Linguistic Alignment Predicts Learning in Small Group Tutoring Sessions

Dorothea French,[1] Robert Moulder,[1] Kelechi Ezema,[1] Katharina von der Wense,[1,2]
and **Sidney D'Mello[1]**
[1]Institute of Cognitive Science, University of Colorado Boulder,
[2]Institute of Computer Science, Johannes Gutenberg University Mainz
{dofr2963, keez8796, romo2943, kaka6430, sidm2155}@colorado.edu

Abstract

Cognitive science offers rich theories of learning and communication, yet these are often difficult to operationalize at scale. We demonstrate how natural language processing can bridge this gap by applying psycholinguistic theories of discourse to real-world educational data. We investigate linguistic alignment – the convergence of conversational partners' word choice, grammar, and meaning – in a longitudinal dataset of real-world tutoring interactions and associated student test scores. We examine (1) the extent of alignment, (2) role-based patterns among tutors and students, and (3) the relationship between alignment and learning outcomes. We find that both tutors and students exhibit lexical, syntactic, and semantic alignment, with tutors aligning more strongly to students. Crucially, tutor lexical alignment predicts student learning gains, while student lexical alignment negatively predicts them. As a lightweight, interpretable metric, linguistic alignment offers practical applications in intelligent tutoring systems, educator dashboards, and tutor training.

1 Introduction

Cognitive science has developed theories of mind, learning, and communication (Pickering and Garrod, 2006; Fusaroli et al., 2014; Biggs, 1996), which aim to explain how individuals construct and share knowledge. These theories, however, are difficult to investigate at scale in real-world contexts which require delving into complex or messy data. Our work illustrates how natural language processing (NLP) techniques can bridge the gap between cognitive science theory, real-world outcomes, and applications in the context of education.

The educational domain lends itself especially well to this exploration, as teachers and students have distinct roles, measurable outcomes, and there is an increasing need for computational tools to support effective learning processes. While many factors contribute to effective learning (Nystrand and Gamoran, 1991; Pollock et al., 2011; Nickow et al., 2020), the success of human tutoring programs underscores the central role of interpersonal communication and individual attention to learning. We argue that, at a basic level, the language used during tutoring can provide meaningful insights into the dynamics and quality of these interactions. This is grounded in the language-as-thought hypothesis: language reflects underlying psychological states and processes, including thoughts, feelings, and motivations within and between individuals (Boroditsky, 2001; D'Mello and Graesser, 2012; Lira et al., 2023; Pennebaker et al., 2003). As language arises naturally in many types of interactions and serves a primary role in communication, it provides unique insights into social processes including shared knowledge construction, perspective taking, and emotional regulation (Clark, 1996; Dowell et al., 2019; Stewart et al., 2019).

In this work we investigate linguistic alignment (Pickering and Garrod, 2004) as a measurable component of interpersonal coordination in small-group tutoring. Prior work suggests that linguistic alignment – convergence on word choice (lexical), grammar (syntactic), and meaning (semantic) – fosters mutual understanding, reduces cognitive load, and enhances conversational outcomes (Micklos and Woensdregt, 2023; Tian et al., 2025; Pickering and Garrod, 2006; Rahimi et al., 2017). While extensively studied in one-on-one dialogues (Kejriwal and Beňuš, 2025; Norman et al., 2022; Sinclair and Schneider, 2021) and, to a lesser extent, in classroom and human–computer interactions (Branigan et al., 2010; French et al., 2024), alignment remains underexplored in human small-group tutoring, particularly in relation to learning outcomes. By investigating the way tutors and students adapt their language to each other, we hope to gain further insight into their communicative roles and strategies, and the implications on learning.

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 15600–15611
November 4-9, 2025

Contribution Using a longitudinal dataset of small-group tutoring sessions and student learning metrics collected over an academic year, we explore the following questions: (1) To what extent does linguistic alignment occur in small group tutoring? (2) How do students and tutors differ in their patterns of alignment? (3) How does linguistic alignment predict student learning outcomes? We find evidence of lexical, syntactic, and semantic alignment from both tutors and students, with tutors exhibiting higher degrees of alignment overall. Importantly, we show that lexical alignment in particular predicts student learning gains. Linguistic alignment offers a lightweight, interpretable, and computationally tractable measure of tutoring interactions. As such, it has a multitude of use cases in tutor training, real-time feedback for tutors, and in the design of intelligent tutoring systems – highlighting the way NLP techniques can be used for both scientific research and in research-based applications.

Novelty We examine linguistic alignment at multiple levels (lexical, sematic and syntactic) and its association with student achievement using an authentic and extensive dataset of small group tutoring. Based on our review of the literature (Section 2), our work is novel in the following areas. First, as opposed to an artificial communicative task (often conducted in the lab), we examine alignment in the context of real-world tutoring sessions, where the goal is to promote engagement and learning. Second, we use a large corpus of data where students from multiple school districts engage in small group tutoring sessions in a unique semi-hybrid setup. The extent to which linguistic alignment emerges in this novel context – both multi-party tutoring and hybrid tutoring – is an open question. Third, to the best of our knowledge, our study is the first to examine whether linguistic alignment predicts student achievement growth over extended time frames, i.e., an entire school year.

2 Related Work

2.1 Interactive Alignment

Conversation is complex as it requires simultaneous activation of multiple skills – memory, speech comprehension, speech production, plus physical movement – the activation of which are closely integrated, both within a single interlocutor's mind and between two or more interlocutors (Micklos and Woensdregt, 2023; Brennan et al., 2010). Clark and Schaefer (1987) establish a classic model of two stage conversational units, where one participant first presents their contribution then the participants work together to establish (accept) that all participants have understood the contribution.Brennan et al. (2010) describe conversation as a joint activity during which participants coordinate their internal mental representations and their actions. The classic interactive alignment model treats conversation as driven by automatic priming: successful communication is achieved when speakers align their linguistic choices and, thus, their internal situation models (Pickering and Garrod, 2006). In contrast, Fusaroli et al. (2014) posit that linguistic processes are not the domain of individual cognitive systems, but synergies between conversational units. This perspective emphasizes that coordinated behavior in dialogue involves both synchrony and complementarity: interlocutors may converge on shared terms and structures, but also take complementary roles conditioned by context and task.

