# Estimation and Correlation of Student Maturity with Social Attributes Using Large Language Models and Transformers

**Ridhima Kar[1], Parijat Kar[2]**

[1]Capistrano Unified School District

[2]University of California

29karr@capousd.org, parijatkar@uci.edu

**Abstract**

Positive academic growth is foundational for a child's well-being and future success. While most children exhibit strong indicators of flourishing, early identification of potential challenges can significantly impact their development. This study utilizes AI-powered natural language processing techniques to assess a child's academic developmental progress based on classwork and conversational data.

Analyzing homework samples from elementary and middle school students in the Capistrano Unified School District, we fine-tuned large language models to classify conversations by grade level. This classification identifies potential discrepancies between a child's developmental stage and conversational maturity.

We found interesting correlations between social parameters like native languages, family income, race, number of siblings, and pets.

Furthermore, we explored the potential of AI-driven interventions. Feedback from a system developed using Large Language Models helped students retain vocabulary and grammatical accuracy.

## Background Work

### Child Development

Language, cognitive, and social development provide insights into general developmental trends. Early language acquisition, encompassing stages from babbling to complex sentence formation, and pragmatics (language use in social contexts), are key aspects of language development. Cognitive development, explored through theories like Piaget's, sheds light on children's thinking processes and understanding of mental states. Social development encompasses social cognition and emotional regulation, crucial for understanding social relationships. (Jiahong Su et all 2023)

A comprehensive evaluation of a child's developmental stage requires considering multiple factors. Physical development, including motor skills, provides clues about age. Social and emotional development, observed through interactions, offers insights into social and emotional maturity. Cognitive development, assessed through problem-solving, memory, and attention span, contributes to a holistic understanding. Consulting with a qualified child development professional is recommended for a precise evaluation. (Lisa Reinhart et all 2024)

### AI and Child Development

Since 2012, AI research has significantly impacted child development in several ways.

Early Identification of Developmental Delays: AI algorithms are being developed to analyze various data sources, such as video recordings of children's play, speech patterns, and even physiological signals, to identify potential developmental delays or disorders like autism spectrum disorder (ASD) and developmental dyslexia at an early stage. (Scassellati, 2007)

Studies have shown promising results in using machine learning to predict the risk of autism in infants based on early brain scans and behavioral observations. (Prentzas, 2013)

### Personalized Learning Experiences

AI-powered educational platforms adapt to individual learning paces and styles, providing personalized learning experiences. Intelligent Tutoring Systems (ITS) utilize AI to provide customized feedback and guidance to students, adjusting the difficulty level and content based on their performance. (Sun et al., 2021)

### Enhanced Social-Emotional Learning

AI-powered tools are being developed to support children's social-emotional learning, such as recognizing and managing emotions, developing empathy, and building healthy relationships. AI-driven platforms can provide interactive exercises and simulations to help children practice social skills and develop emotional intelligence. AI technologies are used to assist children with special needs, such as autism or communication disorders. AI-powered communication devices can help children with limited verbal abilities to communicate more effectively using text-to-speech and image recognition. (Nan, 2020)

### AI and Age Estimation

AI, particularly in natural language processing and speech recognition, shows promise in estimating age based on textual and vocal cues. Text analysis can assess linguistic patterns and vocabulary usage to infer age. For instance, AI can analyze social media posts to gauge maturity levels or assess student writing to tailor educational content. Voice analysis can leverage acoustic features like pitch, intonation, speech rate, and estimating age. This technology could be applied in scenarios such as voice-activated devices adjusting their language complexity or call centers routing calls based on the caller's estimated age. (Syed Ashiqur Rahman et al 2020) It's important that AI is a tool and should be used with human judgment. Ethical considerations and potential biases must be carefully addressed to ensure accurate and fair assessments. (Andy Nguyen et al 2022)

## Data Collection

Initially, we collected 11,600 conversations and 2,600 homework assignments from 211 unidentified students for training fine-tuned LLM. These students are from three elementary schools and two middle schools. These students were only classified by gender and grade. This data was collected from students' WhatsApp chat, and Gmail exchanges. Zoom meeting recordings with built-in speech-to-text were also used. We obtained consent from all participants to use their conversations anonymously in our experiment. Beyond electronic data collection, we collected direct communications from multiple audio and video recordings from social events. We used a speech-to-text tool to convert those to usable data. We have a few sample conversations in table 1. We associated each student with an MBTI personality, gender, and grade (Kar 2025). Prior works have shown gender differences in early language acquisition, focusing on actions, gestures, and vocabulary. While a female advantage is often perceived, studies using varied methods show slight differences, influenced by factors like age, sample size, and individual variability, particularly among boys. These differences likely stem from complex interactions between biological, neuropsychological, and cultural factors (Rinaldi et al 2023).

