

Computational Modeling of Educational Theory in Low-Socioeconomic Contexts

Jadon Swearingen¹, Mustafa Ocal², Md Tarique Hasan Khan³, Labiba Jahan¹

¹Southern Methodist University, Dallas, TX 75205

²Florida International University, Miami, FL 33199

³Tarleton State University, Stephenville, TX 76402

Abstract

This study examines narratives in which students describe challenges they faced in higher education due to low socioeconomic (SES) backgrounds and the strategies they used to overcome them. Using computational text analysis, we operationalize three educational theories, Paulo Freire’s Critical Pedagogy, Urie Bronfenbrenner’s Ecological Systems Theory, and Pierre Bourdieu’s Theory of Capital and Habitus to analyze patterns in these narratives. To strengthen the theory-to-method connection, we incorporate temporal timeline extraction, identifying ordered event sequences and tracking how challenges and forms of capital evolve across a student’s posting history. This temporal lens links theoretical categories (barriers, supports, forms of capital) to when they occur, highlighting moments for timely interventions. By combining theory-driven features with temporal analysis, we evaluate the explanatory capacity of each framework and demonstrate how computational methods can quantitatively examine qualitative lived experience at scale, supporting interdisciplinary research.

1 Introduction

Socioeconomic status (SES) describes a person’s social and economic position, often based on factors like income, education, and occupation (Saegert et al., 2006). SES can shape access to resources, opportunities, and overall quality of life. Students from low-SES backgrounds frequently face additional obstacles in higher education, which can affect their academic success, mental health, and career goals. While researchers in the social sciences particularly education have studied these challenges and suggested potential solutions, STEM fields including Computer Science (CS) have made less progress in addressing these disparities (Kelbessa et al., 2024). Bringing together insights from education and CS can provide a deeper understanding of the barriers low-SES students

face and support collaborative, interdisciplinary approaches to creating meaningful interventions. This study seeks to contribute to that effort.

The landscape of education stands at a critical inflection point, driven by the dual forces of technological advancement and a pressing need for greater equity and inclusion. While significant efforts have been made to diversify the pipeline, students from low socioeconomic status (SES) backgrounds remain persistently underrepresented and face unique, systemic barriers to success (Ohland et al., 2011). The challenge of supporting these students is not merely a moral imperative but an economic one, as broadening participation is essential for fostering innovation and meeting the complex workforce demands of the future (Engineering Education–Workforce Continuum Committee, 2019). Traditional research methods for understanding these barriers have largely relied on quantitative data from surveys and institutional records, which, while valuable, often fail to capture the nuanced, lived experiences of students (Secules et al., 2021). These approaches can reduce complex socioeconomic challenges to easily measurable variables, potentially overlooking the rich contextual factors such as the “hidden curriculum”, “social capital deficits”, and “psychological stress” that profoundly impact a student’s journey (Meyers et al., 2012). As a result, educational interventions, while well-intentioned, may not fully address the root causes of attrition among low-SES students.

Building on this foundation, the present study moves beyond conceptual argument to empirical application. While prior work has established the theoretical imperative for integrating critical frameworks with computational methods (Khan et al., 2026), no study to date has systematically operationalized multiple educational theories at scale to analyze low-SES student narratives. This paper addresses that gap by applying a computational coding framework grounded in Critical Pedagogy

(Freire, 2020), Ecological Systems Theory (Bronfenbrenner, 1979), and Bourdieu’s Theory of Capital and Habitus (Bourdieu, 1986) to a corpus of 387 student narratives. By transforming theoretical constructs into computationally identifiable features, we provide the first large-scale, quantitative mapping of how these theories manifest in students’ lived experiences. Furthermore, we introduce narrative timeline extraction for tracing the temporal progression of students’ challenges and capital trajectories, offering a dynamic view of how low-SES students navigate education over time. This work contributes to education and computer science research by demonstrating how computational methods can bridge the gap between qualitative lived experience and quantitative generalizability, advancing both methodological innovation and theoretical understanding of socioeconomic inequity.

We also incorporate temporal timeline extraction to strengthen the paper’s theory-to-method connection by operationalizing key theoretical constructs, especially trajectory and transition points, in students’ narratives. By extracting ordered event sequences within posts (e.g., financial shock → increased work hours → missed class → academic consequences) and, when possible, tracking how challenges and forms of capital evolve across a student’s posting history, timelines make it possible to observe how barriers emerge, compound, or resolve over time. This temporal lens complements our use of educational theory (e.g., community cultural wealth) by linking theoretically grounded categories (barriers, supports, forms of capital) to when they occur in the narrative, supporting the paper’s precision-intervention logic by highlighting moments where targeted support may be most timely and impactful.

The remainder of this paper is organized as follows. In §2, we review related work. In §3, we describe the dataset and the educational theories examined in this study. Section §4 outlines our methodology, including the experimental setup for theory testing and timeline experiments. We then present our findings in §5 and discuss the results in §6. Section §7 describes sources of error and directions for future work, followed by a discussion of the study’s limitations in §8. Finally, we summarize our contributions in §9 and discuss the ethical and societal implications in §10.

2 Related Work

Impact of Low-SES Background in Education: Socioeconomic inequity in education has been extensively documented, with students from low-SES backgrounds facing persistent barriers that extend beyond financial constraints to include cultural isolation, limited professional networks, and psychological burdens such as imposter syndrome (Ohland et al., 2011; Martin et al., 2013). To understand these multifaceted challenges, education researchers have drawn upon three foundational theoretical frameworks. Critical Pedagogy has been employed to frame student narratives as expressions of agency and critique rather than deficit, revealing how marginalized students name and resist systemic oppression (Freire, 2020; Riley, 2022; Secules et al., 2021). Ecological Systems Theory has provided a spatial vocabulary for locating student challenges across nested contexts from immediate family and work environments (microsystem) to institutional policies (exosystem) and cultural ideologies (macrosystem) demonstrating how pressures at one level amplify difficulties at another (Kahu and Picton, 2022; Bronfenbrenner, 1979). Most influentially, Bourdieu’s Theory of Capital and Habitus has enabled researchers to analyze how inequality operates through multiple resource dimensions: economic capital (financial means), cultural capital (institutionally valued knowledge), social capital (networks and relationships), and symbolic capital (recognition and belonging), while Yosso’s critique has extended this framework to recognize the “community cultural wealth” marginalized students bring (Bourdieu, 1986; Yosso*, 2005). Despite the influence of each framework, they have been applied separately to different populations using different methods, making it impossible to assess their relative explanatory power. The present study addresses this gap by operationalizing all three theories on a common corpus, enabling the first systematic comparison of how each illuminates the lived experiences of low-SES students.