2.2 Linguistic Alignment

Linguistic alignment (also called *entrainment* or *convergence*) refers to the interlocutors' tendency to mutually use similar linguistic patterns over the course of a dialogue (Halberg, 2013; Srivastava et al., 2025). Linguistic alignment is critical to conversation, bolstering understanding between speakers, enabling better cohesion and task success (Friedberg et al., 2012; Gonzales et al., 2010; Ward and Litman, 2007; Xu and Reitter, 2015). When interlocutors reuse exact words and phrases from the recent discourse, they align lexically. Research suggests that this helps speakers draw attention to the salient elements within the conversation (Duran et al., 2024). When speakers converge at similar meanings, they align semantically – for example, *puppy* and *dog* are semantically but not lexically aligned. When speakers reuse syntactic structures from the conversation, they align syntactically (Branigan et al., 2007).

2.3 Linguistic Alignment in Educational Contexts

Effective teaching and learning rely on rich communication between instructors and students. Classrooms that allow time for small-group work, collaborative problem solving, and whole-class discussions provide a more engaging and effective educational experience (Pollock et al., 2011; Nystrand and Gamoran, 1991). Beyond the classroom, one-on-one tutoring offers opportunities for high-quality discourse among students and tutors (Graesser et al., 1995) and is known to be highly effective in promoting learning gains (Nickow et al., 2020). Tutors have more time and ability to focus on individual student needs; they can give personal and timely feedback, diagnose student knowledge and misconceptions, and provide students with opportunities to have guided interactions with the material (Bloom, 1984).

Decades of research on key tutoring practices in novice and expert tutors (D'Mello, 2010) have advanced our understanding of what constitutes effective tutoring (Graesser et al., 2011) and informed the design of intelligent tutoring systems such as AutoTutor (ITSs; D'Mello and Graesser, 2023). However, this research has typically involved labor-intensive human-coding of tutorial transcripts such as different forms of instruction, question asking, and feedback (Graesser et al., 1995). As a complementary approach, we suggest that automated metrics such as linguistic alignment may help provide insights into the communication strategies that underlie the effectiveness of human tutoring.

Within the educational settings, much of the research on linguistic alignment has been focused on a single form (lexical) of alignment, sometimes in conjunction with acoustic or behavioral alignment (Friedberg et al., 2012; Norman et al., 2022; Sinclair and Schneider, 2021; Srivastava et al., 2025; Ward and Litman, 2007). Ward and Litman (2007) examined lexical and acoustic convergence in dyadic tutoring, showing that these metrics were helpful in predicting learning gains for physics concepts, particularly for students with low prior knowledge. Friedberg et al. (2012) examined lexical alignment in teams of 2-5 engineering undergraduates working on an project, showing that high-performing teams were more likely to increase their lexical alignment over time. Duran et al. (2024) studied multiple levels of linguistic alignment (lexical, syntactic, and semantic) amongst triads of undergraduate students working on collaborative problem-solving tasks, finding that low semantic alignment (more diversity of ideas) predicts outcomes.

Outside the educational domain, Rahimi et al. (2017) showed that in teams of 3-4 participants playing a board game, combining multiple alignment types (lexical and acoustic) was a predictor of success. Gonzales et al. (2010) found that linguistic style matching was positively correlated with groups' cohesiveness and performance.

1	Tutor	How are we going to get rid of the exponent for B squared?
2	Tutor	What are we doing?
3	Student	We do opposite operations.
4	Tutor	Which is what, love?
5	Student	Oh, finding the square root.
6	Tutor	The square root.
7	Tutor	So let's go ahead, right and square root, both sides.
8	Tutor	And what do you think we're going to be left with?
9	Tutor	What do you think B equals?
10	Student	B equals the square root of C squared minus A squared.

Table 1: Transcript excerpt

3 Method

3.1 Research Context and Data Collection

We obtain real-world transcripts of students in a high-impact (White et al., 2022) hybrid tutoring setting through a partnership with an organization that provides tutoring services primarily for Title I schools. Groups of 2-5 students are physically present in their school classroom and connect with the tutor virtually 2-3 times a week through an online platform which integrates video conferencing with other essential classroom features, such as a whiteboard, calculators, and chat. The session recording data are collected and deidentified by the partner organization, after which the deidentified transcripts are shared with the research team under a data usage agreement. The data cannot be shared publicly. The overarching project studying tutoring and student learning has been approved by the research organization's Institutional Review Board.

3.2 Dataset

The dataset is collected across a full school year (September 2023 to June 2024) from 9th-grade students in four school districts in the US, totaling 58,383 tutorial sessions spanning 104,379 hours. Given the large dataset, we use the open-source Whisper V2 (Radford et al., 2022) automatic speech recognition system (ASR) to automatically transcribe the audio from the recording. The average word error rate (WER) is 0.33 for tutor speech and 0.57 for student speech, which is consistent with prior work finding the transcription of student speech highly challenging in real-world classrooms (Southwell et al., 2024). There are a total of 21,408,147 transcribed utterances, of which

79.3% were from tutors, 19.3% from other students and 1.4% from speakers (coaches and other tutors joining the session). Students complete a 28-item assessment of both foundational math skills as well as grade-level concepts at five intervals across the school year. Each student's mean score across the 28 items is taken as the measure of their knowledge at each time point.

3.3 Computing Linguistic Alignment

Preprocessing First, the 56,486 transcripts are separated into 495,614 conversation snippets based on inter-utterance gaps of 15 seconds or longer. As found in manual inspection of the transcripts, students often take up to 15 seconds to respond (i.e., when making calculations or thinking about an answer), but when the gap is over 15 seconds, the next utterance is likely unrelated, beginning a new sub-conversation. Snippets with only one participant are dropped, leaving 458,599 for analysis. Next, adjacent turns by the same speaker are merged, resulting in similar quantities (but not average lengths) of total student and tutor utterances.