We collected homework assignments from students with their social attributes to identify correlations with maturity levels. This is discussed in detail in the Experiment section. In table 1 and table 2, Riya, Anika, Maya, Aisha, and Chloe are girls.

Conversation 1: The Extroverted Friend Group

- Riya (ENFP): "Hey, did you go anywhere for Thanksgiving? Like, out of state or something?"
- Arjun (ESFP): "Nah, we just stayed home. But we had a huge family dinner and played board games all night!"

Conversation 2: The Quiet Intellectuals

- Aisha (INTJ): "I finished that new sci-fi book over the break. It was mind-blowing."
- Wei (INTP): "Nice. I spent most of my time coding. I'm almost done with my new game."

Conversation 3: The Sensitive Souls

- Anika (INFJ): "I volunteered at the soup kitchen. It was really rewarding."
- Fahad (INFP): "That's awesome. I just relaxed and read. I needed some quiet time."

Conversation 4: The Energetic Optimists

- Maya (ENFP): "We went to my grandma's house and had a huge feast. We even had a pie-eating contest!"
- Li (ESTP): "That sounds so fun! I went to my cousin's house, and we played video games all day."

Table 1: Example Conversations

## Classification Framework

Our objective in this work is to identify maturity levels in the conversations of middle and late elementary school students. To do this responsibly, filter all new slang that are in bad taste.

We created multiple fine-tuned models for both genders by grade level. We segregated our data by gender to train Large Language Models like GPT4, Llama V3, Claude V2, and Titan G1 Express with labeled grades.

As a longer conversation or homework assignment might demonstrate different maturity levels, we split all our training data with a maximum limit of 100 words.

### Slang Identification

We used a database of profanity from Kaggle (Kaggle English). We also considered a database of well-known and acceptable school slang and informal languages.

We identified all non-dictionary words, 2-grams, and 3-grams around those. We used the English Dictionary database from Kaggle.

Conversation 5: The Analytical Minds

- Aisha (INTJ): "I learned how to code a new programming language."
- Wei (INTP): "Nice. I've been working on a math problem that's been bugging me."

Conversation 6: The Social Butterflies

- Chloe (ESFP): "I went to a huge Thanksgiving party with all my friends. It was so much fun!"
- Arjun (ENFP): "That's awesome! I went to a family reunion. There were like, a hundred people there."

Conversation 7: The Cautious Planners

- Anika (ISFJ): "I helped my mom cook Thanksgiving dinner. It was a lot of work, but it was worth it."
- Fahad (ISTJ): "I helped my dad fix the car. It was a good learning experience."

Conversation 8: The Creative Spirits

- Maya (ENFP): "I wrote a short story over the break. It's about a magical forest."
- Riya (ESFP): "Cool! I painted a picture of a sunset."

Conversation 9: The Quiet Observers

- Aisha (INTJ): "I just relaxed and thought about life."
- Wei (INTP): "I watched some documentaries. I learned a lot about space."

Conversation 10: Helpful Friends

- Chloe (ESFP): "I helped my neighbor rake leaves."
- Alex (ENFP): "I volunteered at the local animal shelter."

Table 2: Example Conversations 2



We computed the probability of any non-dictionary words and their grams to be an offensive word using BERT embeddings. We also use the Large Language Model Titan G1 Express with prompts to identify the meaning of an unknown word within the given context. For example, in Table 3, a cram session means an intense study period. It was not in the Kaggle database, we identified it as a new slang with meaning. We also identified that it’s not a curse or profanity. We did it in two ways. First, we used a direct prompt on each example and noted the outcome. Second, we retrieved all conversations that include this phrase from the Knowledgebase.

Student 1: "**Dude**, I'm so stressed about this upcoming history test. I feel like I'm going to have to pull an all-nighter for a **cram session**."

Student 2: "Same here! I'm drowning in history notes. Maybe we should have a study group at my place after school? We can help each other out."

Student 1: "That sounds great! We can quiz each other and make sure we know all the dates and events."

Student 2: "Yeah, and maybe we can even turn it into a mini **pep rally** to get pumped up for the test!"

Student 1: "Haha, love the idea! We can even make some silly chants to boost our **aloha spirit**."

Student 2: "I'm just hoping I don't get called on to answer a question and completely blank out. I'd be so embarrassed!"

Student 1: "Don't worry, you'll be fine. Just try to relax and focus on what you do know. Besides, the teacher knows we're all a little stressed with all these tests coming up."