Computer Science Approaches to Research on Low-SES Students: Contributions from Computer Science researchers in this area remain limited (Kelbessa et al., 2024). Earlier work by Lee et al. (2008) examined the educational development of low-SES students using data from the National Educational Longitudinal Study and applied categorical regression to analyze patterns in the data. Robinson (2020) investigated STEM intentions

among low-SES students at a high school in California, combining quantitative analysis via logistic regression with qualitative insights from in-depth interviews. Kelbessa et al. (2022) explored higher education contexts, identifying patterns in low-SES student experiences through analysis of social media data. Recently, Abdelgaber et al. (2025) developed a semi-automated dataset of narratives written by low-SES students collected from Reddit, which serves as the foundation for the current study.

TimeML Annotation and Timeline Extraction: TimeML is a widely used specification for annotating temporal information in text, designed to support machine-readable representations of when events happen and how they relate in time (Sauri et al., 2006). It represents events, time expressions (e.g., dates or durations), and temporal relations (e.g., BEFORE/AFTER/INCLUDES), enabling document-level temporal graphs for tasks like narrative understanding, temporal question answering, and timeline construction (Liu et al., 2012; Saquete et al., 2004; Navarro-Colorado and Saquete, 2016). Several automated TimeML pipelines (e.g., TARSQI, CAEVO, ClearTK, CATENA) extract events/times and predict temporal links (Verhagen et al., 2005; Chambers et al., 2014; Bethard, 2013; Mirza and Tonelli, 2016), but their outputs can be structurally flawed. Ocal et al. (2022) report that about one-third of automatically generated TimeML graphs are temporally inconsistent, and they contain roughly 20% more temporal indeterminacy than gold annotations, hurting downstream timeline construction. More recently, large language models have improved temporal extraction, with evidence that recent GPT models can better capture implicit temporal cues than traditional systems (Eirew et al., 2025; Wang and Weiss, 2025). Given TimeML annotations, a natural next step is timeline extraction: converting a set of temporally constrained events into an ordered event sequence that better reflects narrative progression. This conversion is non-trivial because TimeML graphs may be incomplete, underspecified, or inconsistent, and therefore require principled inference and constraint handling. The TLEX algorithm addresses this problem by first converting TimeML annotations into a temporal graph, and then transforming that representation into a temporal constraint graph. Timeline construction is performed by solving the resulting constraints using a constraint satisfaction problem (CSP) solver (Ocal et al., 2026).

3 Data and Theories

Data: We used the corpus introduced by Abdelgaber et al. (2025), consisting of Reddit narratives written by students from low-SES backgrounds. The dataset captures students’ lived experiences in higher education and was created using a hybrid approach combining machine learning and human validation, expanding a small seed set into a larger, high-quality corpus.

The dataset consisted of English Reddit posts collected from U.S.-centric subreddits related to low-SES and student experiences, including communities such as r/college, r/ApplyingTo-College, r/financialaid, and r/fafsa. In the data collection process, strict inclusion criteria ensured relevance: narratives had to describe low-SES experiences, challenges faced, and responses to those challenges. General commentary or advice without personal experience was excluded. Since Reddit users are anonymous, no personally identifiable metadata about the authors was available. Low-SES status was identified based on self-reported narratives containing phrases such as “I grew up poor”, “My parents had no money”, and “I received a Pell Grant”. The corpus is publicly available at <https://github.com/NahedAbdelgaber/NLP-for-Social-Science/tree/main/CoNLL%202025>.

The semi-automated data collection pipeline began with 128 manually labeled posts and iteratively expanded through classifier-based candidate selection, dimensionality reduction and clustering to remove outliers, and human validation to retain accurate narratives. This process produced a final dataset of 387 narratives, which were further processed with LLaMA to extract structured information on background, challenges, and strategies. Sentiment analysis revealed an increase in neutral and negative posts in the expanded dataset, providing broader representation of student experiences.

Theories: We wanted to use a strong theoretical framework for this corpus to move beyond descriptive observations and generate analytically meaningful insights into the socioeconomic gap in higher education. To achieve this, our study is grounded in a synthesis of three complementary theoretical perspectives drawn from the social sciences, following the theoretical synthesis proposed by Khan et al. (2026). These frameworks provide conceptual structure for interpreting students’ narratives and for systematically mapping lived experiences

onto established theoretical constructs. The Table 1 presents a concise overview of each theory and its role in guiding our analytical approach.

4 Methodology

Computational Coding Framework: To operationalize the three theoretical frameworks at scale, we developed a prompt-based coding system using GPT-4.1 (OpenAI, 2025a). The prompt defined all three frameworks with specific coding categories, provided concrete linguistic examples for each code, required JSON-formatted outputs with evidence citations, and instructed the model to code constructs as absent when they were not explicitly discussed rather than inferring deficits.

The prompting keywords used for computational coding were derived directly from the theoretical constructs operationalized in this study rather than selected post hoc to improve model output. For Critical Pedagogy, the prompt emphasized terms related to structural barriers, oppression, agency, resistance, voice, and critical consciousness, reflecting Freire’s framing of education as a process through which marginalized learners identify and challenge oppressive conditions (Freire, 1970; Riley, 2008; Secules et al., 2021). For Ecological Systems Theory, keywords such as microsystem, mesosystem, exosystem, macrosystem, family, work, institutional policy, financial aid, and cultural expectation were used to locate student experiences across nested social and institutional contexts, consistent with Bronfenbrenner’s ecological model (Bronfenbrenner, 1979; Kahu & Picton, 2022). For Bourdieu’s Theory of Capital and Habitus, the prompt included economic capital, cultural capital, social capital, symbolic capital, belonging, recognition, hidden curriculum, professional networks, and institutional familiarity to identify how socioeconomic disadvantage operates through unequal access to material, cultural, relational, and symbolic resources (Bourdieu, 1986; Yosso, 2005; Martin et al., 2013). Thus, the keywords functioned as theory-grounded coding anchors that translated established educational concepts into computationally identifiable narrative features. They were not designed to force clustering, but to ensure that the language model applied each framework consistently across the corpus.