Alignment Calculation We use word and part-of-speech token n-gram overlap as our metrics of lexical and syntactic alignment, respectively, as they are lightweight, comparable, and psychologically relevant (Duran et al., 2019, 2024; Dideriksen et al., 2023). Specifically, we use the validated Align package (Duran et al., 2019) to assist with the data preprocessing and computation of linguistic alignment. The unigram, bigram, and trigram alignment are calculated with both tokenized and lemmatized words for syntactic and lexical alignment. To determine the syntactic alignment between two turns, first the parts of speech (POS) are tagged using the NLTK part of speech tagger (Bird et al., 2009). Next the overlap of the turns is determined by creating a list of all sets of unigram, bigram, and trigram POS tags for each sentence. The n-gram POS tags are vectorized by counting the number of occurrences of each n-gram, and the cosine similarity is computed between the vectors from subsequent turns. The results are between 0 (no overlap) and 1 (complete duplication). This is repeated for the lemmatized version of the turns. A similar process is used for lexical alignment, except that unigram lemmatized words are used instead of parts of speech. Lastly, we use cosine similarity of BERT (Devlin et al., 2019) sentence embedings to calculate semantic similarity.

9	**Tutor**	What do you think B equals?
	Bigrams	What do, do you, you think, think B, B equals
	POS	WP (WH-pronoun), VBP (verb, non-3rd person singular present), PRP (personal pronoun), VB (verb, base form), NN (noun), VBZ (verb, 3rd person singular present)
	Lemmas	What, do, you, think, b, equal, ?
10	**Student**	B equals the square root of C squared minus A squared.
	Bigrams	B equals, equals the, the square, square root, root of, of C, C squared, squared minus, minus A, A squared.
	POS	NN, VBZ, DT (determiner), JJ (adjective), NN, IN (preposition), NN, JJ, CC (coord. conjunction), DT, JJ
	Lemmas	B, equal, the, square, root, of, C, square, minus, A, square, .

Table 2: Transcript excerpt with example unigram alignment processing steps. Bi- and tri-grams are additionally created for POS and lemmas.

Baseline Alignment Because alignment reflects dependencies across adjacent turns, we compute baseline alignment scores across non-adjacent turns. Specifically, we create a shuffled version of each transcript, where the tutor and student speaking order and timing are held constant, but the utterances themselves are shuffled within each speaker type – i.e., all student responses are randomly mixed and reinserted into the utterance order, and all tutor responses are randomly mixed and reinserted. Thus, we maintain the topical content of each transcript but disrupt the flow of the conversation. This accounts for the uniform nature of the semantic and lexical content of the conversation (often restricted to a single domain of math) and allows us to determine if the alignment we find is due to real changes in response to a partner's speech instead of generic speech patterns and word choice in this domain.

3.4 Measuring Student Learning Rate

A multiple-group latent growth curve model (LGCM; Duncan et al., 2013) is employed to examine the trajectory of students' mathematical achievement across the five assessment rounds. The analysis is grouped by three distinct geographic regions accounting for locational variation. A robust maximum likelihood estimator (MLR) is utilized to account for non-normality and heteroscedasticity in the data, and full information maximum likelihood (FIML; Enders and Bandalos, 2001) is implemented to handle missing values, ensuring that students with partial data remained in the analysis, preserving statistical power and minimizing bias due to missingness. The LGCM is specified with two latent factors: the intercept (I), representing students' baseline mathematical achievement and the slope (S), capturing their learning trajectory over the five assessments. The model is defined as follows:

$$\gamma_{ij} = I_j + tS_j + e_{ij}, \tag{1}$$

where γ_{ij} represents the observed assessment score for student at time t, I_j is the student's initial ability level, S_j is the rate of learning, j is the assessment time point, e_{ij} and is the residual error. The factor scores of the intercept and slope are retained as individual-level measures of a student's baseline ability (prior knowledge) and learning rate (student achievement growth), respectively. These estimates provide a measure of each student's starting proficiency and the extent of their academic progress over time, while accounting for statistical error. The average baseline ability is 0.546 (SD = .161). The average learning gain over the year is 0.0512 (SD = 0.017, $p < 0.001$), with an increase of roughly 10% per assessment round.

3.5 Data Cleaning and Merging

Actual and baseline alignment scores are computed across adjacent (or shuffled adjacent) turns, then averaged by partner pair (i.e., tutor A to student X; tutor A to student Y; student X to tutor A; and student Y to tutor A). These averages are computed across all transcripts for each partner pair (i.e., tutor A to student X and vice versa). Any scores involving non-student or non-tutor participants are dropped. As only 3.5% of turns are between students, we do not analyze student–student alignment. The data are then merged with the student achievement growth data (i.e., prior knowledge and learning rates), resulting in data from 1641 students and 103 tutors from 52,800 transcripts. Some students participate in sessions with multiple tutors, leaving 1737 partner pairs. We address outliers by winsorizing (Dixon and Yuen, 1974) all three alignment metrics as well as utterance length and number of utterances to the 99th percentile on the right side.

3.6 Variable Selection and Analytics

We select the following alignment measures for analyses. We use bigram overlap for both syntactic and lexical alignment, as bigrams allow us to measure purposeful overlap (compared to unigrams) without imposing excessive restrictions (trigrams). For syntactic alignment we use the bigram tokens instead of lemmas to preserve fine-grained information of grammatical structures. Conversely, for lexical alignment, we desire information about overlapping word choices, thus allowing different conjugations of the same word with lemmatized lexical bigram alignment. Our measure of semantic alignment is cosine similarity of sentence vectors. We use linear mixed effects (LME) models (Bates et al., 2015) for all statistical analyses due to the nesting of students within tutors. These models include a random intercept for tutor as more complex random effects structures result in convergence errors. We also include utterance length of each participant and number of utterances between a participant pair as a covariate to control for the effects or verbosity. We use the lme4 package version 1.1.35.5 (Bates et al., 2015) in R (Team, 2020) and adopt a two-tailed $p < 0.05$ significance criterion.