Student 2: "I know, but it's still nerve-wracking. I don't want to get detention for not studying!"

Student 1: "Don't even think about detention! Let's just focus on **acing** this test and enjoying the rest of the week. Maybe we can even celebrate with some ice cream after school."

Student 2: "Deal! Now, let's get back to studying. I need to become a **bookworm** for the next few hours."

Table 3: Examples of School Slangs

The meaning of this new slang computed by LLM Claude V2 is in Table 4. Identified 50 new elementary school abbreviations and informal languages.

| |
|---|
| Cram session: Intense period of studying before an exam. |
| Study group: A group of students who study together. |
| Aloha spirit: Aloha is the school theme meaning school spirit |
| Pep rally: An event to boost school spirit. |
| Detention: Punishment for misbehaving. |
| Bookworm: Someone who enjoys reading a lot. |

Table 4: Slang Meanings

### Model Training

We created a homework assignment and conversation classification framework. Our objectives were to reach a high accuracy in grade-level classification so that we could use these fine-tuned models to evaluate the maturity level of homework, assignments, and random conversations of any student.

First, we segregated homework and conversations by grade level and gender. We created BERT embeddings by grade and gender. We computed a cosine similarity score with different BERT embeddings and chose the final classification using SoftMax. This classification was also 97% accurate.

Second, we created fine-tuned LLM models from base models GPT-4, Llama V3, and Claude V2. The fine-tuned LLM models produced a direct classification. We reached 96% to 97% accuracy from all popular large language models.

### Maturity Score Computation

We computed a maturity score based on the weighted average of classifications. If a student has n number of homework assignments, we classify all of those. For any higher-grade classification, we added a score of 0.25. For any lower-grade classification, we subtracted 0.25 (Algorithm 1). The final maturity score is an average of all classification scores. We considered maturity scores above 0.9 to be advanced, and maturity scores below 0.5 to be a deficiency.

Algorithm 1: Compute Maturity Score

```
Input Parameter: Student ID S, List of Assignments L
Total = 0
Number of assignments = |L|
Γ = 0.25 (adjustment rate)
Grade = Student Grade
For all assignments in L:
    If Classify (Li) equals to Grade:
        Total = Total + Grade
    Else if Classify (Li) higher than Grade:
        Total = Total + (Grade - Classify (Li)) * Γ
    Else
        Total = Total + (Classify (Li) - Grade) * Γ
Return Total/|L|
```

## Social Attributes Correlation Experiment

We collected 1600 homework assignments and conversations from 211 students. We chose students from multiple elementary and middle schools. We also collected information about the second language spoken, race, pets, and number of older siblings.

## Observations

Figure 1 shows that the maturity level in writing is better for students with an older sibling.

11% of male students demonstrated advanced maturity levels for their grades. 21% of male students exhibited deficiencies in vocabulary and sentence structure. The results for the fourth-grade students using the GPT-4 classifier are shown in Figure 1.

Female students, on average, demonstrated higher levels of writing maturity. A sentimental analysis of their writings proved to be more optimistic and pragmatic. They had 10% better vocabulary choices. We picked the most frequent 111 adjectives.

The other interesting finding (Figure 2.1 and Figure 2.2) was that the students with 1-2 siblings and pet cats had a higher maturity level than the general population.

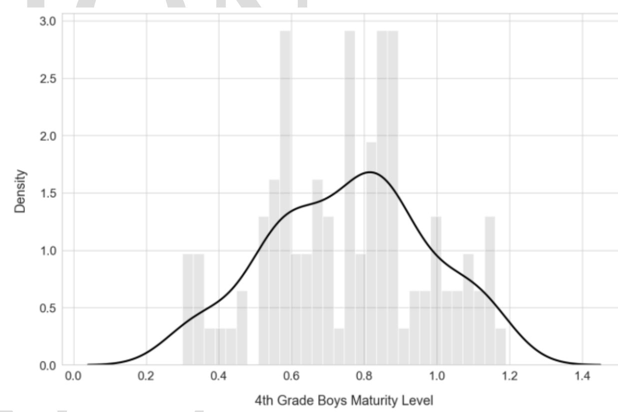


Figure 1: Maturity Level Distribution

## Future Work

This research investigates pre-trained models to facilitate linguistic enhancement and provide mental health support for children. AI-powered conversational agents may align with a child's developmental stage, we aim to cultivate a supportive environment that fosters positive mental well-being and cognitive growth. Furthermore, this study will focus on monitoring the progression of maturity levels within a cohort of children exhibiting developmental trajectories within the first standard deviation of the mean, as depicted in Figure 1.