Because the prompt required the model to assign narratives to predefined theoretical categories rather than generate categories inductively, visual-

izations based on these labels should be interpreted as prompt-guided separations of coded constructs rather than as fully unsupervised discoveries of naturally occurring clusters. GPT-4.1 was selected for the theory-based coding task because the task required consistent, structured classification across the full corpus rather than open-ended temporal reasoning. In preliminary testing, GPT-4.1 produced reliable JSON-formatted outputs for the predefined theoretical categories while allowing the full set of 387 narratives to be processed efficiently.

For Critical Pedagogy, we coded student agency (high/moderate/low), structural barriers (present/absent), and consciousness of barriers (high/moderate/low). For Ecological Systems Theory, we identified the dominant system level (microsystem, mesosystem, exosystem, macrosystem, chronosystem) and secondary influences. For Bourdieu’s Theory, we classified four forms of capital, economic, social, cultural, and symbolic as deficit, asset, or absent. All 387 narratives were successfully processed with 0 errors. Results were exported as structured JSON and parsed into CSV format with full narrative text preserved for validation purposes. The methodology produced a dataset of 387 multi-dimensional theoretical profiles, enabling comparative analysis across frameworks.

Validation and Data Quality: The improved prompt design explicitly distinguishes between narratives that discuss a construct as a deficit versus narratives that do not discuss it at all (absent). This refinement, validated on a sample of 51 narratives before full run execution, improved the model’s ability to reflect actual narrative content without over interpreting missing information. To support reproducibility, the full analysis pipeline, prompts, and data processing scripts used in this study are publicly available at <https://github.com/Jadonsw123/low-ses-narrative-analysis>.

Applying TLEX: We applied TLEX to construct explicit temporal structure from students’ posts by combining LLM-based TimeML annotation with constraint-based timeline extraction. Specifically, we used few-shot prompting with GPT-5.2 Thinking (OpenAI, 2025b) to produce TimeML-style annotations for each post. GPT-5.2 Thinking was used for the TimeML annotation task because temporal extraction requires more complex reasoning over implicit event ordering, temporal relations, and narrative sequence. For the few-shot exemplars, we selected fully TimeML-annotated documents from the AQUAINT corpus (Consortium

Theory	Core Idea	Function in This Work	Relevance to Low-SES Student Narratives
Critical Pedagogy (Freire)	Education is a political act that can reproduce or challenge inequality.	Frames the study as an act of empowerment amplifying student voices rather than diagnosing deficits.	Positions low-SES students as agents whose narratives expose structural barriers in education.
Ecological Systems Theory (Bronfenbrenner)	Student experiences are shaped by nested systems (micro → macro).	Provides a multi-layered lens to interpret narrative data across personal, institutional, and societal contexts.	Helps map each narrative theme (family, work, policy) to specific systemic layers influencing student outcomes.
Bourdieu’s Theory of Capital & Habitus (Bourdieu)	Inequality persists through economic, cultural, and social “capitals” and internalized dispositions (habitus).	Explains how resource deficits cascade into behavioral, academic, and psychological outcomes.	Reveals how habitus mismatch and unrecognized “community cultural wealth” shape belonging and persistence.
Integrative Role		Combines Critical Pedagogy (why voices matter), Ecological Systems Theory (where challenges occur), and Capital Theory (how inequality operates).	Produces a unified interpretive framework connecting personal experience, systemic structure, and institutional culture.

Table 1: Theoretical Frameworks and Their Analytical Functions

et al., 2002), which provide high-quality demonstrations of how events, temporal expressions, and temporal links are encoded in standard TimeML notation. Given these examples, we prompted GPT-5.2 to annotate each post by identifying events, time expressions, and the temporal relations connecting them. In our experiments, GPT-5.2 generated annotations in a well-formed TimeML format, enabling direct downstream processing.

Next, we fed the resulting TimeML annotations into jTLEX (Ocal et al., 2023, the Java library for the TLEX algorithm). jTLEX first converts the annotations into temporal graphs and performs temporal consistency checking to identify violations in the induced temporal constraints. Finally, using the consistent constraint structure, jTLEX extracts within-post timelines by ordering events and time expressions into a coherent temporal sequence, which we use as the basis for analyzing narrative progression and identifying temporal trajectories in students’ accounts.

5 Results

Critical Pedagogy: Agency and Consciousness: Two key dimensions emerge from the Critical Pedagogy framework: student agency and consciousness of structural barriers. As shown in Figure 1, while 71.6% of students demonstrate high consciousness of the systemic barriers they face, only 48.1% exhibit high agency in responding to those barriers. This 23.5 percentage point gap suggests a distinction between critical awareness and empowered action, echoing Freire’s argument that reflection must be joined with action, or praxis, in struggles against oppression (Freire, 2020).

The clear separation in Figure 1 likely reflects the discrete coding structure imposed by the

prompt, which asked the model to classify agency and consciousness into predefined levels. Therefore, the figure should be interpreted as a visualization of model-coded theoretical categories rather than evidence that the narratives naturally separate into distinct groups without theoretical guidance.

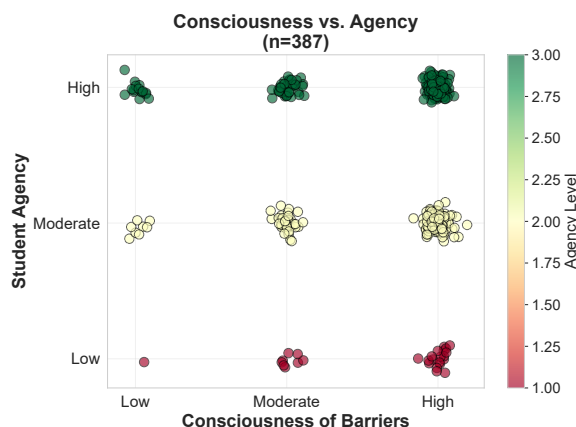


Figure 1: The Awareness Action Gap: Student consciousness of barriers (x-axis) versus student agency (y-axis) reveal that high consciousness is more prevalent than high agency, suggesting that awareness alone is insufficient for consistent empowered action. Each point represents one narrative (n=387).