4 Results

4.1 Occurrence of Alignment (RQ1)

The first research question, if there is alignment in small tutor groups, is addressed by comparing actual to baseline alignment values. The histograms (Figures 1 and 2) comparing actual and baseline values for each alignment score (linguistic, syntactic, semantic) and direction (tutor to student and vice versa) show a clear separation of actual from baseline values. The separation of the distributions is most prominent for lexical alignment, indicating it is heavily turn-order dependent, and somewhat larger for syntactic than for semantic alignment.

We fit a series of mixed effects linear regression models to test for actual vs. baseline differences in alignment scores for all six variables (three alignment scores × two alignment directions). The models have the following structure:

$$\begin{aligned}\gamma_{ij} =& \beta_0 + \beta_1 Actual[vs.Baseline] \\ &+ \beta_2 UttLen + \beta_3 UttNum + u_{0j} + e_{ij}\end{aligned} \tag{2}$$

where γ_{ij} is the alignment variable, β_0 is the intercept, $\beta_1 Actual[vs.Baseline]$ are the alignment differences between the baseline and actual conditions, $\beta_2 UttLen$ and $\beta_3 UttNum$ are the effects of utterance number and length, u_{0j} is a random

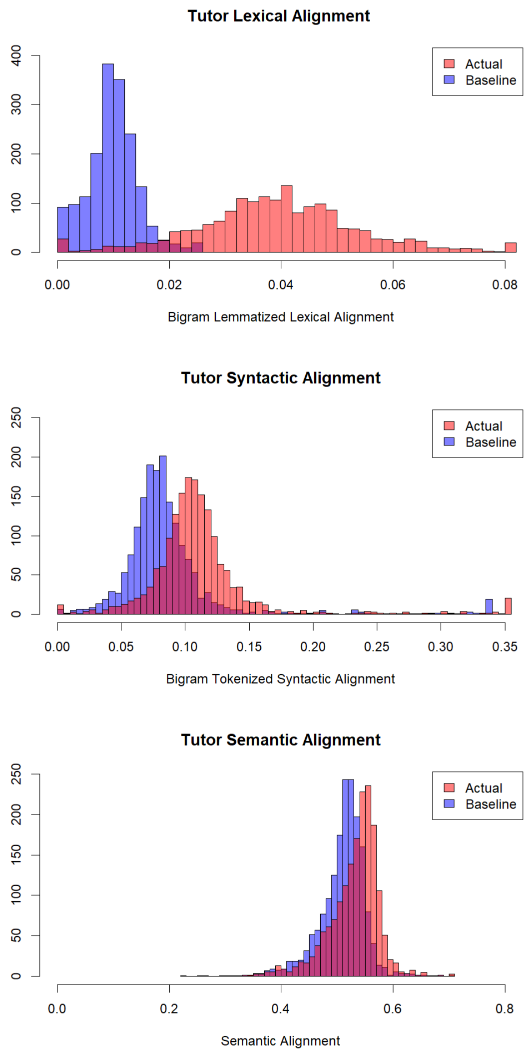

Figure 1: Distribution of tutor alignment scores

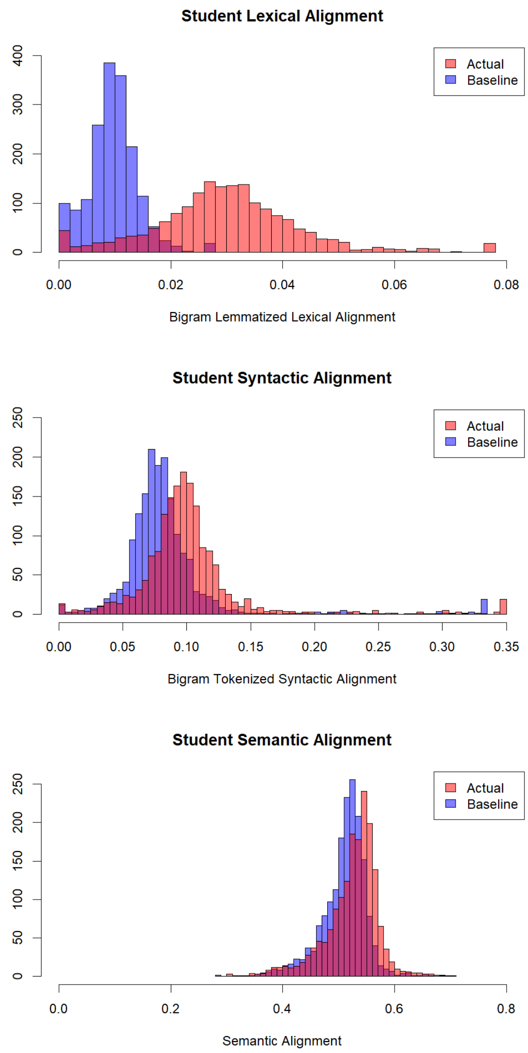

Figure 2: Distribution of student alignment scores

intercept for each tutor, and e_{ij} is an error term for student-tutor pair i under tutor j.

Results indicate that actual alignment scores are significantly higher than baseline alignment for all models (Figure 3). Effect sizes (βs), i.e., how much the actual data over the baseline effects the value of the alignment metric, conform to the following general pattern for the alignment measures (student, tutors): lexical (β_1 = 1.38, 1.61) > syntactic (β_1 = 0.31, 0.53) > semantic (β_1 = 0.14, 0.47) alignment (all ps <0.001). This shows that word choice, or lexical alignment, is more turn-order dependent, suggesting a short-term dependency compared to syntactic and semantic alignment. The effect sizes for the students are also lower in all metrics, indicating student alignment is less sensitive to turn-ordering than the tutors.

