspective for understanding the evolving patterns of public attention and sentiment toward the Gaokao and their relationship with socioeconomic factors.

**Keywords:** big data; social media; Weibo; college entrance examination; attention data analysis; sentiment component analysis

## Introduction

The rapid development of big data technologies has revolutionized social science research, particularly in analyzing social media data, substantially expanding research scope and depth (Du Zhijuan et al., 2017; Wu Shengtao et al., 2023). For instance, Zhu Tingshao et al. (2013) proposed a group event risk warning model using social media data, demonstrating the practical application of big data in dynamic social monitoring. Huang Feng et al. (2020) examined the relationship between economic development and collective emotions and morality, validating the utility of big data in analyzing socioeconomic phenomena.

Weibo (weibo.com), as one of mainland China's most influential social media platforms, has become a critical source for obtaining public emotion and behavioral data due to its extensive user base and rapid information dissemination capabilities, and has been widely adopted by domestic scholars (Chang Jianxia & Li Junyi, 2021; Chen Xingshu et al., 2020; Dai Yifang, 2023; Fan Xiaolei, 2023; Xiao Jiarui et al., 2021). In the international academic community, Weibo-based research is also noteworthy: Zhao and Wang (2023) revealed the framing interactions between the public and government media during the pandemic by

analyzing Weibo data, while Xiao et al. (2018) achieved real-time monitoring of urban rainstorm waterlogging disasters affecting citizens' lives through Weibo data.

Social media platforms like Weibo have become important venues for public expression of independent discourse. Chang Jianxia and Li Junyi (2021) analyzed the temporal-spatial distribution of public anxiety during the COVID-19 pandemic through Weibo data, demonstrating that social media data can sensitively reflect social emotional fluctuations. Xiao Jiarui et al. (2021) retrieved representative typical texts from Weibo data to explore the relationship between kindergarten entry anxiety and mothers' personality traits and anxiety levels.

The Gaokao is mainland China's educational examination and admission system (unless otherwise specified, "nationwide" hereinafter refers to mainland China, comprising 31 provincial-level administrative regions, excluding Hong Kong, Macau, and Taiwan, which do not implement the same Gaokao policies). It is also an annual social hotspot event, and Gaokao reform has consistently been a focal topic in Chinese education, concerning not only educational equity but also extensive social psychological elements (Liu Haifeng, 2009). In recent years, the implementation of the new Gaokao reform, with its policy complexity and regional differences, has triggered widespread public discussion and debate (Yan Kun & Wu Han, 2022; Zhong Binglin & Wang Xinfeng, 2019). For example, Zhang Lan (2024) analyzed public opinion data on Jiangsu's new Gaokao subject selection policy, revealing the public's main concerns about the reform and proposing recommendations for strengthening policy communication and optimizing implementation.

Yang Xianmin et al. (2019) noted that big data technology plays a significant role in Gaokao reform, including supporting student subject selection decisions, volunteer application assistance, and teaching evaluation optimization. Big data-supported public opinion analysis provides policymakers with powerful tools to understand public needs and opinions, thereby enhancing policy scientificity and feasibility (Zhang Lan, 2024). Additionally, Zhong Binglin and Wang Xinfeng (2019) argued that the complexity of Gaokao reform requires the policy-making process to be data-driven while attending to public psychological reactions and social impacts. Zhou Xu (2021) further illustrated through analysis of parental anxiety under the "Double Reduction" policy that educational policy implementation must consider social psychological impacts to avoid triggering unnecessary public emotional fluctuations.

Sentiment analysis represents an important direction in natural language processing, revealing public emotional trends through textual content analysis (Wang, 2023). The lexicon-based method is a crucial technique that relies on existing lexical resources and is widely applied in sentiment analysis, topic modeling, and semantic parsing. Using pre-compiled dictionaries, researchers can rapidly analyze massive texts through word segmentation programs. In recent years, with advances in machine learning and natural language processing technologies, sentiment analysis of Weibo data has significantly

improved in accuracy and breadth. However, data processing still faces numerous challenges, such as filtering data noise, identifying diverse emotional expressions, and analyzing unstructured text (Chen Xingshu et al., 2020; Wu Zepeng, 2020).