Additionally, 92.5% of narratives explicitly mention structural barriers such as poverty, institutional policies, discrimination, or family instability. Moderate agency (44.7%) represents students who attempt to navigate barriers repeatedly but encounter constraints, while high agency (48.1%) reflects decisive, sustained action despite obstacles. Low agency (7.2%) appears in narratives expressing resignation or powerlessness. Figure 2 and Figure 3 show the distributions of agency and consciousness respectively.

Ecological Systems: Where Challenges Occur: Bronfenbrenner’s framework reveals that stu-

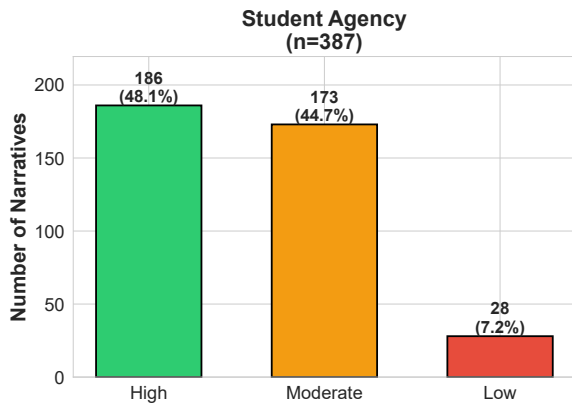


Figure 2: Agency Distribution: Nearly half of students (48.1%) demonstrate high agency, with 44.7% moderate and only 7.2% low, indicating that most low-SES students actively work to overcome barriers despite structural constraints.

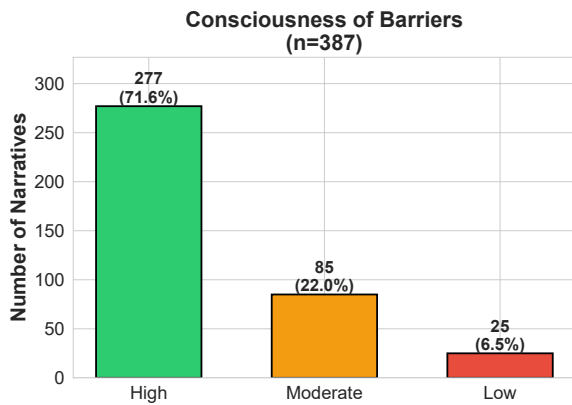


Figure 3: Consciousness of Barriers: The majority of students (71.6%) show high consciousness of systemic barriers affecting their education, aligned with Freire's concept of 'conscientization,' or the awakening to structural injustice.

dents' experiences are dominated by their immediate environments. As shown in Figure 4, the microsystem (family, school, peer relationships, direct work environment) is the dominant influence for 62.8% of narratives, followed by the exosystem (institutional policies, parental employment, financial aid systems) at 24.0%, and macrosystem (societal ideologies, laws, class structures) at 12.9%.

Bourdieu's Capital: Resources and Inequality: The most striking finding emerges from Bourdieu's theory: the distribution of capital types reveals not just deficits, but systematic absence of discussion around certain forms of capital. Figure 6 shows the distribution of students across different combinations of the following capitals.

Economic Capital: 85.8% deficit (as expected for low-SES students), 8.3% asset, 5.9% absent. Few narratives omit money concerns.

Social Capital: Highly polarized with 42.1% ab-

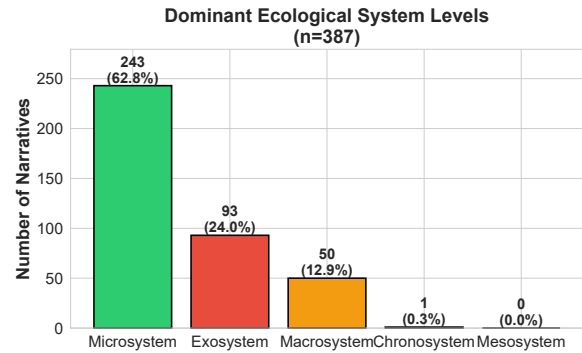


Figure 4: Ecological System Levels: Microsystems dominate (62.8%), indicating that intimate relationships and immediate contexts matter most. However, exosystem factors (policies, indirect institutional influences) affect nearly a quarter of narratives, demonstrating that structural barriers beyond the student's direct control are significant.

sent, 32.6% asset, 25.3% deficit. Many students either describe robust social support or do not mention relationships at all.

Cultural Capital: Nearly split with 46.0% asset (students discuss education, credentials, knowledge), 42.1% absent, 11.9% deficit. This suggests that while many low-SES students build educational capital, others' narratives remain silent on this dimension.

Symbolic Capital: The most absent 48.6%, indicating that students rarely narrate their status, recognition, or sense of belonging. When discussed, 32.3% frame it as deficit (shame, imposter syndrome) and only 19.1% as asset (pride, recognition).

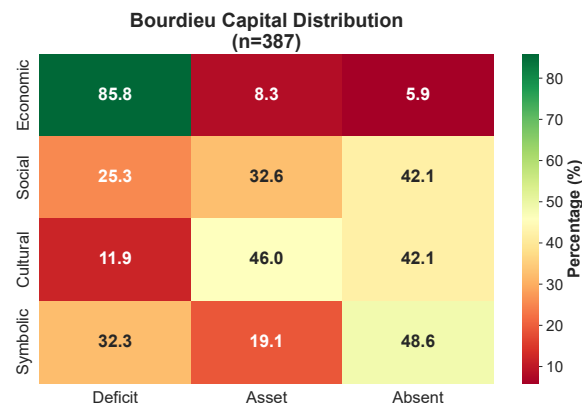


Figure 5: Bourdieu's Capital Distribution (n=387): Economic capital shows near universal deficit (85.8%), confirming low-SES background. However, social capital, cultural capital, and symbolic capital are frequently absent from narratives (42.1%, 42.1%, and 48.6% respectively), indicating students either lack these resources or do not articulate them. Cultural capital is split nearly evenly between asset (46.0%) and absent (42.1%), suggesting education/credentials are either leveraged strategically or invisible in their narratives.

