Beyond Static Measures: Temporal Analysis of Lexical Alignment in Human-Human Learning With a Teachable Robot

Paras Sharma[1(✉)], Daniel Fritsch[1], Yuya Asano[1], Quentin King-Shepard[1], Tyree Langley[1], Tristan Maidment[1], Diane Litman[1], Timothy Nokes-Malach[1], Adriana Kovashka[1], Nikki Lobczowski[2], and Erin Walker[1]

[1] University of Pittsburgh, Pennsylvania, USA
{pas252,djf122,yua17,quk1,tyl27,tristan.maidment,dlitman,nokes,kovashka,eawalker}@pitt.edu
[2] McGill University, Quebec, Canada
nikki.lobczowski@mcgill.ca

Abstract. Lexical alignment occurs when conversational partners converge on similar linguistic patterns. In collaborative learning settings, lexical alignment could indicate rapport, which can further predict learning and the collaborators' evolving shared understanding. Traditional approaches to alignment computation often focus more on the summary statistics computed at the end of the conversation, which usually do not capture the conversational dynamics effectively. This work investigates how alignment evolves in a conversation by modeling lexical alignment trajectories between human dyads while interacting with a teachable robot. We find that, along with the summary statistics, the alignment curve parameters and the time taken to reach key alignment moments significantly predict rapport. We further see significant relationships between the early turns in the conversation and the overall alignment trajectories, indicating the importance of modeling conversational dynamics to plan real-time interventions in the robot, with the goal of altering the alignment trajectories and, consequently, learning outcomes.

Keywords: Collaborative Learning · Temporal Alignment · Teachable Robot · Rapport

1 Introduction and Related Work

Many AIED systems focus on the interactions between one or more students and an intelligent agent to shape the learning experience [1,18,21]. Research has found that the rapport between the learners, in a collaborative setting, or the learner and the agent significantly influences engagement and learning outcomes

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. I. Cristea et al. (Eds.): AIED 2025, LNAI 15880, pp. 277–291, 2025.
https://doi.org/10.1007/978-3-031-98459-4_20

[10,12]. Lexical alignment, a phenomenon where interlocutors adopt similar vocabulary for the same underlying concepts [20], serves as a foundational mechanism for establishing this rapport [2]. This alignment leads to a higher overlap in the shared mental models of the interlocutors [15] facilitating mutual understanding, higher engagement [13], and high task-success rates [6,8,16,17,23]. A strategic manipulation of lexical alignment, e.g. by tuning an agent's lexical behaviors at certain moments in a conversation, can enhance learner understanding and engagement. However, in many dialogue studies, alignment is typically reduced to a single summary measure, for example, the final turn's alignment value [2,5], thus missing the rich dynamic patterns unfolding throughout a conversation.

Early alignment assessment methods focused on word repetition within a specific time window [24] or vocabulary alignment [9], capturing short-term effects. A crucial limitation of these alignment models is their retrospective nature in analyzing alignment, focusing on the end-of-conversation metrics rather than the alignment trajectories. Some methods focus on temporal analysis by applying a windowing strategy to analyze fine-grained fluctuations in the alignment [4,6,7]. However, they usually struggle with selecting appropriate window sizes and potentially miss the evolving trends over time due to these temporal segmentations. We suggest that instead of focusing solely on certain temporal windows, analyzing learner alignment trajectories and their properties can better predict their conversational behaviors. If relationships can be identified between different moments in these trajectories and rapport, real-time interventions, allowing for proactive modifications to conversational strategies, can be implemented in agents to influence final alignment patterns and improve rapport.

In this work, we extend the ideas of alignment in human–human-robot collaborative learning context, investigating the continuous evolution of alignment through modeling *alignment trajectories* and evaluating their ability to predict partner rapport. In our collaborative learning scenario, two students work with a teachable NAO robot to solve ratio problems. Within the scope of this paper, we focus our analysis on the human–human dyads. We first assess whether classical alignment metrics, like the final turn's value, can predict learning outcomes. Next, we perform an exploratory analysis by fitting a logarithmic model to the alignment values throughout the conversation, allowing us to extract properties of the curve, e.g., rate of convergence, intercept, slopes, and time taken to reach certain moments in alignment, and evaluate their relationship with rapport. Finally, we fit curves using a small subset of the whole conversation data and assess the predictive power of these early curves against the complete trajectory, asking whether early behavior can forecast the ultimate alignment patterns. These early indicators of learners' behaviors can enable real-time interventions in collaborative learning systems with robots. For example, in our context, the robot can continuously model these human alignment behaviors at the beginning of the conversations and then intervene to influence the outcomes. We aim to answer the following research questions:

1. **RQ1**: What alignment curves' properties predict learner behaviors?