We also want to explore this framework to evaluate other works like sketching, arts, and music to their maturity levels. We are using image classifiers with custom labels for it. Finally, we want to merge this work with our other work by using generative text and chatbots to accelerate development for children behind minimum standards.
the same framework.

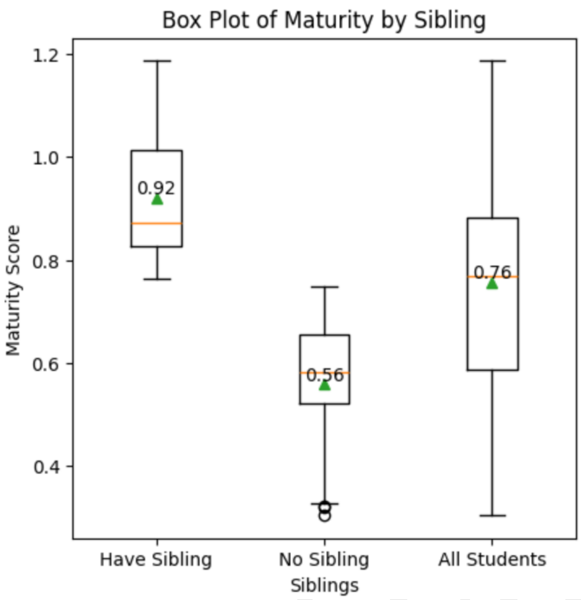


Figure 2.1: Sibling Impact

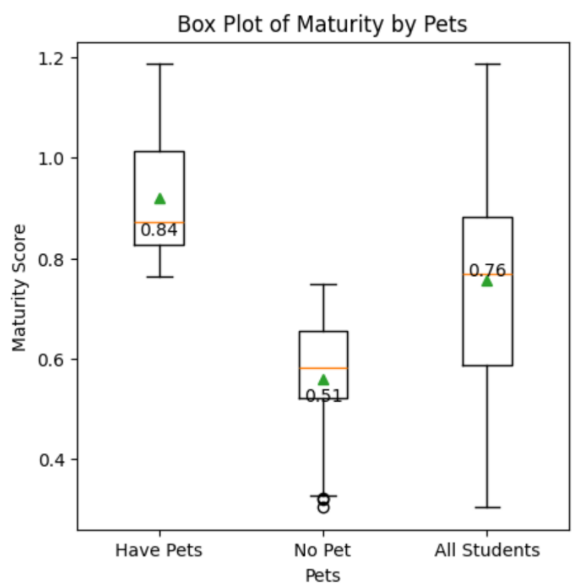


Figure 2.2: Pet Impact

## References


Lisa Reinhart et all 2024 Intelligence Medicine Volume 9 - Artificial Intelligence in Child Development Monitoring: A Systematic Review on Usage, outcomes, and Acceptance
Jiahong Su et all 2023 Volume 3 Computers and Education: Artificial Intelligence - Artificial intelligence in early childhood education: A scoping review
Nan, 2020 J. Nan Research of application of artificial intelligence in. Preschool education. The 2020 international symposium on electronic information technology and communication engineering (2020, June 19-21)
Prentzas, 2013 J. Prentzas Artificial intelligence methods in early childhood education. X.-S. Yang (Ed.), Artificial Intelligence, evolutionary computing and metaheuristics, Springer (2013), pp. 169-199
Scassellati, 2007 B. Scassellati How social robots will help us to diagnose, treat, and understand autism. Robotics research, Springer (2007), pp. 552-563
Shin et al., 2008 K.C. Shin, N. Kuppuswamy, H.C. Jung Network-based service robot for education EKC2008 proceedings of the EU-Korea conference on science and technology, Springer (2008), pp. 307-313
Sun et al., 2021 Z. Sun, M. Anbarasan, D. Praveen Kumar Design of online intelligent English teaching platform based on artificial intelligence techniques. Computational Intelligence.
Syed Ashiqur Rahman et al 2020. Deep learning for bio-logical age estimation. PubMed.
Andy Nguyen et al 2022. Ethical principles for artificial intelligence in education. PubMed.
Kar P. Convergence of Personalities Using Reinforcement Learning and Text Generation. AAAI Spring Symposium 2025.
Rinaldi P, Pasqualetti P, Volterra V, Caselli MC. Gender differences in early stages of language development. Some evidence and possible explanations. J Neurosci Res. 2023 May;101(5):643-653. doi: 10.1002/jnr.24914. Epub 2021 Jul 9. PMID: 34240751.