Chang Jianxia and Li Junyi (2021) demonstrated that public anxiety on Weibo during the pandemic was significantly positively correlated with newly confirmed case numbers, further confirming the importance of sentiment analysis in revealing social emotional patterns. Zhang Fang and Gan Haochen (2020) found that psychological distance significantly impacted public emotions by analyzing the temporal and spatial distribution of Weibo sentiment during the pandemic. Dai Yifang (2023) emphasized that accurately capturing sentiment information in Weibo data can provide valuable support for policy adjustments. Furthermore, Zhu Tingshao et al. (2013) noted that sentiment analysis based on social media data can also be used for risk warning of emergencies, providing powerful tools for social governance.

Existing research remains insufficient in utilizing social media big data to analyze the relationship between public sentiment and Gaokao policy, particularly lacking nationwide and long-term analyses. Such big data analysis requires comprehensive consideration of local Gaokao reform progress, macroeconomic indicators, and the number of current-year examinees, as well as the impact of the COVID-19 pandemic after 2020 on Gaokao public opinion.

This study employs both cross-sectional and longitudinal designs to examine temporal-spatial variations in attention and textual sentiment toward the Gaokao on Sina Weibo from 2019 to 2024 across provincial-level administrative regions in mainland China. The research proceeds in two phases: Study One will use the 2019-2023 Weibo text database constructed by Zhu Tingshao's research team at the Institute of Psychology, Chinese Academy of Sciences, and web crawlers to obtain Weibo datasets from January 5 and June 5 of 2019-2024, combined with Baidu Index ([index.baidu.com](http://index.baidu.com)), to examine potential attention changes before and after the new Gaokao reform. Through statistical analysis of attention data and macro-level indicators, it preliminarily explores real-world factors influencing Gaokao public opinion attention. Building on this, Study Two will use the Lincui Analysis System for word segmentation and sentiment factor extraction based on the Dalian University of Technology Sentiment Lexicon (Wang Damao, 2021) to examine the temporal-spatial distribution of sentiment in Gaokao-related Weibo texts, investigate its temporal changes in relation to the pandemic and Gaokao reform, and explore its spatial differences and statistical associations with macroeconomic indicators, attempting to interpret real-world factors influencing emotional attitudes toward the Gaokao in public discourse.

## Study One: Attention Variation

### Datasets

**Dataset A:** Extracted from the 2019-2023 Weibo text database constructed by Zhu Tingshao’s research team at the Institute of Psychology, Chinese Academy of Sciences, using “Gaokao” as the keyword, with entries containing “Global Gaokao” (the title of an online novel with high discussion volume on Weibo) removed. This yielded Dataset A, comprising 425,219 entries and 35,661,285 characters, with an average length of 83.87 characters. This dataset originates from a full-time-period crawled Weibo text database with temporal-spatial continuity. However, since the initial crawl was incomplete and could not cover all texts on the entire Weibo platform, the extracted volume of entries containing “Gaokao” is constrained by the crawling methodology, presenting certain limitations.

**Dataset B:** Using Python-based web crawlers, we obtained texts from the entire Weibo platform on January 5 and June 5 of 2019-2024 (the Gaokao is typically held June 7-11; we selected June 5 before the exam and January 5 before the final winter break, totaling 12 days; hereinafter “Gaokao season” refers to June-July when the Gaokao and volunteer applications occur) containing “Gaokao” but not “Global Gaokao.” This yielded Dataset B, comprising 101,842 entries and 10,860,655 characters, with an average length of 106.64 characters. This dataset consists of single-day crawls without temporal-spatial continuity, reflecting greater contingency in data volume (attention), but achieved comprehensive coverage of all relevant texts on the entire Weibo platform for single days.

### Longitudinal Attention Trends

This study employs text entry count as the attention metric. Dataset A trends show: All provinces exhibit highly homogeneous entry count variation trends, with inexplicably low attention in 2022 and 2023 across all provinces. Therefore, this dataset is deemed unable to reflect genuine temporal-spatial attention variations, presumably due to the crawling methodology of the original text database for Dataset A.

Dataset B trends show: Due to limitations of single-day crawling, this dataset also cannot reflect genuine temporal-spatial attention variations.