2. **RQ2**: Does temporal information in alignment curves improve predictions over the summary metrics?
3. **RQ3**: Can early alignment behaviors predict the full trajectory?

Table 1. The table shows a set of exchanges between students A & B following a statement from Emma. The colored bounding boxes represent the repeated expressions in this student-student exchange.

Speaker	Utterance	Locutor
Emma	Now that we know how long one battery will last, can you help me figure out how many batteries I need total?	B
A	Okay. How did she come to the last one? She flipped fifteen, she did fifteen over two times?	B
B	Times two over fifteen.	A
A	Wouldn't that just make one?	B
B	Yes.	A
A	Yeah. For one whole, yeah.	B
B	In order to make one.	A
...	...	...
B	Hey, Emma, try to do the same unit for step zero and step one.	Emma
Emma	I'm not sure I understand though. Student A, can you explain more?	A

2 Methodology

2.1 System and Data

We analyze human-human conversations collected from interactions between pairs of human participants and a teachable NAO robot (**Emma**) [2]. The first study involved 40 participants, 35 women, 5 men, mean age: 19.64 (1.25); interacting virtually over 30-minute Zoom sessions, recruited from a public university in the USA. The students identified their race as White (17), Asian (13), Black (5), Latinx (1), and no answer (4). Only 28 of them worked in dyads, out of which we excluded the data for one dyad because a student in that pair did not interact with Emma. The second study was conducted with 28 participants, 18 women, 1 non-binary, 9 men, mean age: 20.67 (SD = 2.38); over 45-minute sessions in a controlled lab setting, recruited from the same university. The students identified their race as Asian (13), Black (2), Nigerian (2), White (10), and Multiracial (1). We only consider the data from 13 dyads as the other 2 students worked with a facilitator due to the absence of their partner. 3 pairs in Study 2 knew each other from before, as they were allowed to sign up with friends.

In both sessions, the learners interacted with each other and Emma to collaboratively help her solve ratio word problems posed by her, following a learning by teaching pedagogy. The interactions were relatively asymmetrical, as the learners were encouraged to discuss the solution between them before communicating with Emma. Emma functions using a combination of keyword matching and the current problem step, often helping learners refine the established understanding within her constraints, prompting further human-human dialogues. Table 1 shows an example interaction. All interactions were video recorded, and system logs were captured to track participants' task progress. All our studies followed ethical guidelines approved by the Institutional Review Board.

Each student in the dyad filled a 12 item post-survey on a 6 point Likert scale, ranging from strongly disagree to strongly agree, with questions about perceived partner rapport based on previous literature [11,19,22]. We then computed *average partner rapport* by averaging across all rapport answers to obtain a rapport metric for a student, further averaging for the students in the dyad. The mean and SD of *average partner rapport* (max possible: 6) were 5.20 (0.49) in Study 1 and 5.11 (0.41) in Study 2. We also computed average normalized learning gain (max possible: 1) based on a student's pre-test and post-test scores, with mean and SD of 0.39 (0.29) in Study 1 and 0.28 (0.29) in Study 2. It is important to note that while the rapport instrument was the same across both studies, the questions in pre-tests and post-tests, and the problem order when interacting with the robot were different. We refer to *average partner rapport* as *rapport* for the rest of this paper.

2.2 Alignment Computation

We computed lexical alignment using the concept of *shared expressions*, recurring text patterns identified at the utterance level and produced by both speakers in a dialogue [5]. An expression in a dyadic conversation becomes a "shared expression" when it is introduced by a speaker and reused by another speaker later in the conversation. We used this notion of shared expressions defined in [5] to devise our metric for alignment computation. We calculated a speaker's average repetition rate as their contribution to the alignment process. Our metric differs from [5] by differentiating between cases of speakers aligning on individual words versus sequences of n consecutive words (n-grams). This approach provides a more precise measure of alignment strength and quality, as aligning on longer sequences may reflect a stronger shared linguistic pattern than simply repeating single words. For a given speaker S, the repetition rate is defined as:

$$RepetitionRate_{S_t} = \frac{\text{Number of Repeated Expressions by S till turn t}}{\text{Total number of n-grams by S till turn t}} \quad (1)$$

where repeated expressions are n-grams shared between the speakers. This measure captures how frequently speakers *reuse expressions*, and evolves throughout the conversation as more phrases get added to the shared vocabulary.