4.2 Alignment Trends (RQ2)

Alignment by Student and Tutor We investigate patterns of alignment, first by comparing how alignment varies between and within each partner direction. Specifically, we fit three mixed effects models comparing student versus tutor alignment for each metric. The structures of these models are as follows:

$$\begin{aligned}\gamma_{ij} =& \beta_0 + \beta_1 Student[vs.Tutor] \\ &+ \beta_2 UttLen + \beta_3 UttNum + u_{0j} + e_{ij}\end{aligned} \tag{3}$$

where γ_{ij} is the alignment metric, β_0 is the intercept, $\beta_1 Student[vs.Tutor]$ are the alignment differences between tutors and students.

The results reveal that all three metrics are significantly different between tutors and students. Both semantic and lexical alignment are relatively more dependent on partner direction, with the effect sizes

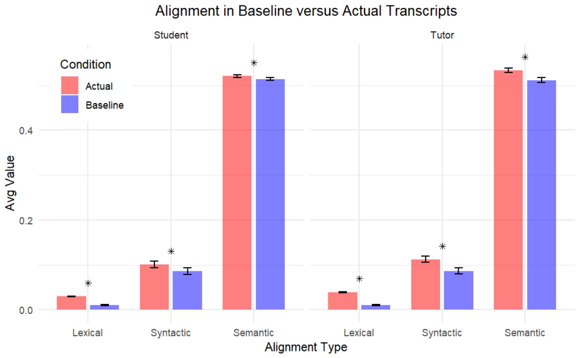

Figure 3: Mean actual alignment value for students and tutor versus the baseline ; * p < .001.

indicating greater tutor-to-student over student-to-tutor alignment ($\beta_1 = 0.49$ and $0.43, p < 0.001$) compared to syntactic alignment ($\beta_1 = 0.21$, $p < 0.001$).

	Lexical $\bar{x}$: 0.039 σ: 0.014	**Syntactic** $\bar{x}$: 0.112 σ: 0.047	**Semantic** $\bar{x}$: 0.531 σ: 0.047
Lexical $\bar{x}$: 0.030 σ: 0.013	0.457	0.300	0.589
Syntactic $\bar{x}$: 0.102 σ: 0.047	0.264	0.893	0.533
Semantic $\bar{x}$: 0.524 σ: 0.048	0.531	0.336	0.927

Table 3: Pearson correlations of conversation metrics within tutor (blue; upper triangle) and student (red; lower triangle), and between student and tutor (diagonal). The mean ($\bar{x}$) and standard deviation (σ) are shown below each variable, with tutors in the column headings and students in the rows.

Correlations among Alignment Measures Next, we investigate correlations between the three types of alignment scores for students and tutors along with student–tutor correlations. The correlation matrix, Table 3, reveals that the correlation of semantic to lexical alignment is moderate for both students and tutors (rs of 0.531 and 0.589), whereas the correlation of syntactic to semantic alignment is higher for tutors (r of 0.533) as compared to students (r of 0.336). Correlations between lexical and syntactic alignment are lower and comparable for both students and tutors (rs between 0.264 and 0.300). Syntactic and semantic alignment between students and tutors are very highly correlated (rs of 0.893 and 0.927) compared to lexical alignment (r of 0.457).

4.3 Alignment and Student Outcomes (RQ3)

Prior Knowledge and Alignment We first investigate if students' prior knowledge predicts alignment, fitting six mixed effects models:

$$\gamma_{ij} = \beta_0 + \beta_1 PriorKnowledge + \beta_2 UttLen + \beta_3 UttNum + u_{0j} + e_{ij} \quad (4)$$

where γ_{ij} is the alignment metric, β_0 is an intercept term, and $\beta_1 PriorKnowledge$ is the effect of students' prior knowledge on alignment.

We find that prior knowledge predicts alignment, even when accounting for the effects of utterance length and number (Table 4). Specifically, student prior knowledge negatively predicts the degree to which a student aligns with their tutor along all three measures of alignment. Conversely, student prior knowledge positively predicts the degree to which the tutor aligns (lexically and semantically), but not syntactically, with that student. Thus, the more knowledge a student has, the more their tutor aligns with them, and the less they align with the tutor.

Alignment and Student Learning Next, we regress student achievement growth on the alignment scores (controlling for prior knowledge and utterance length and number). To avoid multi-collinearity we average the student and tutor values for: syntactic alignment, semantic alignment, and number of utterances. Results (Table 5) indicate that tutor lexical alignment is a predictor of student achievement growth, and student lexical alignment is a negative predictor thereof. Neither syntactic nor semantic alignment are predictors. Prior knowledge is not a predictor due to being incorporated into the growth curve modeling framework used to measure student learning gains (Section 4.4). While utterance length does not predict student score growth, number of utterances (how much they participate) does.

5 Discussion

Conversation is a joint task that involves participants presenting information, then seeking or indicating understanding; a process referred to as grounding (Clark and Schaefer, 1987; Brennan, 1990), which involves some degree of both internal and synergistic cognitive processes (Dale et al.,

	Lexical Student		Syntactic Student		Semantic Student	
Predictors	β	p	β	p	β	p
(Intercept)	0.02	0.571	-0.02	0.770	0.00	0.991
Prior Knowledge	-0.08	**<0.001**	-0.09	**<0.001**	-0.07	**<0.001**
Avg Num Utt.	0.05	**0.025**	-0.02	0.483	0.09	**<0.001**
Utt Len Student	0.33	**<0.001**	0.30	**<0.001**	0.67	**<0.001**
	Lexical Tutor		**Syntactic Tutor**		**Semantic Tutor**	
Predictors	β	p	β	p	β	p
(Intercept)	0.03	0.659	0.01	0.880	0.03	0.453
Prior Knowledge	0.15	**<0.001**	0.01	0.542	0.12	**<0.001**
Avg Num Utt	0.06	**0.011**	0.00	0.975	0.12	**<0.001**
Utt Len Tutor	0.01	0.683	-0.09	**0.001**	-0.42	**<0.001**

