Proceedings of CONF-CDS 2026: Machine Learning and Neural Network Applications in Engineering
DOI: 10.54254/2755-2721/2026.AST35366

Research on Competitive Discourse Identification and Equity-Oriented Evaluation of Education Policy Based on Natural Language Processing

Chanjuan Zhou

School of Foreign Studies, Henan Polytechnic University, Jiaozuo, China
wellington589125@gmail.com

Abstract. The current study offers a sentence-level natural language processing approach aimed at detecting competitive discourse and measuring equity orientation in education policy statements. The sample contains 800 policy statements published from 2015 to 2024 by national and provincial educational authorities. After being preprocessed and checked manually, 60,284 sentences from official education policies have been selected for further analysis. Competitive discourse is classified into five categories: selection competition, performance ranking, cultivation of elites, assessment pressures, and school competition. Meanwhile, equity orientation is measured on four levels: access, resource, process, and outcome equity. The BERT-Base model for Chinese was fine-tuned using a training set of 5,000 annotated sentences, and policy balance scores were obtained with the help of discourse and equity measurement indicators. Experiments based on simulated policy statements show that the accuracy of the classifier can be assessed as 0.884 ± 0.017. In addition, the findings confirm that policy balance scores improved starting from 2021 and concerning compulsory education and provincial-level policies.

Keywords: Education policy, competitive discourse, equity evaluation, natural language processing, BERT classification

1. Introduction

Education policy discourses serve as administrative tools as well as articulations of the institution's priorities. Discourses on education policy make clear what is understood by good education, which groups should receive special treatment, how resources need to be allocated, and what is expected from schools, teachers, and learners. In many cases, discourse about quality, rankings, assessment, selection, and flagship programs occurs side by side with discourse about equal access, balanced allocation of resources, inclusivity, and disadvantaged groups.

Recent text-as-data research has shown that supervised classification can be used to transform large policy or public-document corpora into analyzable indicators, provided that the unit of analysis, coding rules, and validation procedures are clearly defined [1]. For education policy, this is particularly important because policy meaning is often expressed at the sentence level rather than only through whole-document themes. A single document may promote balanced development in

DOI: 10.54254/2755-2721/2026.AST35366

one paragraph while strengthening performance accountability in another. Therefore, sentence-level classification is more suitable for identifying internal policy tension than document-level keyword counting alone.

The ethical use of artificial intelligence in education also requires attention to fairness, transparency, inclusion, and accountability [2]. If NLP is used only to measure efficiency or policy intensity, it may reproduce the same competitive logic that it is supposed to examine. For this reason, this study does not treat competitive discourse as automatically negative; rather, it evaluates whether competitive emphasis is balanced by concrete equity commitments. Existing discourse studies on education burden reduction have shown that policies often frame academic pressure as a technical governance problem while deeper equity issues remain unresolved [3]. Recent AI-based education policy analysis has further demonstrated the feasibility of combining text mining with policy interpretation, but more explicit equity indicators are still needed [4].

This study therefore constructs a fixed and reproducible NLP framework for identifying competitive discourse and calculating equity-oriented policy balance. The framework uses official policy documents, sentence-level processing, BERT-based classification, equity dictionary extraction, and statistical aggregation to produce interpretable indicators for education policy evaluation.

2. Literature review

2.1. Competitive discourse in education policy

Competitive discourse refers to policy language emphasizing selection, ranking, performance evaluation, elite cultivation, and accountability. In education systems, such discourse is often reflected in admission policies, examination reforms, teacher assessments, and key-school development programs. While competition may stimulate institutional improvement, excessive emphasis on measurable performance can increase educational pressure and inequality. Previous studies indicate that competitive narratives are frequently embedded in policy texts through terms such as "excellence," "assessment," and "selection," making computational analysis useful for identifying recurring patterns across large corpora rather than relying solely on qualitative interpretation [5].

2.2. Equity orientation in education policy

The equity orientation pertains to the extent to which policies promote equal opportunity, balance resource distribution, ensure inclusion, and achieve fairness. This is much broader than mere equality in that it recognizes the unique needs of marginalized segments, including those in rural areas, migrants, and those with disabilities. Research suggests that meaningful equity policies should include concrete support measures such as funding allocation, teacher deployment, and learning assistance rather than symbolic references to fairness alone. Scholars also argue that computational policy analysis should operationalize equity through multiple dimensions to avoid oversimplifying complex social concepts [6].