To calculate n-grams, we preprocessed all the conversation data to make it lowercase and removed all the punctuation. Further, we limited the n-grams

to a maximum sequence of 3 words in a phrase, as many of our domain-specific vocabulary is expressed in 3-word phrases. After extracting n-grams, we removed all *stop words*: commonly used words that contribute little meaning to the conversation, such as fillers, based on NLTK's English stop word lists [3]. However, these were only removed from single-word phrases, while multi-word phrases containing stop words remained in the final n-gram set. Stop word removal was done to ensure alignment metric focuses on content-rich words reflecting learners' meaningful lexical choices, rather than just the filler words that might artificially inflate similarity measures. An example n-grams generation: Utterance:"*Hi. How are you?*", Generated n-grams: '*hi*', '*how*', '*are*', '*you*', '*hi how*', '*how are*', '*are you*', '*hi how are*', and '*how are you*'.

This measure was computed separately for both speakers at every turn in the conversation. To compute an overall alignment at any specific turn in the dialogue, we computed a weighted average of both speakers' contributions:

$$AlignmentScore_t = \frac{RepetitionRate_{A_t} \times NGrams_{A_t} + RepetitionRate_{B_t} \times NGrams_{B_t}}{NGrams_{A_t} + NGrams_{B_t}} \quad (2)$$

where $RepetitionRate_{A_t}$ and $RepetitionRate_{B_t}$ represent the respective contributions of speakers A and B till turn t. This metric is a proxy for lexical alignment over time, capturing how alignment evolves across the dialogue, with a greater number of *shared expressions*, correlating with stronger lexical alignment.

2.3 Analysis Framework

Classical Analysis. The classical analysis quantifies lexical alignment using a single alignment score, computed at the end of the conversation. We used the *Last Turn Value* metric in this paper. This metric uses the cumulative alignment score (Eq. 2) for both speakers, computed at the end of the conversation, considering the entire conversation as a context for repeated expressions calculation. Similar metrics have been used to describe alignment in collaborative learning [2], and correlate well with learning outcomes [8,17]. We also looked at *Mean Value* as a cumulative alignment score for both speakers at the end of the conversation, but found it is highly correlated with the last turn value. Since alignment increases overall in our context, we also looked at *Total Turns* in a conversation to ensure that any observed effect is not merely just its factor.

Temporal Analysis. We observed that our alignment metric follows a *logarithmic pattern* (see Fig. 1), where alignment increases rapidly in the early turns and then stabilizes over time. This log pattern emerges from how our metric is computed. As it is normalized by n-grams, there is an initial period of high alignment growth, followed by relatively stable alignment levels as the number of n-grams increases. Therefore, the curve properties, like the stabilization rate, could inform how quickly learners align with each other. To model this dynamic, we fit a *linear logarithmic curve* (Eq. 3) to the alignment score over time:

$$Y(t) = a.ln(t) + b \quad (3)$$

– $Y(t)$ represents the alignment score at turn t,

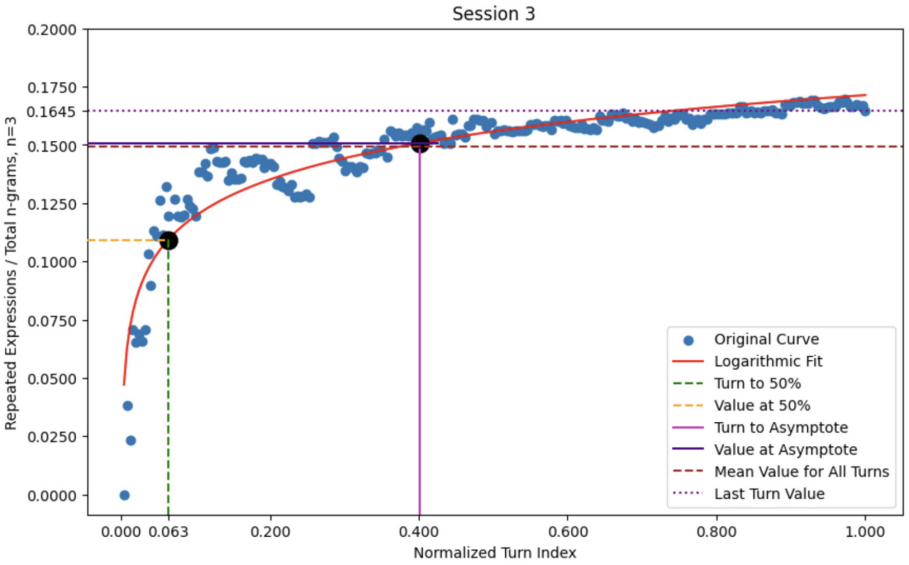


Fig. 1. The figure shows the alignment and the fitted logarithmic curves for one dyad with points marked to reflect: Mean and Last Turn Values(Actual curve) and Turn to 50%, Value at 50%, Turn to Asymptote, and Value at Asymptote (Fitted curve).

- a captures the rate of lexical alignment,
- b is the stabilization alignment level, and
- $ln(t)$ denotes the decreasing rate of change in the alignment over time.