Capital Profiles (n=387)

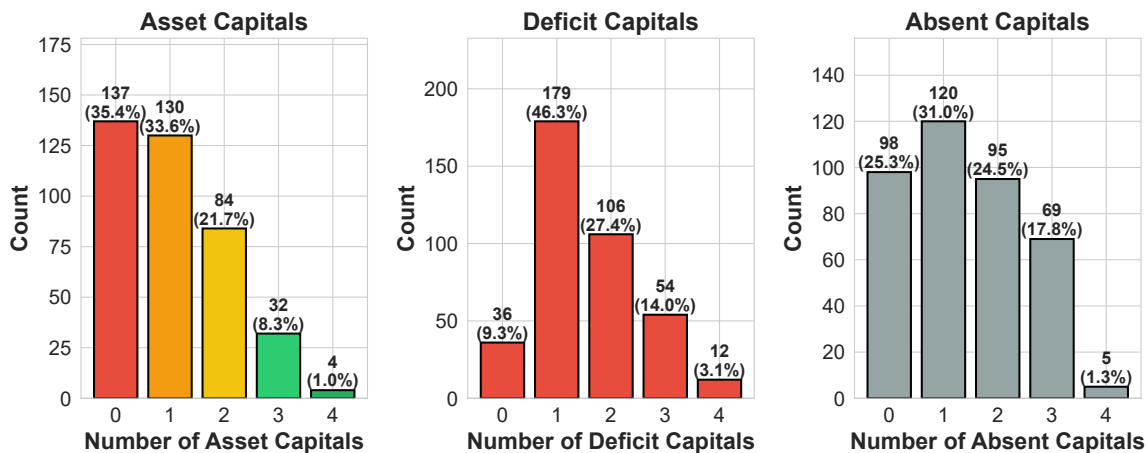


Figure 6: Capital Profile Diversity (n=387): The left panel shows asset distribution, the middle panel shows deficit distribution, and the right panel shows absence distribution across the four capital categories.

Narrative Timelines: In addition to producing narrative timelines, we observe a clear skew in how often writers narrate outcomes beyond schooling. Out of 387 extracted narrative timelines, only 21 include explicit discussion of the author’s post-education life, suggesting that most posts remain focused on in-school challenges rather than longer-term resolution. Among those 21 post-education timelines, 11 describe a positive outcome. We labeled a timeline as positive when the writer explicitly reports outcomes such as obtaining a good job, advancing in their career, achieving financial stability (e.g., earning more than their parents, paying off loans), expressing happiness with where they ended up, or explicitly framing their struggles as “worth it.” Several of the remaining timelines were mixed rather than clearly positive. For instance, authors describe doing better overall while still reporting ongoing debt, stress, or lingering hardship.

6 Discussion

The three theory framework reveals complementary insights: Critical Pedagogy exposes a consciousness-action gap where students are aware but constrained; Ecological Systems Theory pinpoints the microsystem as the primary site of intervention but acknowledges exosystem barriers; and Bourdieu’s Theory reveals that symbolic capital (sense of belonging, recognition, and self worth) is the most invisible and potentially most damaging absence in low SES students’ narratives.

The prevalence of absent capital categories suggests that students either genuinely lack those re-

sources, or do not articulate or frame their experiences using those constructs. Each interpretation has implications: it points to resource scarcity and a potential mismatch between students’ self narratives and institutional recognition frameworks.

This study used three influential educational theories on a common corpus of 387 low SES student narratives, enabling the first systematic comparison of their explanatory power. Our findings reveal that Bourdieu’s Theory of Capital and Habitus provides the most comprehensive framework for understanding socioeconomic inequity in education. While Critical Pedagogy illuminates students’ critical consciousness (71.6% high awareness of structural barriers) and Ecological Systems Theory maps the locations of struggle (62.8% microsystem dominance), Bourdieu’s framework uniquely explains how disadvantage operates through multiple resource dimensions. The near universal economic capital deficit (85.8%) confirms material hardship, but more significantly, Bourdieu reveals the hidden dimensions of inequality: social capital polarized between asset (32.6%) and absence (42.1%), cultural capital nearly split between asset (46.0%) and absence (42.1%), and most strikingly, symbolic capital belonging and recognition either absent (48.6%) or framed as deficit (32.3%). These findings demonstrate that low SES students face not merely financial constraints but a multi dimensional capital deprivation that quantitative metrics alone cannot capture.

However, each theory also exhibits distinctive blind spots that justify a synthetic approach with

Bourdieu as foundation. Bourdieu's framework, with its emphasis on structural reproduction, risks portraying students as passive carriers of capital rather than active critics of unjust systems. Critical Pedagogy corrects this by revealing that students exercise remarkable agency despite constraints (48.1% high agency) and explicitly name structural barriers (92.5% of narratives). Yet Critical Pedagogy alone cannot specify what resources students lack or where interventions should target. Ecological Systems Theory supplies spatial precision, locating 62.8% of challenges in the microsystem (immediate environments) and 24.0% in the exosystem (policies and institutions), but cannot explain how problems operate within those contexts. Bourdieu supplies the missing mechanism: capital deficits cascade through ecological levels, transforming material scarcity into social isolation, cultural confusion, and psychological non-belonging.

The implications for education are clear. First, interventions must address multiple capitals simultaneously: financial support (economic capital), explicit teaching of hidden curricula (cultural capital), intentional network-building structures (social capital), and active affirmation of students' worth (symbolic capital). Second, the consciousness-action gap documented through Critical Pedagogy warns that raising awareness alone is insufficient, and students need tangible resources, not just recognition of injustice. Third, the ecological distribution of narratives reminds us that microsystem supports cannot compensate for exosystem barriers like financial aid policies or rigid scheduling. A synthesized theoretical framework Bourdieu as the diagnostic lens for what students lack, Freire for understanding how they critique their situation, and Bronfenbrenner for identifying where to intervene offers educators the most powerful toolkit for designing equitable, precisely targeted interventions.