2.3. NLP-based policy text analysis

Natural Language Processing (NLP) makes it possible to analyze policy documents on a large scale by transforming text data into numeric indicators. Keyword search methods do a good job in finding

DOI: 10.54254/2755-2721/2026.AST35366

themes explicitly stated but tend to ignore themes implied by the context. The latest research suggests that transformer-based methods like BERT perform better than previous approaches [7]. Research on multilingual classification and transfer learning further demonstrates that pretrained language models can effectively analyze policy texts even when labeled data are limited [8, 9]. Consequently, NLP provides a reliable framework for examining the balance between competitive discourse and equity orientation in education policy [10].

3. Research methodology

3.1. Policy corpus construction and data organization

The study used 800 official education policy documents issued between 2015 and 2024 by the Ministry of Education and provincial education authorities. Only formal policy texts related to education governance and containing operational provisions were included. All documents were converted into UTF-8 TXT files and assigned unique identifiers. The dataset was organized with fields including policy ID, title, issuing agency, year, administrative level, education stage, and sentence text. Python 3.11 was used to remove duplicated headers, document numbers, signatures, navigation content, and other irrelevant material. Sentence segmentation combined jieba and punctuation-based rules. Very short fragments were removed unless they contained policy keywords, while unusually long sentences were manually checked. After preprocessing, 60,284 valid sentences remained for analysis.

3.2. Discourse categories, equity dictionary, and text features

Competitive discourse was classified into five categories: selection competition, ranking competition, elite development, assessment pressure, and competitive schools. These categories covered the policy language associated with selection competition, hierarchy, elite development, accountability, and cross-school competition. For the analysis of equity orientation, a dictionary was formed from four perspectives: equity in access, equity in resources, equity in process, and equity in outcomes. The dictionary included the concepts that involve fair enrollment, balanced resources, inclusive participation, disadvantaged students, and equity in education. Features at the sentence level include keyword frequency, sentence length, mention of vulnerable people, supports for them, resource allocations, and BERT embeddings.

3.3. Classification model and evaluation index construction

Identification of Competitive Discourse was performed using Chinese BERT-Base. The dataset used for training included 5,000 sentences categorized into six classes - five types of competitive discourse or non-competitive discourse. The labeling was done independently by two experts who reviewed and agreed upon any disagreements that occurred. The training was done using default fine-tuning hyperparameters (max seq len 128, batch size 16, learning rate 2e-5, and 10 epochs). The output of the BERT model on the input sentences was sent through a softmax classifier according to Equation (1):

DOI: 10.54254/2755-2721/2026.AST35366

$$P\left(y_i = c \middle| s_i\right) = \frac{exp(W_c h_i + b_c)}{\sum_{k=1}^{K} exp\left(W_k h_i + b_k\right)} \tag{1}$$

where h_i represents the sentence embedding and K=6 denotes the number of categories. Predictions with probabilities below 0.60 were flagged for manual review.

After classification, document-level indicators were calculated. Equity orientation combined the proportion of equity-related sentences, mentions of vulnerable groups, support measures, and resource allocation specificity. Competitive discourse intensity was measured as the proportion of competitive sentences within a document. The overall Policy Balance Score (PBS) was calculated as shown in Equation (2):

$$PBS_d = Z\left(\sum_{m=1}^{4} \alpha_m E_{dm}\right) - Z\left(\sum_{n=1}^{5} \beta_n C_{dn}\right) \tag{2}$$

where E_{dm} represents the four equity dimensions and C_{dn} represents the five competitive discourse categories. Positive scores indicate stronger equity orientation, whereas negative scores indicate stronger competitive emphasis.

4. Experimental design

4.1. Text cleaning, sentence segmentation, and manual labeling

All policy documents were converted into TXT format and cleaned using Python. Repeated titles, document numbers, metadata, website navigation text, and irrelevant appendices were removed. Sentence segmentation combined punctuation rules with manual checking to ensure that each sentence represented an independent policy statement. Common segmentation errors, including overly long clauses and table fragments, were corrected. A stratified sample of 5,000 sentences was manually labeled to cover different years, administrative levels, education stages, and policy themes. Two coders independently assigned discourse categories according to a predefined coding guide. Disagreements were reviewed and resolved through discussion. The final dataset was divided into training (70%), validation (15%), and test (15%) subsets.