To ensure comparability across sessions of varying lengths, we normalized the turn counts by dividing each point in time by that session's total turns, thereby bringing all conversations to the same scale. This was done to rule out any effect of conversations being too short or too long. We then performed an exploratory analysis on the *curve parameters* (growth rate and stabilization alignment level) and curve properties to evaluate their ability to predict rapport. We look at the following specific properties:

1. **Pseudo Inflection Point (Midpoint behavior)**: A logarithmic curve does not have a true inflection point (since its second derivative never changes sign), so we approximate a transition zone where the growth starts to slow down significantly. We call this point a pseudo-inflection point and approximate it as half of its asymptotic value (maximum value in bounds). We then evaluate the behavior of these curves to reach this inflection point, the time to reach this point (*Turn to 50%*), and the alignment value at this point (*Value at 50%*). This point is important as this marks a significant transition when the alignment shifts from rapid growth to a more stable, slow-changing state. Conversations, where this midpoint occurs early, could suggest fast alignment, while a later midpoint could indicate slower alignment. Based on Eq. 4, we

calculate these values as follows:

$$Y(t_{50}) = \frac{Y_{max} + Y_{min}}{2},\ t_{50} = e^{\frac{Y(t_{50})-b}{a}} \tag{4}$$

where Y_{min} and Y_{max} are the first and the last turn values of the fitted curve.

2. **Near Asymptote Point (Saturation Behavior)**: A logarithmic curve theoretically grows indefinitely but at an ever-decreasing rate. However, for our conversational scenario, the curve tends to form a plateau near later turns. We define the asymptote as the point when the slope/growth is minimal (1%) compared to the starting slope of the curve, meaning that the growth is extremely slow at this point and the curve almost reaches an asymptote. We then calculate the time to reach this point (*Turn to Asymptote*), and the alignment value at this point (*Value at Asymptote*). This point is important as it can tell if a certain conversation potentially would have reached even higher alignment values if it had lasted a bit longer. We see this in our results, for 9 sessions, this asymptote point lies outside the conversation duration, indicating potential for possible interventions for increasing alignment. Based on Eq. 5, we calculate these values as follows:

$$t_{asymptote} = \frac{a}{slope_{asymptote}},\ Y(t_{asymptote}) = a.ln(t_{asymptote}) + b \tag{5}$$

where a and b are the parameters of the fitted curve.

3. **Area Under the Curve (AUC)**: AUC is the area under the fitted logarithmic curve and captures the cumulative effect of each conversational turn in alignment. It is less sensitive to short-term fluctuations in alignment values. We calculate AUC as shown in Eq. 6:

$$AUC_t = (a \times t \times ln(t)) - (a \times t) + (b \times t), AUC = AUC_{\frac{T}{T}} - AUC_{\frac{1}{T}} \tag{6}$$

where AUC_t is the AUC until turn t, a and b are the curve parameters, T represents the total number of utterances in the session.

Early Curves Trajectories. To take advantage of the dynamic nature of the conversation and see if the early trends (e.g., first few turns) can predict the behaviors across the conversation, we fit a new set of logarithmic curves, called *early curves*. We experimented with various turn counts and found 25 to be the minimal number sufficient to represent the full alignment trajectory, so we used it as the dataset to build these curves. This also roughly coincided with the first problem boundary in our context. With a session mean of 140.65 turns (SD = 73.28), 25 turns usually covered about 18% of a session. We then compare the parameters of early curves to original curves, aiming to find trends in the early conversation that can determine rapport, answering **RQ3**.

3 Results

3.1 What Alignment Curves' Properties Predict Learner Behaviors?

We combined the samples from both studies, resulting in a dataset of 26 dyads, and used the study type as a predictor in our analysis. A Pearson's correlation

Table 2. The table shows the means and standard deviations for all the metrics used in the analysis: classical metrics, temporal metrics, early & complete curve parameters.

Classical Metrics					
	Last Turn Value		**Mean Value**		**Total Turns**
Mean (SD)	0.17 (0.05)		0.13 (0.04)		140.65 (73.28)
Complete Curve Parameters					
	RMSE		**R^2**	**a**	**b**
Mean (SD)	0.13 (0.04)		0.86 (0.1)	0.04 (0.01)	0.17 (0.05)
Early Curve Parameters					
	RMSE		**R^2**	**a**	**b**
Mean (SD)	0.07 (0.02)		0.74 (0.16)	0.03 (0.01)	−0.02 (0.02)
Temporal Metrics for Complete Curve					
	AUC	**Turn to 50%**	**Value at 50%**	**Turn to Asy.**	**Value at Asy.**
Mean (SD)	0.13 (0.05)	0.1 (0.03)	0.07 (0.02)	1.03 (0.85)	0.16 (0.04)

between combined rapport and learning (r: −0.05, p: 0.81) indicated no statistically significant relation. Given the scope of this paper, we do not report any results related to learning, as it did not have a significant relationship with rapport or our metrics. We computed all the classical metrics using the formulas in Sect. 2.3. Table 2 shows their descriptive statistics. We further calculated regressions between all the classical metrics with study type as another predictor: Last Turn Value & Mean Value (F: 144.2, p: 0.00), Last Turn Value & Total Turns (F: 8.51, p: 0.002), and Mean Value & Total Turns (F: 10.73, p: 0.00). We see that all these metrics are highly correlated with each other. Hence, we only report the results with the Last Turn Value in further analysis.