Narrative Timelines: Qualitative Trajectories and Transition Points

The extracted within-post timelines reveal how students narrate trajectories, not only in terms of events, but also in how experiences accumulate and shift over time. Even when posts contain few explicit time expressions, the GPT-5.2→TLEX pipeline often recovers coherent temporal ordering by leveraging implicit temporal cues (e.g., discourse structure, causal phrasing, and aspectual markers). This is especially valuable for educational-theory-grounded analysis because it helps operationalize transition points, making

it easier to identify when barriers emerge, intensify, or begin to resolve. In addition, timelines frequently make mood/affect shifts visible across the narrative sequence: in one illustrative post, the author describes a difficult period during their college years and then contrasts it with their present circumstances, noting that 17 years later they are happy with their life. That turning point appears clearly in the timeline, with earlier events associated with hardship and later events reflecting a more positive state, highlighting narrative change in a way that aligns with the paper's emphasis on trajectories and theoretically meaningful moments for potential intervention.

7 Error Analysis and Future Work

The dataset for this study was drawn from prior work (Abdelgaber et al., 2025). Although the original data collection included human validation, some inconsistencies or low-quality narratives may remain. During exploratory analysis, we identified at least one narrative that only weakly matched the dataset's inclusion criteria. To maintain methodological consistency with the prior study, these data points were retained, though they may introduce minor noise. Future work will aim to expand the dataset with additional manually validated narratives and implement stricter filtering to ensure all posts are accurate.

In the theory experiments, GPT-4.1 enabled scalable analysis across the 387 narratives; however, the use of a large language model introduces potential error. Despite carefully designed prompts specifying theoretical constructs and required evidence, LLMs may still misinterpret ambiguous language or incorrectly map content. A notable example involves distinguishing between absent and deficit categories, which often depend on subtle contextual cues. Future work will incorporate human validation to improve reliability. Another improvement is to add an automated evaluation step for the labeling procedure. For example, repeated runs on the same narratives could be compared to assess label stability, and model-generated labels could be checked against a smaller human-coded validation set using agreement metrics such as Cohen's kappa. This would provide a more explicit measure of consistency in the theoretical coding process.

Our analysis primarily reports relative frequencies of theoretically coded categories. Future work

should move beyond descriptive distributions by examining correlations or associations among coded factors, such as the relationship between agency, consciousness of barriers, ecological system level, and different forms of capital. For example, additional analysis could test whether economic capital deficits co-occur with lower agency, whether symbolic capital deficits are associated with higher consciousness of barriers, or whether certain ecological system levels correspond to specific capital profiles. This would provide a stronger quantitative account of how the theoretical dimensions interact rather than treating each category separately.

A further limitation is that two GPT model versions were used for different subtasks. Although this choice reflected the different complexity and resource demands of the tasks, future work should include a controlled comparison in which both models are evaluated on the same subset of narratives. This would help determine whether model choice affects coding consistency, temporal annotation quality, or downstream findings.

Two primary limitations affect narrative timeline extraction. For cross-post ordering, future improvements include enhanced detection of temporal expressions and the use of these references as alignment points. When explicit anchors are unavailable, each post's platform timestamp can serve as a proxy to reconstruct a higher-level sequence, enabling longitudinal tracking of barriers and affect. To reduce within-post fragmentation, we will introduce a linking step that identifies bridging temporal cues such as connectives ("later", "after that", "by then"), causal indicators, and narrative shifts, generating candidate links between otherwise disconnected event chains. This approach is expected to yield more complete and interpretable timelines.

8 Limitation

First, the dataset is relatively small with 387 narratives and drawn from Reddit, which may bias the sample toward students comfortable sharing experiences publicly. Second, the study relies on large language models for theoretical coding. While effective at identifying patterns, these automated interpretations approximate rather than perfectly reproduce expert qualitative judgment. Third, narrative content is inherently selective. Students may emphasize some experiences and omit others, so the absence of a concept does not imply absence in reality. Finally, temporal timeline analysis has

constraints. Cross-post timelines are difficult to construct due to missing or ambiguous dates, and some within-post timelines produce disconnected event chains, limiting the ability to fully capture sequential progression in a single narrative.

9 Contributions

We have four major contributions. First, this study takes an interdisciplinary approach, bridging Computer Science and Education by combining computational methods with established educational theories. In particular, it translates core constructs from Freire's Critical Pedagogy, Bronfenbrenner's Ecological Systems Theory, and Bourdieu's Theory of Capital and Habitus into computationally measurable features, enabling systematic analysis of low-SES student narratives. Second, the paper incorporates temporal timeline extraction, capturing ordered event sequences and showing how challenges and forms of capital evolve over time, linking theoretical categories such as barriers, supports, and resources to specific moments in students' experiences. Third, it provides a comparative assessment of the educational frameworks, evaluating how each theory explains different dimensions of students' lived experiences and responses. Finally, we made our code public for reproducibility.

10 Ethical and Societal Impact

This research has important ethical and societal implications. By analyzing narratives from low-SES students, it highlights systemic barriers and inequities in higher education, helping educators, policymakers, and institutions better understand the challenges faced by marginalized populations. Ethically, the study ensures privacy and anonymity by using only publicly available data, removing identifying information, and aggregating findings to prevent exposure of individual students. Moreover, the work contributes to equity-centered interventions by identifying patterns of barriers, supports, and forms of capital across students' educational experiences. By linking theoretical frameworks to practical, temporally structured insights, this research can inform targeted support strategies, mentorship programs, and policy initiatives aimed at reducing educational disparities. Additionally, the interdisciplinary approach demonstrates how computational methods can be ethically applied to socially relevant educational research.