We used Baidu Index (index.baidu.com, based on comprehensive analysis of Baidu user behavior data) search index (using Baidu search volume as the data foundation, with “Gaokao” as the statistical object, analyzing and calculating the weighted sum of search frequency in Baidu web searches; hereinafter Baidu Index refers to Baidu search index) as the metric reflecting temporal-spatial attention variations. Considering the temporal-spatial differences in new Gaokao reform implementation:

**Batch 1 (2017):** Shanghai, Zhejiang — 3+3 model

**Batch 2 (2020):** Beijing, Tianjin, Shandong, Hainan — 3+3 model

**Batch 3 (2021):** Hebei, Jiangsu, Fujian, Hubei, Hunan, Guangdong, Chongqing, Liaoning — 3+1+2 model

**Batch 4 (2024):** Jilin, Anhui, Jiangxi, Guangxi, Guizhou, Gansu, Heilongjiang — 3+1+2 model

**Batch 5 (2025):** Shanxi, Henan, Shaanxi, Sichuan, Yunnan, Ningxia, Inner Mongolia, Qinghai — 3+1+2 model

Extracting Baidu Index trend graphs for two regions per batch reveals: Despite different Gaokao reform timelines, regional search index variations remain highly homogeneous. The 2019 Gaokao season index was extremely low (presumably due to incomplete raw data records required for Baidu Index algorithm in 2019), higher in 2022 and 2023, and relatively lower in 2021 and 2024. Meanwhile, intra-season variations show distinct double peaks: the first peak occurs during the examination week, and the second appears during score release and volunteer application periods. All provinces exhibit a phenomenon where the first peak exceeded the second before 2022, while the second peak exceeded the first from 2022 onward.

Since Baidu has not disclosed the specific search index algorithm, explaining this homogeneity is difficult. If the algorithm itself does not have a homogenizing tendency, these data may indicate that the new Gaokao reform's impact on Chinese internet attention toward the Gaokao is negligible. Considering that most current-year examinees and their parents may not understand Gaokao reforms and policy changes, this internet phenomenon can be explained by real-world circumstances. In other words, annual Gaokao season attention is primarily driven by the Gaokao itself, with longitudinal differences in internet data more influenced by the broader internet environment of that period than by policy differences between current and previous years.

Since relevant data such as regional Gaokao applicant numbers and fiscal budgets have grown steadily and slowly over the six years, making statistical analysis like correlation testing difficult with the irregularly varying attention data across years.

### Cross-Sectional Factor Exploration

The primary factor influencing the spatial distribution of Gaokao-related public opinion attention should be regional population size, i.e., the annual number of Gaokao applicants in each provincial-level administrative region. Through information released on local education official websites, we obtained the number of Gaokao applicants for 31 provincial-level administrative regions from 2019 to 2024.

Conducting correlation tests between the total five-year applicant numbers and total heat from Dataset A yielded Pearson's  $r = 0.291$ ,  $p = 0.112$ . Observing the scatter plot, we noticed three outlier data points above the regression line, which upon inspection corresponded to Beijing, Shanghai, and Guangdong—the four first-tier “Beijing-Shanghai-Guangzhou-Shenzhen” cities with more developed

economies and higher informatization levels. After removing these three regions and retesting, Pearson's  $r = 0.547$ ,  $p = 0.003$ .

Correlating 2019-2024 total applicant numbers with Baidu Index values after removing the aforementioned three regions yielded Pearson's  $r = 0.834$ ,  $p < 0.001$ ; Spearman's  $\rho = 0.909$ ,  $p < 0.001$ . This indicates that Baidu's search index has a stronger correlation with applicant numbers, while Weibo's user composition and text entry count are more influenced by regional development and internet usage depth, presenting certain limitations that may prevent it from accurately reflecting real-world regional Gaokao attention.

Economic conditions and local development evidently affect educational investment and public internet usage, thereby influencing Gaokao public opinion attention. We conducted statistical tests using 2019-2020 macroeconomic indicators (total local fiscal expenditure on education for aggregate measures, and per capita gross regional product reflecting development level for per capita measures) against total Baidu Index values and total heat from Dataset A for all 31 provincial-level administrative regions during the same period.

**Dataset A Heat vs. Total Local Fiscal Expenditure on Education:** Pearson's  $r = 0.718$ ,  $p < 0.001$ ; Spearman's  $\rho = 0.744$ ,  $p < 0.001$ .

**Dataset A Heat vs. Per Capita Gross Regional Product:** Pearson's  $r = 0.650$ ,  $p < 0.001$ ; Spearman's  $\rho = 0.660$ ,  $p < 0.001$ .

**Baidu Index vs. Total Local Fiscal Expenditure on Education:** Pearson's  $r = 0.968$ ,  $p < 0.001$ ; Spearman's  $\rho = 0.936$ ,  $p < 0.001$ .

**Baidu Index vs. Per Capita Gross Regional Product:** Pearson's  $r = 0.302$ ,  $p = 0.099$ ; Spearman's  $\rho = 0.452$ ,  $p = 0.011$ .