We fit a separate linear logarithmic curve for each dyad. We then calculated temporal metrics for all the dyads using the fitted curves based on the definitions in Sect. 2.3. Table 2 shows the descriptive statistics for the curve parameters (Root-mean squared error: RMSE, R^2, slope values: a, and intercept values: b) and the temporal metrics. We observed an R^2 close to 1 and an RMSE close to 0, representing a good fit. However, not all alignment trajectories conform perfectly to a logarithmic trend and exhibit fluctuations, a log model does not fully capture. Such deviations could be due to noise inherent in our n-gram alignment metric or the inherently dynamic nature of conversational alignment [4]. We further performed regression analysis (Equation 7) with these metrics as predictors and rapport as the dependent variable in separate regressions. We also added data collection as a dummy variable ($P3$) in these regressions (Study 1: 0, Study 2: 1) given the methodological changes between Study 1 & Study 2.

$$Rapport = a0 + a1 * P1 + a2 * P2 + cond * P3 \tag{7}$$

$P1$, $P2$, $P3$ are predictors, $a1$, $a2$, *cond* are respectively the coefficients of these predictors, $a0$ is the constant term, and *Rapport* is the dependent variable. This equation changes based on the number of predictors involved in the regression.

Table 3. The table shows the regression results (coefficients, p-values, and f-statistics) between all the metrics, data collection, and rapport. $a1$ is the coefficient of the predictor variable(s) in the first column and *cond* is the coefficient of the data collection variable. For curve parameters (a & b), we report 2 coefficients ($a1$ & $a2$) apart from *cond*, as we have 3 predictors (a, b, and data collection) for this regression. Values marked with * and ** are statistically significant at $p < .05$ & $p < .01$ respectively. Values with $^{+}$ have marginal significance at $p = 0.05$.

	F-statistic ($\boldsymbol{p}$)		a1 ($\boldsymbol{p}$)	cond ($\boldsymbol{p}$)
Last Turn	2.85 (0.07)		3.69^{*}(0.03)	−0.13 (0.42)
AUC	2.99 (0.07)		4.80^{*}(0.02)	−0.13 (0.42)
Turn to 50%	3.69^{*}(0.04)		-6.60^{*}(0.01)	−0.20 (0.23)
Value at 50%	3.42^{+}(0.05)		8.77^{*}(0.02)	−0.13 (0.40)
Turn to Asy.	2.50 (0.10)		0.22^{*}(0.04)	−0.17 (0.32)
Value at Asy.	0.38 (0.69)		1.86 (0.42)	−0.05 (0.76)
Curve Params a & b	**F-statistic ($\boldsymbol{p}$)**	**a1 ($\boldsymbol{p}$)**	**a2 ($\boldsymbol{p}$)**	**cond ($\boldsymbol{p}$)**
	3.89^{*}(0.02)	-25.05^{*}(0.02)	8.71^{**}(0.003)	−0.21(0.19)

Table 3 shows the results of these regressions. We see no significant effect of data collection on rapport in any case. The coefficients for *Last Turn Value*, *AUC*, *Turn to 50%*, *Value at 50%*, and *Turn to Asymptote* all have a **statistically significant relationship** ($p < 0.05$) with rapport. Notably, Turn to 50% showed a *negative* relationship with rapport, suggesting that quickly achieving a moderate alignment is more indicative of rapport [14]. The overall regression model is only statistically significant ($p < 0.05$) in the case of Turn to 50% and data collection as predictors, and marginally significant with Value at 50% and data collection as predictors. For the curve parameters, we see **statistically significant relationships** between the **coefficients of parameters** a ($p < 0.05$), b ($p < 0.01$) and rapport, along with the whole regression model being statistically significant ($p < 0.05$).