Table 4: Regression results showing how prior knowledge and utterance length and number predict alignment for students and tutors

Predictors	β	p
(Intercept)	-0.02	0.687
Prior Knowledge	-0.01	0.597
Lexical Tutor	0.06	**0.023**
Lexical Student	-0.08	**0.004**
Syntactic Avg	-0.02	0.510
Semantic Avg	-0.01	0.729
Avg Num Utts	0.12	**<0.001**
Utt Len Tutor	0.04	0.220
Utt Len Student	-0.00	0.902
Random Effects		
σ^2	0.63	
τ_{00} tutor	0.32	
ICC	0.34	
N_{tutor}	103	
Observations	1736	
Marginal R^2 / Conditional R^2	0.019 / 0.353	

Table 5: Student Achievement Growth

2013). Unlike general conversation, in tutoring, the goal is not only mutual understanding of a narrative, but facilitating student learning of content. Thus, one basic role of the tutor is to provide explanations, assess student understanding, and provide feedback, while also scaffolding the conversation and adapting to the student (Cade et al., 2008). Conversely, a key role of the student is to listen, comprehend the content, and demonstrate understanding through problem solving or question answering. They do not especially need to adapt to the tutor. These distinct pedagogical roles of tutors and students likely influence their alignment behaviors as we elaborate on below.

Alignment Reliably Occurs in Small-Group Tutoring We first ask whether alignment occurs in small group tutoring sessions. In these sessions, the content is highly topically constrained, and the roles and amount of speech are imbalanced between students as tutors. To determine if the degree of alignment extends beyond chance overlap of phrasing, we compare actual alignment to a within-session shuffled baseline and find a significant difference between the two. Semantic alignment is most similar to the baseline, suggesting longer-term temporal dependencies of the semantic content. Lexical alignment, however, is substantially above baseline, indicating that while the content of the discourse remains stable, the specific word choices within turns fluctuates considerably. Syntactic alignment is between the two, suggesting a moderate degree of repetition of syntactic structures across the discourse.

Students Demonstrate Lower Alignment Compared to Tutors Students consistently align less than tutors across all alignment measures. As tutors are responsible for guiding the discussion and scaffolding problem solving, they might be more likely to match student phrasing and terminology and follow their lead to facilitate student comprehension and reduce the cognitive load of processing novel speech. It might also signal tutors' uptake (i.e., building on) student ideas, which is a key pedagogical strategy (MacNeilley, 1998).

Prior Knowledge predicts Lexical Alignment We further found that the degree of alignment of both students and tutors were predicted by the students' prior knowledge. The more knowledge a student has, the more (lexical and semantic) alignment they elicit from the tutor, and the less they lexically, syntactically, and semantically align to the tutor. One possibility is that students with more prior knowledge give more varied responses, such as answering questions or asking new ones. And the tutor might be acknowledging their responses through repetition. But further research is needed to confirm this and other patterns.

Lexical Alignment Predicts Student Achievement Growth We also find that, the student lexical alignment negatively predicted their learning . We speculate that high levels of student alignment might suggest superficial mimicry of the tutor's speech rather than more elaborative discussions such as such as asking questions or providing elaborations that have been shown to positively correlate with achievement (Webb, 1991). Conversely, more tutor alignment predicts better student outcomes. Perhaps more tutor alignment signals higher quality discourse and pedagogical strategies (such as rephrasal or increased uptake) where tutors use repetition to build on students' responses.

That said, further cross-dataset and causal analyses are required to thoroughly understand the underlying cognitive and behavioral mechanisms that may underlie these patterns.

5.1 Applications

Our findings have important implications regarding tutor training, automated feedback systems for tutors and other educators, and Intelligent Tutoring Systems (ITSs). While tutoring is known to be very effective, expert tutors are scarce and in high demand, and less experienced tutors may struggle to accurately assess student understanding and diagnose misconceptions (Nickow et al., 2020; Chi et al., 2004). However, tutors who receive pedagogical training engage better with students even compared to tutors trained in the subject matter (Hsiao et al., 2015), suggesting the efficacy of tutor professional learning. Tutor professional learning programs and automated feedback systems could use alignment as an additional tool to understand and guide interactions with students. Specifically, as found here, student alignment is related to their prior knowledge and learning, suggesting that it might provide a useful cue for tutoring interactions. Similarly, because tutor alignment also predicts student knowledge, it may serve as a potential feedback mechanism into the quality of their interactions. Of course, alignment should not be the only measure used, but could be coupled with other measures of tutorial discourse (Sawaya et al., 2025). In particular, lexical alignment could be important as a measure of the dynamics of ongoing tutoring sessions since it prospectively predicts learning outcomes. Thus, alignment can complement existing metrics in these curricula and systems, such as student-to-tutor talk ratio, uptake (building on ideas), and discourse moves, such as academically productive talk (Booth et al., 2024; Demszky et al., 2024). Moreover, comparing alignment scores across students within the same group could provide a measure into balanced interactions to foster more collaborative and successful tutoring sessions (Friedberg et al., 2012).

Lastly, alignment measures can also be included in ITS evaluations, which facilitate access to adaptive and personalized learning. For example, research can examine the extent to which conversational ITSs elicit similar patterns of alignment as human tutors to ascertain their naturalness.

For both tutor feedback tools and ITSs, the computational efficiency of linguistic alignment makes it particularly useful. Unlike data-intensive AI approaches, computing alignment requires no training data and is subject domain agnostic.