These results show that for total Gaokao attention, Baidu Index demonstrates stronger correlation with total local fiscal expenditure on education but weaker correlation with per capita gross regional product compared to Dataset A from Weibo. Conversely, Weibo attention shows significant correlation with per capita economic data. This further confirms our earlier perspective: Baidu's platform, particularly its search function, is more universally used with lower barriers to entry, making its search index more influenced by total internet user volume. Weibo, as social media, has publication and interaction features that make its attention more influenced by regional development level and internet usage depth. Consequently, its correlation with total local fiscal expenditure on education is weaker than Baidu Index's, but its correlation with per capita economic data is stronger.

## Study Two: Sentiment Components

We used the Lincui Analysis System<sup>1</sup> for word segmentation and sentiment factor extraction to obtain relative sentiment component values, i.e., the proportion of each sentiment component in the text content<sup>2</sup>. The lexicon employed was the Dalian University of Technology Chinese Sentiment Lexicon Ontology (Wang Damao, 2021). This ontology, compiled and annotated through the col-

lective efforts of the Information Retrieval Laboratory at Dalian University of Technology under Professor Lin Hongfei’s guidance, describes Chinese words or phrases from multiple dimensions, including part-of-speech categories, sentiment categories, sentiment intensity, and polarity. The sentiment classification system is based on Ekman’s six basic emotions, with the addition of the “good” category for more nuanced positive sentiment classification, resulting in seven major categories and 21 subcategories<sup>3</sup>.

**Dataset A:** Total word count of 35,661,285; average dictionary coverage rate of 0.076. We examined subcategories with higher nationwide totals across all time periods compared to other categories, then compared these results with sentiment components from randomly sampled Weibo texts without keywords, selecting categories where nationwide totals showed larger gaps. Ultimately, nine categories were included: happiness (PA), peace (PE), trust (PG), wish (PK), panic (NI), fear (NC), annoyance (NE), blame (NN), and doubt (NL).

**Dataset B:** Total word count of 10,860,654; average dictionary coverage rate of 0.091. The examination process and final consideration scope for sentiment components were identical to Dataset A.

### Temporal Variation Patterns

First, we examined sentiment component changes for 2019-2023 combined totals from Dataset A. Happiness (PA) and blame (NN) show similar variation trends, opposite to peace (PE). Examining monthly combined sentiment component changes for each year from 2019-2023 in Dataset A reveals that in May (approaching the Gaokao) and June (Gaokao month), happiness (PA) and wish (PK) increase, while blame (NN) components are higher in August when exam scores are released.

Focusing on January and June and comparing with sentiment component analysis results from Dataset B shows that compared to January, texts from June 5 (immediately before the Gaokao) and June (during and after the Gaokao) contain higher components of happiness (PA), wish (PK), and panic (NI), but lower components of blame (NN) and doubt (NL). This aligns with common sense and demonstrates that Weibo, as an online media and social platform, still follows real-world logic in textual sentiment variation to some extent.

### Regional Difference Exploration

We selected five macro-level indicators reflecting local economic development and educational investment: per capita gross regional product, per capita disposable income of residents, per capita education expenditure, high school student-teacher ratio, and regular higher education institution student-teacher ratio (all for 2019-2020). These were correlated with the nine key sentiment components in Weibo texts by region.

**Dataset A (2019-2020 data):** The analysis shows per capita education ex-

penditure is most closely related to sentiment components. Trust (PG), wish (PK), fear (NC), and doubt (NL) are significantly positively correlated with it, while panic (NI), annoyance (NE), and blame (NN) also show some correlation. This may indicate that regions with higher per capita education expenditure have societies and families that value education more, correspondingly attaching greater importance to the Gaokao and more likely expressing stronger emotions such as trust, wishes, fear, and even doubt online, with correspondingly stronger feelings of panic and annoyance. This manifests in texts through social and family values, and may also result from associations between education expenditure and economic development level as well as internet penetration and usage depth. For example, families with high income and strong educational values have high education expenditures for their children, high importance attached to the Gaokao, and economic advantages in internet usage depth for both parents and children.

Additionally, peace (PE) is negatively correlated with per capita gross regional product and disposable income, though the correlation is not strong or particularly significant. This may be because economically less developed regions have more moderate public attitudes toward the Gaokao, with social attitudes being more gentle. The social texts posted about the Gaokao reflect more simple ideas that do not focus on various negative topics such as fairness-related exam regulations and cheating behaviors. Attention to and participation in such discussions may evoke trust, fear, doubt, and other emotions, thereby reducing the proportion of peace components.