3.2 Does Temporal Information Found in the Alignment Curves Improve predictions over the Summary Metrics?

We fit 2 sets of regression models, one with Last Turn Value & data collection, and the other with Last Turn Value, data collection, and one temporal metric as predictors. We compared them using the Adjusted R^2 and F-test p-values (Comparison p) as shown in Table 4. Only the regression models including (i) Curve Parameters ($p < 0.01$) and (ii) Value at Asymptote ($p < 0.05$), along with

Last Turn Value and data collection, significantly predicted rapport. Combining Curve Parameters or Value at Asymptote, with the Last Turn Value produced a better Adjusted R^2 in both cases, with statistically significant F-test results (comparison $p < 0.01$ and < 0.05 respectively). We do see better Adjusted R^2 for some other combinations too, but the results are not statistically significant. We did not observe any multicollinearity between the pairs of predictors.

Table 4. The table compares Base regression model (Last Turn Val + Condition) with the combination regression models (Last Turn Val + Condition + Other Predictors). F-statistic and Adjusted R^2 represent the parameters of these fitted regression models. "Comparison p" represents an F-test comparing the base and combination regression models with the null hypothesis that the combination model does not add any significance over the base model. Greater Adjusted R^2 values represent better models. Values marked with * and ** are statistically significant at $p < .05$ & $p < .01$ respectively.

Predictors	F-statistic ($\boldsymbol{p}$)	Adjusted R^2	Comparison $\boldsymbol{p}$
Last Turn Val & Condition	2.85 (0.07)	0.13	
+ Curve Params (a & b)	5.69^{**}(0.003)	**0.43**	0.005^{**}
+ AUC	1.92 (0.16)	0.10	0.63
+ Turn to 50%	2.66 (0.07)	**0.17**	0.17
+ Val at 50%	2.50 (0.09)	**0.15**	0.21
+ Turn to Asy.	2.20 (0.12)	0.13	0.35
+ Val at Asy.	4.56^{*}(0.01)	**0.30**	0.02^{*}

3.3 Can Early Alignment Behaviors Predict the Full Trajectory?

We fitted "Early Logarithmic Curves" using first 25 turns of each dyadic session. Table 2 shows the descriptives of these early curves' parameters. RMSE and R^2 for these curves represent a good fit. We further compare regression models between the early and the original curve (fitted using the entire conversation data) parameters to see if early alignment behaviors could predict the overall trajectories. Table 5 shows the regression results. The coefficients of early curve parameters (a and b) are strong predictors (statistically significant with $p < 0.05$ and < 0.01 respectively) of original curve parameters, with the regression model also being statistically significant ($p < 0.05$). This suggests that early interaction dynamics can serve as predictive markers for final trajectories. We then evaluated if the alignment value at the early turn (Turn 25; Mean: 0.09, SD: 0.03)

Table 5. The table displays regression models between early and late curve parameters, early and last turn value, and early value and rapport, with study as another predictor. *a*1 & *cond* are coefficients of predictors and data collection respectively. Values marked with * and ** are statistically significant at $p < .05$ & $p < .01$ respectively.

Predictor Variable	Dependent Variable	F-statistic (***p***)	a1 (***p***)	cond (***p***)
Early Curve a	Complete Curve a	3.58*(0.04)	0.59*(0.01)	0.001 (0.79)
Early Curve b	Complete Curve b	4.49*(0.02)	−1.3**(0.009)	0.03 (0.11)
Early (Turn 25) Value	Last Turn Value	7.93**(0.002)	1.04**(0.001)	0.03 (0.10)
Early (Turn 25) Value	Rapport	1.23 (0.29)	4.39 (0.13)	−0.02 (0.89)

could predict the Last Turn Value (Mean: 0.17, SD: 0.05), and found a statistically significant relationship ($p < 0.01$; Table 5). We also assessed whether the early turn value could predict final rapport, but found no statistically significant relationship. However, since this analysis focused only on a single turn, it's uncertain whether similar relationships hold at other points in the trajectory.

4 Discussion

In this work, we analyzed the temporal behaviors of human-human lexical alignment in a teachable human–human-robot setting. We focused on human-human dialogues, as knowledge construction mostly happens in these dialogues in our context. Due to Emma's limited conversational capabilities, learners primarily interacted with each other before responding to her, fostering richer discussions.

Our curve-fitting approach showed promising results, with temporal features such as AUC, Turn to 50%, Value at 50%, and Turn to Asymptote significantly predicting rapport. A negative relationship between Turn to 50% and rapport suggests quickly achieving moderate alignment generates greater rapport [14], enabling future agentic AIED systems to plan interventions promoting early 50% alignment. Conversely, a positive association between Turn to Asymptote and rapport indicates that continuous alignment evolution signals sustained engagement in conversational partners, refining their understanding, fostering deeper rapport. These findings jointly suggest that early moderate alignment, followed by prolonged refinement of mutual understanding, supports stronger rapport.

The curve parameters, specifically, a negative slope coefficient (slower growth rate in alignment) and a higher intercept coefficient (higher stabilization alignment), were strongly associated with partner rapport. The rate at which alignment occurs and stabilizes may highlight important socio-cognitive processes fostering rapport, e.g., rapid mutual alignment may reflect quick mutual understanding, enhancing future interactions and building trust and connection.