6 Conclusion

This study investigates linguistic alignment in small group tutoring and its association with student learning. Our findings indicate that linguistic alignment occurs naturally in tutoring interactions, with students exhibiting lower levels of alignment than tutors. We show that lexical alignment varies more in the short term, while semantic content is more consistent. We demonstrate that lexical alignment predicts learning outcomes and the relationship varies based on the source of the alignment (student vs. tutor), suggesting that alignment may serve as a meaningful indicator of tutoring processes and outcomes. The practical implications of these findings extend to tutor training, automatic feedback tools, and ITSs.

Limitations

While this work provides insights into linguistic alignment in small-group tutoring, several limitations should be acknowledged. One primary limitation is that the tutoring sessions analyzed in this study were conducted in a hybrid environment where students were in person and tutors were online, which may not fully generalize to fully in-person or remote settings. Differences in student engagement, tutor interaction styles, and external distractions may influence linguistic alignment in ways that were not captured in this dataset. Future research could compare alignment patterns across different interaction modalities to determine the extent to which these findings apply broadly. Another limitation pertains to the reliance on automatically transcribed data from an ASR system, which, despite moderate accuracy, may still introduce errors in alignment measurements. Certain linguistic features, lower-frequency vocabulary, or subtle syntactic structures may not be accurately captured, potentially affecting the alignment scores. Additionally, while we examined student–tutor alignment, we could not account for student–student interactions within the same sessions, since they occurred too infrequently. Peer discussions and collaborative problem-solving activities play a role in learning outcomes, and further investigation is needed to explore how alignment between students contributes to overall comprehension and engagement. Our study design was correlational so relationships between alignment scores and student outcomes cannot be interpreted causally. Finally, the transcripts used only represented one specific subject area of mathematics, thereby limiting the generalizability of the findings to other domains.

References

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, 67(1).

John Biggs. 1996. Enhancing teaching through constructive alignment. *Higher Education*, 32(3):347–364.

Benjamin S Bloom. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6):4–16.

Brandon M. Booth, Jennifer Jacobs, Jeffrey B. Bush, Brent Milne, Tom Fischaber, and Sidney K. DMello. 2024. Human-tutor coaching technology (HTCT): Automated discourse analytics in a coached tutoring model. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, LAK '24, pages 725–735. Association for Computing Machinery.

Lera Boroditsky. 2001. Does language shape thought?: Mandarin and english speakers' conceptions of time. *Cognitive Psychology*, 43(1):1–22.

Holly P. Branigan, Martin J. Pickering, Janet F. McLean, and Alexandra A. Cleland. 2007. Syntactic alignment and participant role in dialogue. *Cognition*, 104(2):163–197.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9):2355–2368.

Susan E. Brennan, Alexia Galati, and Anna K. Kuhlen. 2010. Chapter 8 - two minds, one dialog: Coordinating speaking and understanding. In Brian H. Ross, editor, *Psychology of Learning and Motivation*, volume 53 of *The Psychology of Learning and Motivation: Advances in Research and Theory*, pages 301–344. Academic Press.

Susan Elise Brennan. 1990. *Seeking and providing evidence for mutual understanding*. Stanford University.

Whitney L Cade, Jessica L Copeland, Natalie K Person, and Sidney K D'Mello. 2008. Dialogue modes in expert tutoring. In *International conference on intelligent tutoring systems*, pages 470–479. Springer.

Michelene T. H. Chi, Stephanie A. Siler, and Heisawn Jeong. 2004. Can tutors monitor students' understanding accurately? *Cognition and Instruction*, 22(3):363–387.

Herbert H Clark. 1996. *Using language*. Cambridge university press.

Herbert H Clark and Edward F Schaefer. 1987. Collaborating on contributions to conversations. *Language and cognitive processes*, 2(1):19–41.

Rick Dale, Riccardo Fusaroli, Nicholas D. Duran, and Daniel C. Richardson. 2013. Chapter two - the self-organization of human interaction. In Brian H. Ross, editor, *Psychology of Learning and Motivation*, volume 59, pages 43–95. Academic Press.

Dorottya Demszky, Jing Liu, Heather C. Hill, Dan Jurafsky, and Chris Piech. 2024. Can automated feedback improve teachers' uptake of student ideas? evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation and Policy Analysis*, 46(3):483–505.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arxiv:1810.04805 [cs].

Christina Dideriksen, Morten H Christiansen, Kristian Tylén, Mark Dingemanse, and Riccardo Fusaroli. 2023. Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, 152(3):864.

W. J. Dixon and K. K. Yuen. 1974. Trimming and winsorization: A review. *Statistische Hefte*, 15(2):157–170.

Sidney D'Mello. 2010. Expert tutors feedback is immediate, direct, and discriminating.

Sidney K. D'Mello and Art Graesser. 2012. Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5(4):304–317.

Sidney K. D'Mello and Art Graesser. 2023. Intelligent tutoring systems: How computers achieve learning gains that rival human tutors. In *Handbook of Educational Psychology*, 4 edition, pages 603–629. Routledge. Num Pages: 27.

Nia M. M. Dowell, Tristan M. Nixon, and Arthur C. Graesser. 2019. Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, 51(3):1007–1041.

Terry E Duncan, Susan C Duncan, and Lisa A Strycker. 2013. *An introduction to latent variable growth curve modeling: Concepts, issues, and application*. Routledge.

Nicholas D. Duran, Amie Paige, and Sidney K. D'Mello. 2024. Multi-level linguistic alignment in a dynamic collaborative problem-solving task. *Cognitive Science*, 48(1):e13398.

Nicholas D. Duran, Alexandra Paxton, and Riccardo Fusaroli. 2019. ALIGN: Analyzing linguistic interactions with generalizable techNiques—a python library. *Psychological Methods*, 24(4):419–438. Place: US Publisher: American Psychological Association.

Craig K Enders and Deborah L Bandalos. 2001. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling*, 8(3):430–457. Publisher: Taylor & Francis.

Dorothea French, Sidney D'Mello, and Katharina Wense. 2024. Aligning to adults is easy, aligning to children is hard: A study of linguistic alignment in dialogue systems. In *Proceedings of the 1st Human-Centered Large Language Modeling Workshop*, pages 81–87.