**Dataset B:** Due to insufficient data from only four days in 2019-2020, we used all data combined for correlation analysis. The analysis shows peace (PE) is significantly positively correlated with the first two indicators, while per capita education expenditure shows almost no significant correlation with any sentiment component. The correlation results from Dataset A thus appear unstable. However, since Dataset B consists of single-day crawls with many contingent factors and high randomness, and represents averaged data from 2019-2024 correlated with 2019-2020 macro data, it carries a certain predictive nature. Therefore, these analytical results cannot overturn the conclusions drawn from Dataset A's analysis.

## Discussion

This study systematically explores temporal-spatial variation patterns in Gaokao public opinion attention and sentiment components by integrating social media data, sentiment analysis, and local macro-level indicators, partially addressing previous research limitations of insufficient temporal span and incomplete spatial coverage. Through text processing and sentiment analysis of Weibo big data, we find that different platforms (Baidu, Weibo) and different regions exhibit varying user cognitive tendencies and internet usage habits that affect attention and emotional expression. Regional economic development level, educational investment, and internet usage depth also

influence the spatial distribution and sentiment components of Gaokao-related public opinion. The multi-data-source, multi-dataset comparative analytical approach employed in this study also provides a reference for future research on public opinion patterns regarding major social public issues.

Specifically, for longitudinal attention research, Datasets A and B are limited by crawling methodology and cannot reflect genuine temporal attention variation. Baidu Index, as a reliable indicator reflecting temporal attention variation, shows highly homogeneous search index changes across regions, suggesting that Gaokao attention on the Baidu platform is primarily driven by the Gaokao itself, with minimal influence from the new Gaokao reform. This reminds us that promoting the new Gaokao reform requires full consideration of local economic levels and social emotions, with targeted publicity and public opinion work to ensure policy implementation fairness and public acceptance. Cross-sectional attention research shows that applicant numbers and local fiscal education expenditure correlate strongly with Baidu Index but weakly with Dataset A, while per capita gross regional product shows opposite patterns across the two data sources. This may reflect user group differences across platforms. Sentiment component analysis of Weibo text data indicates that sentiment variation trends align with the real-world logic of the Gaokao as a specific social hotspot, reflecting public emotional fluctuations during major events. Regional economic and educational investment levels affect sentiment components, but different emotional categories show distinct influence patterns. For instance, the indicator “per capita education expenditure” significantly influences sentiment components in Gaokao-related public opinion, while peace (PE) components are lower in economically developed regions, possibly related to their cognitive characteristics.

Despite revealing numerous patterns in public opinion attention and sentiment components, this study still has several limitations. First, Weibo text collection methods and analysis strategies may cause certain data biases. The opacity of Baidu Index algorithms, potential issues with crawler functionality, and the crawling approach of existing databases all limit “unbiased sampling” of Weibo text data. Particularly, the anomalous entry counts in Dataset A for 2022 and 2023 undermine the reliability and interpretability of sentiment analysis results derived from Dataset A from a side perspective, greatly limiting its longitudinal analysis with macro-level indicators.

Furthermore, sentiment analysis relies on dictionary coverage rates. The dictionary coverage rate of texts used in this study is generally below 0.1 overall, with some temporal-spatial text ranges approaching 0.05, making the analysis insufficiently reliable. On the other hand, with internet language constantly and rapidly evolving, and youth group “internal languages” emerging endlessly, sentiment lexicons compiled based on real-life applicability may have limited suitability for Weibo texts and the specific Gaokao context.

Future big data research on periodic social hotspot issues like the Gaokao should further expand data sources, such as integrating data from internet social plat-

forms like WeChat, Xiaohongshu, and Zhihu, to explore differences in attention and sentiment components across different user groups. Machine learning models could be employed for more refined sentiment component prediction to improve sentiment analysis accuracy and breadth. Additionally, if multi-platform tracking studies can be conducted to explore associations between Gaokao public opinion and local development and social emotions over longer time spans and across more social media platforms, a more complete research framework could be constructed to further understand the formation mechanisms of Gaokao public opinion and its relationship with socioeconomic factors, providing more robust data support for relevant policy-making.

**Declaration of Generative AI Technology Application:** After completing this manuscript, the authors used ChatGPT-4o to polish portions of the text. The generated content was reviewed and edited as needed and manually revised for inclusion in the paper.