9 out of the 26 dyadic sessions did not reach their asymptote, indicating variability in the alignment process. This absence of a clear stabilization point may reflect potential issues such as inconsistent communication, lack of engagement, or conflicting interaction styles, which may inhibit rapport development. This insight can help agentic AIED systems plan targeted interventions, facilitating smoother and more consistent alignment during conversation.

Combining fitted curve parameters and Value at Asymptote with Last Turn value produces better predictive models for rapport. While the last turn value shows the endpoint of alignment, the curve parameters depict the alignment path, and together they paint a complete picture of the alignment process. Notably, the Value at Asymptote alone did not significantly predict rapport without the last turn value in the regression model. To further investigate this finding, we analyzed the difference between the asymptote and the final observed alignment for each dyad using regression. This regression revealed a strong, statistically significant negative relationship ($p < 0.005$) with rapport, suggesting that dyads whose alignment fell short of their asymptotic potential reported lower rapport, with greater divergence associated with lower felt rapport. This result reinforces the importance of reaching alignment saturation and provides opportunities for agentic AIED systems to develop targeted interventions that help dyads achieve and sustain this asymptote alignment level throughout the interaction.

Further, the statistically significant relationship between early and complete curves indicates that early interaction dynamics already contain important cues about the eventual alignment trajectory. This presents an opportunity to design early interventions in agentic AIED systems, allowing the system to detect and alter the trajectory in real time, consequently updating rapport.

Last Turn value was a significant predictor of rapport. However, while it is a good measure for retrospective analysis, it remains limited in its applicability for real-time interventions in AIED systems. We also see that *total turns* are highly correlated with the last turn value and rapport, raising questions about the role of alignment in this relationship. This strong correlation may stem from our specific human-human-robot learning context, where longer conversations foster convergence. These extended interactions provide learners with more ways to fine-tune their mutual understanding, leading to more rapport.

Limitations and Future Work. Our study has several limitations. First, our sample size is small. Second, we only focused on human-human conversations within the human-human-robot context, which may limit generalizability. Further, our n-gram-based alignment metric may not fully capture conversational complexity. Moreover, assuming alignment curves follow a logarithmic trend may oversimplify the true underlying dynamics, as real data often exhibit fluctuations that our model smooths over. Lastly, we did not find a relationship between rapport or alignment metrics and learning. We think our assessments, graded as

correct or incorrect, are not capturing the learning people are experiencing, and we are working to refine these assessments. Despite these limitations, our work offers novel insights into the temporal dynamics of lexical alignment in collaborative human-human-robot learning interactions. This lays the foundation for our future work to design real-time, alignment-aware interventions in adaptive human-human-robot learning systems based on alignment trajectories.

Conclusion. In this work, we explore the temporal dynamics of human-human lexical alignment in a human-human-robot teaching context. We fit logarithmic curves to visualize these dynamics and use the properties of the fitted curves to predict rapport. We find that these temporal dynamics are an important predictor of how the alignment evolves and can be used to plan interventions to alter the outcomes. By integrating temporal alignment measures into our analysis, we aim to contribute to a richer understanding of how conversational coordination influences rapport and learning in human–human-robot settings.

Acknowledgments. This work was supported by Grant No. 2024645 from the National Science Foundation, Grant No. 220020483 from the James S. McDonnell Foundation, and a University of Pittsburgh Learning Research and Development Center internal award.

Disclosure of Interests. Authors have no competing interests.

References

1. de Araujo, A., Martens, J., Papadopoulos, P.M.: Enhancing student dialogue productivity with learning analytics and fuzzy rules. In: Olney, A.M., Chounta, I.A., Liu, Z., Santos, O.C., Bittencourt, I.I. (eds.) Artificial Intelligence in Education, pp. 397–404. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-64299-9_36
2. Asano, Y., et al.: Comparison of lexical alignment with a teachable robot in human-robot and human-human-robot interactions. In: Lemon, O., et al. (eds.) Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 615–622. Association for Computational Linguistics, Edinburgh, UK (2022). https://doi.org/10.18653/v1/2022.sigdial-1.57, https://aclanthology.org/2022.sigdial-1.57/
3. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python. O'Reilly Media, Inc., 1st edn. (2009)
4. Bonin, F., et al.: Investigating fine temporal dynamics of prosodic and lexical accommodation. In: Interspeech (2013). https://api.semanticscholar.org/CorpusID:12027122
5. Dubuisson Duplessis, G., Clavel, C., Landragin, F.: Automatic measures to characterise verbal alignment in human-agent interaction. In: Jokinen, K., Stede, M., DeVault, D., Louis, A. (eds.) Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, pp. 71–81. Association for Computational Linguistics, Saarbrücken, Germany (2017). https://doi.org/10.18653/v1/W17-5510, https://aclanthology.org/W17-5510/