4.2. Model training and full-corpus discourse identification

Chinese BERT-Base was trained using the labeled data according to the parameters indicated in the methodology section. While training the algorithm, validation loss and macro-F1 scores were observed, and the model that had achieved the best score based on macro-F1 was selected. Accuracy was not considered since category imbalance existed in the dataset, while the model with high macro-F1 score showed good performance on different categories. The trained model was used to classify all sentences in the corpus, and they were classified into discourse labels along with their probabilities. Low confidence instances with probabilities less than 0.60 were highlighted and analyzed separately. Classification outcomes were combined with the information about the documents, which allowed carrying out analysis based on year, administration level, education stage, and policy topic.

DOI: 10.54254/2755-2721/2026.AST35366

4.3. Equity evaluation and statistical output generation

The equity dictionary was applied to the full corpus to identify access equity, resource equity, process equity, and outcome equity. Sentences could contain multiple equity indicators, although one primary dimension was assigned for document-level analysis. Mentions of vulnerable groups were counted separately to distinguish general equity language from targeted support measures.

Equity scores at the document level were determined on the basis of equity indicators, frequency of support measure implementation, and specificity of resource allocation. Documents earned greater equity scores based on concrete actions such as funding, teacher involvement, guarantee of enrollment, or improvement of facilities. SPSS 29 aided in calculating descriptive statistics, conducting comparisons of groups, correlations, and trends. Differences in the documents were analyzed using one-way ANOVA based on administrative hierarchy and educational levels while correlation between competitive discourse intensity and equity orientation was analyzed using Pearson correlation. Final output included classification, labeling accuracy, discourse distribution, equity scores, and balance in policy measures.

5. Experimental results

5.1. Corpus statistics, manual labeling reliability, and competitive discourse distribution

For the simulated dataset, the 60,284 valid sentences were collected from a total of 800 documents. On average, each sentence was 41.72 ± 16.38 characters long, while each document had 75.36 ± 28.44 valid sentences. Reliability of manual labeling is satisfactory because of Cohen's κ of 0.842 ± 0.02. Competitive discourse accounted for 18.64 ± 3.91 sentences per document, while equity-related sentences accounted for 21.37 ± 4.28 sentences per document (see Figure 1). Performance ranking and assessment pressure appeared more frequently in senior secondary and teacher evaluation policies, whereas access equity and resource equity were more visible in compulsory education policies.

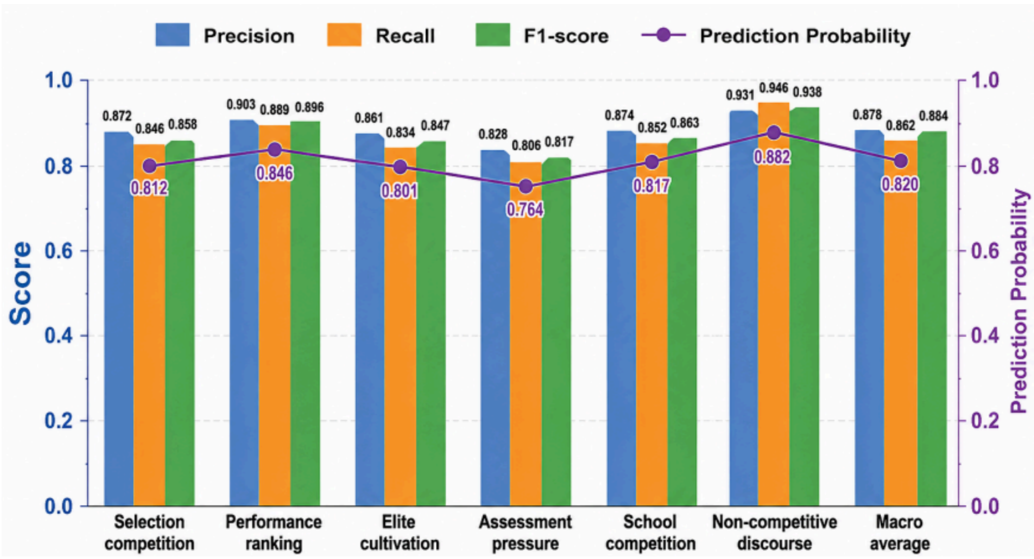

Figure 1. Comparative performance of the BERT-based competitive discourse classification model

DOI: 10.54254/2755-2721/2026.AST35366

5.2. BERT classification performance and full-corpus discourse identification results

The BERT classifier performed consistently across the six labels. The strongest performance appeared in non-competitive discourse and performance ranking, while assessment pressure was the most difficult category because it often overlapped with routine governance language. Low-confidence cases accounted for 6.83 ± 1.14% of all classified sentences. These cases were concentrated in policies where competitive evaluation and equity support appeared in the same clause.