References

- Nahed Abdelgaber, Labiba Jahan, Arham Vinit Doshi, Rishi Suri, Hamza Reza Pavel, and Jia Zhang. 2025. Bridging the socioeconomic gap in education: A hybrid ai and human annotation approach. In *Proceedings of the 29th Conference on Computational Natural Language Learning*, pages 348–364.
- Steven Bethard. 2013. ClearTK-timeml: A minimalist approach to tempeval 2013. In *Second joint conference on lexical and computational semantics (*SEM)*, volume 2: *proceedings of the seventh international workshop on semantic evaluation (SemEval 2013)*, pages 10–14.
- Pierre Bourdieu. 1986. The forms of capital. In John Richardson, editor, *Handbook of Theory and Research for the Sociology of Education*, pages 241–258. Greenwood Press.
- Urie Bronfenbrenner. 1979. *The ecology of human development: Experiments by nature and design*. Harvard university press.
- Nathanael Chambers, Taylor Cassidy, Bill McDowell, and Steven Bethard. 2014. [Dense event ordering with a multi-pass architecture](#). *Transactions of the Association for Computational Linguistics*, 2:273–284.
- Linguistic Data Consortium and 1 others. 2002. The aquaint corpus of english news text.
- Alon Eirew, Kfir Bar, and Ido Dagan. 2025. Beyond pairwise: Global zero-shot temporal graph generation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 31428–31446.
- Engineering Education–Workforce Continuum Committee. 2019. *Understanding the educational and career pathways of engineers*. National Academies Press.
- Paulo Freire. 2020. Pedagogy of the oppressed. In *Toward a sociology of education*, pages 374–386. Routledge.
- Ella R Kahu and Catherine Picton. 2022. Using photo elicitation to understand first-year student experiences: Student metaphors of life, university and learning. *Active Learning in Higher Education*, 23(1):35–47.
- Motti Kelbessa, Ilyas Jamil, and Labiba Jahan. 2022. Impacts of low socio-economic status on educational outcomes: A narrative based analysis. In *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*, pages 59–69.
- Motti Kelbessa, Estephanos Jebessa, and Labiba Jahan. 2024. [Addressing educational inequalities of low ses students: Leveraging natural language processing for impact](#). In *Proceedings of the 17th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '24*, page 388–391, New York, NY, USA. Association for Computing Machinery.
- Md. Tarique Hasan Khan, Labiba Jahan, and Mustafa Ocal. 2026. An inflection point in equity research: Leveraging ai and educational theory to decode the socioeconomic gap in engineering education. In *Proceedings of the ASEE Gulf-Southwest Annual Conference*. American Society for Engineering Education (ASEE).
- Sang Min Lee, M Harry Daniels, Ana Puig, Rebecca A Newgent, and Suk Kyung Nam. 2008. A data-based model to predict postsecondary educational attainment of low-socioeconomic-status students. *Professional School Counseling*, 11(5):2156759X0801100504.
- Shixia Liu, Michelle X. Zhou, Shimei Pan, Yangqiu Song, Weihong Qian, Weijia Cai, and Xiaoxiao Lian. 2012. [Tiara: Interactive, topic-based visual text summarization and analysis](#). *ACM Trans. Intell. Syst. Technol.*, 3(2):25:1–25:28.
- Julie P Martin, Denise R Simmons, and Shirley L Yu. 2013. The role of social capital in the experiences of hispanic women engineering majors. *Journal of Engineering Education*, 102(2):227–243.
- Kerry L Meyers, Matthew W Ohland, Alice L Pawley, Stephen E Silliman, and Karl A Smith. 2012. Factors relating to engineering identity. *Global Journal of Engineering Education*, 14(1):119–131.
- Paramita Mirza and Sara Tonelli. 2016. Catena: Causal and temporal relation extraction from natural language texts. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 64–75.
- Borja Navarro-Colorado and Estela Saquete. 2016. [Cross-document event ordering through temporal, lexical and distributional knowledge](#). *Know.-Based Syst.*, 110(C):244–254.
- Mustafa Ocal, Adrian Perez, Antonela Radas, and Mark Finlayson. 2022. Holistic evaluation of automatic timeml annotators. In *Proceedings of the thirteenth language resources and evaluation conference*, pages 1444–1453.
- Mustafa Ocal, Akul Singh, Jared Hummer, Antonela Radas, and Mark Finlayson. 2023. [jtlex: a java library for timeline extraction](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 27–34.
- Mustafa Ocal, Ning Xie, and Mark A Finlayson. 2026. Tlex: an efficient method for extracting exact timelines from timeml temporal graphs. *Language Resources and Evaluation*, 60(1):10.
- Matthew W Ohland, Catherine E Brawner, Michelle M Camacho, Richard A Layton, Russell A Long, Susan M Lord, and Mara H Wasburn. 2011. Race, gender, and measures of success in engineering education. *Journal of engineering education*, 100(2):225–252.

OpenAI. 2025a. Introducing gpt-4.1 in the api. <https://openai.com/index/gpt-4-1/>. Accessed 2026-05-08.

OpenAI. 2025b. Introducing gpt-5.2. <https://openai.com/index/introducing-gpt-5-2/>. Accessed 2026-05-08.

Donna Riley. 2022. Engineering and social justice. In *Engineering and social justice*, pages 47–106. Springer.

Laura Robinson. 2020. The stem selfing process: Nondigital and digital determinants of aspirational stem futures. *American Behavioral Scientist*, 64(7):950–972.

Susan C Saegert, Nancy E Adler, Heather E Bullock, Ana Mari Cauce, William Ming Liu, and Karen F Wyche. 2006. [Report of the apa task force on socioeconomic status](#). Retrieved from the *American Psychological Association website*.

E. Saquete, P. Martínez-Barco, R. Muñoz, and J. L. Vicedo. 2004. [Splitting complex temporal questions for question answering systems](#). In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Roser Sauri, Jessica Littman, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. 2006. TimeML annotation guidelines, version 1.2.1. https://catalog.ldc.upenn.edu/docs/LDC2006T08/timeml_anguide_1.2.1.pdf.

Stephen Secules, Cassandra McCall, Joel Alejandro Mejia, Chanel Beebe, Adam S Masters, Matilde L. Sánchez-Peña, and Martina Svyantek. 2021. Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community. *Journal of Engineering Education*, 110(1):19–43.

Marc Verhagen, Inderjeet Mani, Roser Saurí, Jessica Littman, Robert Knippen, Seok Bae Jang, Anna Rumshisky, John Phillips, and James Pustejovsky. 2005. [Automating temporal annotation with tarsqi](#). In *ACL*, pages 81–84.

Jing Wang and Jeremy C Weiss. 2025. A large-language model framework for relative timeline extraction from pubmed case reports. *ArXiv*, pages arXiv–2504.

Tara J Yosso*. 2005. Whose culture has capital? a critical race theory discussion of community cultural wealth. *Race ethnicity and education*, 8(1):69–91.