Heather Friedberg, Diane Litman, and Susannah B. F. Paletz. 2012. Lexical entrainment and success in student engineering groups. In *2012 IEEE Spoken Language Technology Workshop (SLT)*, pages 404–409. IEEE.

Riccardo Fusaroli, Joanna Raczaszek-Leonardi, and Kristian Tylén. 2014. Dialogue as interpersonal synergy. *New Ideas in Psychology*, 32:147–157.

Amy L. Gonzales, Jeffrey T. Hancock, and James W. Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37(1):3–19.

Arthur C Graesser, SIDNEY D'MELLO, and Whitney Cade. 2011. Instruction based on tutoring. *Handbook of research on learning and instruction*, pages 422–440.

Arthur C. Graesser, Natalie K. Person, and Joseph P. Magliano. 1995. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9(6):495–522.

Gabrielle Manya Halberg. 2013. Linguistic alignment in natural language generation.

Ya Ping (Amy) Hsiao, Francis Brouns, Jan van Bruggen, and Peter B. Sloep. 2015. Effects of training peer tutors in content knowledge versus tutoring skills on giving feedback to help tutees' complex tasks. *Educational Studies*, 41(5):499–512. Publisher: Routledge _eprint: https://doi.org/10.1080/03055698.2015.1062079.

Jay Kejriwal and Štefan Beňuš. 2025. Lexical, syntactic, semantic and acoustic entrainment in slovak, spanish, english, and hungarian: A cross-linguistic comparison. *Speech Communication*, page 103240.

Benjamin Lira, Margo Gardner, Abigail Quirk, Cathlyn Stone, Arjun Rao, Lyle Ungar, Stephen Hutt, Louis Hickman, Sidney K. D'Mello, and Angela L. Duckworth. 2023. Using artificial intelligence to assess personal qualities in college admissions. *Science Advances*, 9(41):eadg9405.

Lisa H MacNeilley. 1998. Opening dialogue: understanding the dynamics of language and learning in the english classroom by martin nystrand with adam gamoran, robert kachur, and catherine prendergast. *Language*, 74(2):444–445.

Ashley Micklos and Marieke Woensdregt. 2023. Cognitive and interactive mechanisms for mutual understanding in conversation. In *Oxford Research Encyclopedia of Communication*.

Andre Nickow, Philip Oreopoulos, and Vincent Quan. 2020. The impressive effects of tutoring on PreK-12 learning: A systematic review and meta-analysis of the experimental evidence.

Utku Norman, Tanvi Dinkar, Barbara Bruno, and Chloé Clavel. 2022. Studying alignment in a collaborative learning activity via automatic methods: The link between what we say and do. *Preprint*, arxiv:2104.04429 [cs].

Martin Nystrand and Adam Gamoran. 1991. Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 25(3):261–290.

James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1):547–577.

Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.

Martin J. Pickering and Simon Garrod. 2006. Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2):203–228.

Philip H. Pollock, Kerstin Hamann, and Bruce M. Wilson. 2011. Learning through discussions: Comparing the benefits of small-group and large-class settings. *Journal of Political Science Education*, 7(1):48–64.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision. Version Number: 1.

Zahra Rahimi, Anish Kumar, Diane J Litman, Susannah Paletz, and Mingzhi Yu. 2017. Entrainment in multi-party spoken dialogues at multiple linguistic levels. In *Interspeech*, pages 1696–1700.

Sandra Sawaya, Jennifer Jacobs, Robert Moulder, Chelsea Chandler, Brent Milne, Tom Fischaber, and Sidney K D'Mello. 2025. Improving tutor discourse practices via ai-enhanced coaching: A piecewise latent growth curve modeling approach. In *International Conference on Artificial Intelligence in Education*, pages 76–89. Springer.

Arabella J Sinclair and Bertrand Schneider. 2021. Linguistic and gestural coordination: Do learners converge in collaborative dialogue?. *International Educational Data Mining Society*. Publisher: ERIC.

Rosy Southwell, Wayne Ward, Viet Anh Trinh, Charis Clevenger, Clay Clevenger, Emily Watts, Jason Reitman, Sidney D'Mello, and Jacob Whitehill. 2024. Automatic speech recognition tuned for child speech in the classroom. In *Icassp 2024-2024 ieee international conference on acoustics, speech and signal processing (icassp)*, pages 12291–12295. IEEE.

Sumit Srivastava, Suzanna D. Wentzel, Alejandro Catala, and Mariët Theune. 2025. Measuring and implementing lexical alignment: A systematic literature review. *Computer Speech & Language*, 90:101731.

Angela E.B. Stewart, Hana Vrzakova, Chen Sun, Jade Yonehiro, Cathlyn Adele Stone, Nicholas D. Duran, Valerie Shute, and Sidney K. D'Mello. 2019. I say, you say, we say: Using spoken language to model socio-cognitive processes during computer-supported collaborative problem solving. *Proceedings of the ACM on Human-Computer Interaction*, 3:1–19.

R Core Team. 2020. R: A language and environment for statistical computing.

Xiaoyi Tian, Amanda E. Griffith, Zane Price, Kristy Elizabeth Boyer, and Kevin Tang. 2025. Investigating linguistic alignment in collaborative dialogue: A study of syntactic and lexical patterns in middle school students. *Language and Speech*, 68(1):63–86.

Arthur Ward and Diane Litman. 2007. Dialog convergence and learning.

Noreen Webb. 1991. Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22:366–.

Sara White, Leiah Groom-Thomas, and Susanna Loeb. 2022. Undertaking complex but effective instructional supports for students: A systematic review of research on high-impact tutoring planning and implementation. EdWorkingPaper no. 22-652. ERIC Number: ED625876.

Yang Xu and David Reitter. 2015. An evaluation and comparison of linguistic alignment measures. In *Proceedings of the 6th workshop on cognitive modeling and computational linguistics*, pages 58–67.