6. Duran, N.D., Paige, A., D'Mello, S.K.: Multi-level linguistic alignment in a dynamic collaborative problem-solving task. Cogn. Sci. **48**(1), e13398 (2024). https://doi.org/10.1111/cogs.13398
7. Foltz, A., Gaspers, J., Meyer, C., Thiele, K., Cimiano, P., Stenneken, P.: Temporal effects of alignment in text-based, task-oriented discourse. Discourse Process. **52**(8), 609–641 (2015). https://doi.org/10.1080/0163853X.2014.977696
8. Friedberg, H., Litman, D., Paletz, S.B.: Lexical entrainment and success in student engineering groups. In: 2012 IEEE spoken language technology workshop (SLT), pp. 404–409. IEEE (2012). https://doi.org/10.1109/SLT.2012.6424258
9. Hirschberg, J.B., Nenkova, A., Gravano, A.: High frequency word entrainment in spoken dialogue. In: Presentation Powerpoint Slides (2008). https://doi.org/10.7916/D8NZ8DZ4
10. Li, M.: Modeling the role of rapport and classroom climate in EMI students' classroom engagement. Acta Psychologica **245**, 104209 (2024). https://doi.org/10.1016/j.actpsy.2024.104209, https://www.sciencedirect.com/science/article/pii/S0001691824000866
11. Lubold, N.: Producing Acoustic-Prosodic Entrainment in a Robotic Learning Companion to Build Learner Rapport. Ph.D. thesis, Arizona State University (2018). https://www.proquest.com/docview/2154434278?sourcetype=Dissertations%20&%20Theses, copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works. Updated 01 Nov 2023
12. Madaio, M., Peng, K., Ogan, A., Cassell, J.: A climate of support: a process-oriented analysis of the impact of rapport on peer tutoring. Grantee Submission (2018)
13. Niederhoffer, K.G., Pennebaker, J.W.: Linguistic style matching in social interaction. J. Lang. Soc. Psychol. **21**(4), 337–360 (2002). https://doi.org/10.1177/026192702237953
14. Norman, U., Dinkar, T., Bruno, B., Clavel, C.: Studying alignment in a collaborative learning activity via automatic methods: the link between what we say and do. Dialogue & Discourse **13**, 1–48 (2022). https://doi.org/10.5210/dad.2022.201
15. Perner, J., Lang, B.: Theory of mind. Dev. Psychol. Achievements Prospects, 205–230 (1999)
16. Rahimi, Z., Kumar, A., Litman, D.J., Paletz, S., Yu, M.: Entrainment in multi-party spoken dialogues at multiple linguistic levels. In: Interspeech, pp. 1696–1700 (2017). https://doi.org/10.21437/Interspeech.2017-1568
17. Reitter, D., Moore, J.D.: Alignment and task success in spoken dialogue. J. Mem. Lang. **76**, 29–46 (2014). https://doi.org/10.1016/j.jml.2014.05.008
18. Shahriar, T., Matsuda, N.: "i am confused! how to differentiate between...?" Adaptive follow-up questions facilitate tutor learning with effective time-on-task. In: Olney, A.M., Chounta, I.A., Liu, Z., Santos, O.C., Bittencourt, I.I. (eds.) Artificial Intelligence in Education, pp. 17–30. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-64299-9_2
19. Sinha, T., Cassell, J.: We click, we align, we learn: impact of influence and convergence processes on student learning and rapport building. In: Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And InfLuence, pp. 13–20. INTERPERSONAL '15, Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2823513.2823516
20. Srivastava, S., Wentzel, S.D., Catala, A., Theune, M.: Measuring and implementing lexical alignment: a systematic literature review. Comput. Speech Lang., 101731 (2024). https://doi.org/10.1016/j.csl.2024.101731

21. van Stee, E.G., Heath, T., Baker, R.S., Andres, J.M.A.L., Ocumpaugh, J.: Help seekers vs. help accepters: understanding student engagement with a mentor agent. In: Wang, N., Rebolledo-Mendez, G., Matsuda, N., Santos, O.C., Dimitrova, V. (eds.) Artificial Intelligence in Education, pp. 139–150. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-36272-9_12
22. Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Psychol. Inquiry **1**(4), 285–293 (1990). https://doi.org/10.1207/s15327965pli0104_1, http://www.jstor.org/stable/1449345
23. Ward, A., Litman, D.: Dialog convergence and learning. Front. Artif. Intell. Appl. **158**, 262 (2007)
24. Ward, A., Litman, D.J.: Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In: SLaTE Workshop on Speech and Language Technology in Education (2007)