A Appendix

Prompt Used for Theory-Based Coding: The full prompt used for GPT-based theoretical coding is included below to support reproducibility.

```
# GPT-4 Three-Theory Coding Prompt

You are an expert in educational theory and qualitative research coding. Your task is to classify a student narrative against three major theoretical frameworks.

## THEORETICAL FRAMEWORKS

### 1. CRITICAL PEDAGOGY (Freire)
Education is a political act that can reproduce or challenge inequality. Focus on how narratives reveal:
- **Student Agency**:: Evidence of student voice, empowerment, or willingness to act
- **Structural Barriers**:: Systemic obstacles, injustice, or inequitable conditions
- **Conscientization**:: Student awareness of how social systems affect them
- **Liberation/Constraint**:: Movement toward freedom or deeper constraint

**Coding Instructions**:: Look for language showing consciousness of barriers, agentic responses, or expressions of empowerment/disempowerment.

---

### 2. ECOLOGICAL SYSTEMS THEORY (Bronfenbrenner)
Student experiences are shaped by nested systems influencing each other. Identify which level(s) most affect the student:
- **Microsystem**:: Immediate environment (family, school, peer relationships, workplace)
- **Mesosystem**:: Connections between microsystems (e.g., family-school relationships)
- **Exosystem**:: Indirect influences (parents' workplace, educational policies, economic conditions)
- **Macrosystem**:: Societal ideologies, values, culture, laws
- **Chronosystem**:: Historical/temporal context (life transitions, major events)

**Coding Instructions**:: Identify which level(s) the student emphasizes most. Most narratives will mention multiple levels-identify the 1-2 DOMINANT levels.

---

### 3. BOURDIEU'S THEORY OF CAPITAL & HABITUS
Students accumulate four forms of capital that determine educational and social outcomes:
- **Economic Capital**:: Financial resources, income, material assets, employment status
- **Social Capital**:: Networks, relationships, institutional connections, support systems
- **Cultural Capital**:: Educational credentials, knowledge, language skills, cultural practices
- **Symbolic Capital**:: Status, prestige, recognition, respect, legitimacy in society

For each capital type, classify as:
- **DEFICIT**:: Student explicitly mentions lacking, struggling with, or needing this capital
- **ASSET**:: Student explicitly mentions having, building, or leveraging this capital
- **ABSENT**:: Capital is NOT mentioned or discussed in the narrative at all

**Coding Instructions** (Use specific language/themes as evidence, not speculation):
- **Economic deficit** (explicit): "can't afford," "need to work," "financial stress," "money problems," "struggling financially"
- **Economic asset** (explicit): "got scholarship," "parents support," "earning income," "financial stability"
- **Economic absent**:: Narrative never discusses finances, money, work, or employment
- **Social deficit** (explicit): "no support," "isolated," "family problems," "no mentor," "alone," "don't have connections"
- **Social asset** (explicit): "family helped," "friends support," "mentor/professor helped," "community support"
```

```

- **Social absent**: Narrative never discusses relationships, support systems, or social networks
- **Cultural deficit** (explicit): "unprepared," "don't know," "no credentials," "behind," "underprepared"
- **Cultural asset** (explicit): "good student," "high GPA," "learning," "degree," "educated," "knowledge"
- **Cultural absent**: Narrative never discusses education, credentials, knowledge, or academic readiness
- **Symbolic deficit** (explicit): "shame," "not good enough," "imposter," "expected to fail," "invisible," "doubt"
- **Symbolic asset** (explicit): "proud," "overcame," "resilient," "recognized," "valued," "confident"
- **Symbolic absent**: Narrative never discusses status, recognition, respect, or self-perception

---

## STUDENT NARRATIVE

{NARRATIVE_TEXT}

---

## REQUIRED OUTPUT FORMAT

Return ONLY valid JSON (no markdown, no extra text) with this exact structure:

```json
{
 "critical_pedagogy": {
 "student_agency": "high" | "moderate" | "low" | "absent",
 "structural_barriers_evident": true | false,
 "consciousness_of_barriers": "high" | "moderate" | "low",
 "summary": "Brief 1-2 sentence summary of student's agency/constraint",
 "quotes": ["quote1", "quote2"]
 },
 "ecological_systems": {
 "dominant_level": "microsystem" | "mesosystem" | "exosystem" | "macrosystem" | "chronosystem",
 "secondary_level": "microsystem" | "mesosystem" | "exosystem" | "macrosystem" | "chronosystem" | null,
 "microsystem_elements": ["element1", "element2"],
 "summary": "Brief explanation of dominant system level(s)"
 },
 "bourdieu_capital": {
 "economic": "deficit" | "asset" | "absent",
 "social": "deficit" | "asset" | "absent",
 "cultural": "deficit" | "asset" | "absent",
 "symbolic": "deficit" | "asset" | "absent",
 "dominant_capital_need": "economic" | "social" | "cultural" | "symbolic",
 "dominant_capital_strength": "economic" | "social" | "cultural" | "symbolic" | null,
 "summary": "Brief summary of capital profile"
 },
 "overall_summary": "2-3 sentence synthesis: How do all three theories help us understand this student's experience?"
}
```

---

## INSTRUCTIONS FOR CODING

1. **Read the entire narrative carefully** before making classifications.
2. **Use evidence from the text** - Do NOT make assumptions or inferences beyond what is explicitly stated.
3. **For Bourdieu**: IMPORTANT - If a capital type is never mentioned or discussed, code it as "

```

```

absent" rather than inferring deficit or asset.
A narrative can have BOTH deficits and assets in the same capital type, OR it can omit the capital entirely.
4. **For Ecological Systems**: Most narratives will show multiple levels-pick the 1-2 that most dominate student's experience.
5. **For Critical Pedagogy**: Look for language of consciousness, agency, empowerment, or constraint.
6. **CRITICAL RULE**: Never force a classification when sufficient evidence is absent. Use "absent," "low," or "null" rather than guessing.
7. **Be specific in summaries**: Don't just repeat definitions-show what THIS student's narrative reveals using direct language from the text.

---

## VALIDITY CHECK

Before returning JSON:
- [ ] Does the JSON parse and validate?
- [ ] Have I cited evidence from the narrative?
- [ ] Am I using the student's own language/voice?
- [ ] Does the summary synthesize across theories?

```