## References

- Wang, R. (2023). *Sentiment Analysis Technology of Chinese Weibo under the Background of Big Data*.
- Xiao, Y., Li, B., & Gong, Z. (2018). Real-time identification of urban rainstorm waterlogging disasters based on Weibo big data. *Natural Hazards*, 94(2), 833-847.
- Zhao, X., & Wang, X. (2023). Dynamics of Networked Framing: Automated Frame Analysis of Government Media and the Public on Weibo With Pandemic Big Data. *Journalism & Mass Communication Quarterly*, 100(1), 100-122.
- Chang, J., & Li, J. (2021). A study on the temporal and spatial differentiation of COVID-19 and public anxiety: An analysis based on Weibo data. *Human Geography*, 36(03), 47-57+166.
- Chen, X., Chang, T., Wang, H., Zhao, Z., & Zhang, J. (2020). Temporal and spatial analysis of public opinion on “COVID-19” based on Weibo data. *Journal of Sichuan University (Natural Science Edition)*, 57(02), 409-416.
- Dai, Y. (2023). Artificial intelligence network public opinion analysis based on Weibo data: A case study of ChatGPT topic. *Communication and Copy-right*(21), 93-95.
- Du, Z., Wang, S., Wang, Q., & Meng, X. (2017). A survey on big data analysis of social media. *Computer Science and Exploration*, 11(01), 1-23.
- Fan, X. (2023). *Public Opinion Analysis Based on Weibo Data Under Public Emergencies* [Master’s thesis].
- Huang, F., Li, H., Ding, H., Wu, S., Liu, M., Liu, T., Liu, X., & Zhu, T. (2020). The impact pattern of economic development on collective morality from the

perspective of social media big data. *Chinese Science Bulletin*, 65(19), 1979-1989.

Liu, H. (2009). Ideas, principles, and policy recommendations for Gaokao reform. *Educational Research*, 30(07), 3-7.

Wang, D. (2021). *Dalian University of Technology Chinese Sentiment Lexicon Ontology*.

Wu, S., Mao, Y., Wu, S., Feng, J., Zhang, Q., Xie, T., Chen, H., & Zhu, T. (2023). Cultural-psychological analysis based on big data. *Advances in Psychological Science*, 31(03), 317-329.

Wu, Z. (2020). The dilemma of big data analysis and solutions from a contextualism perspective. *Theory Research*(03), 60-62.

Xiao, J., Wang, K., Liu, M., Shangguan, F., & Zhu, T. (2021). The relationship between kindergarten entry anxiety and maternal emotional characteristics: A study based on Sina Weibo. *Chinese Journal of Women and Children Health*, 12(03), 11-15.

Yan, K., & Wu, H. (2022). Realistic dilemmas and breakthrough strategies of the new Gaokao reform. *China Higher Education Research*(02), 13-21.

Yang, X., Guo, L., Jin, X., & Gu, J. (2019). Big data assisting new Gaokao reform: Framework design and implementation path. *E-Education Research*, 40(02), 30-37.

Zhang, F., & Gan, H. (2020). A study on the impact of psychological temporal-spatial distance on public emotions during the pandemic: A computational analysis based on panel data from Weibo texts during COVID-19. *Press Circles*(06), 39-49.

Zhang, L. (2024). Research on the interactive relationship between online public opinion and Gaokao policy: A case analysis based on Jiangsu's new Gaokao subject selection public opinion management. *China Examinations*(05), 28-35.

Zhong, B., & Wang, X. (2019). Realistic dilemmas, rational compliance, and strategic choices of the new Gaokao. *Journal of Educational Studies*, 15(05), 62-69.

Zhou, X. (2021). Family capital and academic anxiety: A discussion on parental anxiety issues triggered by the "Double Reduction" policy. *Journal of Guangxi Normal University (Philosophy and Social Sciences Edition)*, 57(06), 96-106.

Zhu, T., Bai, S., & Li, A. (2013). Group event risk warning based on social media big data. *Psychology and Innovation Capacity Enhancement—The 16th National Psychology Academic Conference*, Nanjing, Jiangsu, China.

(Corresponding author: Zhu Tingshao, Email: tszhu@psych.ac.cn)

<sup>1</sup> Lincui Analysis System

<sup>2</sup> The sentiment component data used hereinafter represent simple summation of proportional values (for missing items or total component value comparisons) or arithmetic mean (for inter-component comparisons with missing items), having comparative meaning only within each chart.

<sup>3</sup> This ontology and introductory text are taken from the Dalian University of Technology Chinese Sentiment Lexicon Ontology - Heywhale.com (<https://www.heywhale.com/mw/dataset/611383fc911b33001748107d>).

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*