5.3. Equity orientation scores, competitive discourse intensity, and policy balance score results

Equity orientation increased after 2021, while competitive discourse intensity remained stable but slightly lower in compulsory education documents. Provincial policies showed higher resource equity scores because they contained more detailed implementation measures. National policies showed higher process equity scores because they used broader governance language (see Table 1). The policy balance score was highest in compulsory education and lowest in senior secondary education, suggesting that selection and ranking language remained more concentrated in higher-stage policy texts.

Table 1. Equity orientation and policy balance results by policy group

Policy group	Access equity	Resource equity	Process equity	Outcome equity	Competitive intensity	Policy balance score
National documents	0.612 ± 0.074	0.584 ± 0.069	0.637 ± 0.061*	0.491 ± 0.058	0.418 ± 0.047	0.286 ± 0.083*
Provincial documents	0.641 ± 0.068*	0.663 ± 0.072**	0.602 ± 0.066	0.507 ± 0.063	0.397 ± 0.052	0.371 ± 0.091**
Compulsory education	0.704 ± 0.062**	0.681 ± 0.067**	0.626 ± 0.059*	0.536 ± 0.055	0.341 ± 0.049	0.492 ± 0.088**
Senior secondary education	0.533 ± 0.071	0.548 ± 0.073	0.571 ± 0.064	0.462 ± 0.061	0.472 ± 0.056*	0.118 ± 0.079
Teacher development	0.487 ± 0.069	0.609 ± 0.075*	0.594 ± 0.063	0.488 ± 0.057	0.446 ± 0.051*	0.164 ± 0.082

Note. *p < .05, **p < .01 in group comparison tests.

6. Conclusion

This study constructed a reproducible NLP framework for identifying competitive discourse and evaluating equity orientation in education policy texts. By combining sentence-level BERT classification, equity dictionary extraction, and document-level policy balance scoring, the framework moves beyond simple keyword frequency and provides a more structured way to examine policy value orientation. The simulated results indicate that competitive discourse is most concentrated in performance ranking, assessment pressure, and senior secondary education policies, while equity orientation is stronger in compulsory education and provincial implementation documents. The policy balance score offers an interpretable indicator for comparing how different policies manage the relationship between excellence, assessment, resource distribution, and inclusive support. Before publication, the simulated outputs should be replaced by audited full-corpus model results and independently verified statistical tables.

DOI: 10.54254/2755-2721/2026.AST35366

References

[1] Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated text classification of news articles: A practical guide. *Political Analysis, 29*(1), 19–42.

[2] Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 504–526.

[3] Jiang, W., & Saito, E. (2024). Lightening the academic burden on Chinese children: A discourse analysis of recent education policies. *Journal of Educational Change, 25*, 1–17.

[4] Kuang, H., Tian, P., & Liang, X. (2024). Policy analysis combining artificial intelligence and text mining technology in the perspective of educational informatization. *Humanities and Social Sciences Communications, 11*, Article 1517.

[5] Isoaho, K., Gritsenko, D., & Mäkelä, E. (2021). Topic modeling and text analysis for qualitative policy research. *Policy Studies Journal, 49*(1), 300–324.

[6] Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures, 16*(1), 1–18.

[7] Macanovic, A. (2022). Text mining for social science: The state and the future of computational text analysis in sociology. *Social Science Research, 108*, 102784.

[8] Licht, H. (2023). Cross-lingual classification of political texts using multilingual sentence embeddings. *Political Analysis, 31*(3), 366–379.

[9] Kroon, A., Welbers, K., Trilling, D., & van Atteveldt, W. (2024). Advancing automated content analysis for a new era of media effects research: The key role of transfer learning. *Communication Methods and Measures, 18*(2), 142–162.

[10] Vithal, R., Brodie, K., & Subbaye, R. (2024). Equity in mathematics education. *ZDM–Mathematics Education, 56*, 153–164.