Elisabeth André · Ryan Baker
Xiangen Hu · Ma. Mercedes T. Rodrigo
Benedict du Boulay (Eds.)

LNAI 10331

Artificial Intelligence in Education

18th International Conference, AIED 2017
Wuhan, China, June 28 – July 1, 2017
Proceedings

Springer

Elisabeth André · Ryan Baker
Xiangen Hu · Ma. Mercedes T. Rodrigo
Benedict du Boulay (Eds.)

Artificial Intelligence in Education

18th International Conference, AIED 2017
Wuhan, China, June 28 – July 1, 2017
Proceedings

Springer

Impact of Pedagogical Agents' Conversational Formality on Learning and Engagement

Haiying Li[1(✉)] and Art Graesser[2]

[1] Graduate School of Education, Rutgers University, New Brunswick, NJ 08904, USA
Haiying.li@gse.rutgers.edu
[2] Department of Psychology, University of Memphis, Memphis, TN 38152, USA
Graesser@memphis.edu

Abstract. This study investigated the impact of pedagogical agents' conversational formality on learning and engagement in a trialog-based intelligent tutoring system (ITS). Participants (N = 167) were randomly assigned into one of three conditions to learn summarization strategies with the conversational agents: (1) a *formal* condition in which both the teacher agent and the student agent spoke with a formal language style, (2) an *informal* condition in which both agents spoke informally, and (3) a *mixed* condition in which the teacher agent spoke formally, whereas the student agent spoke informally. Result showed that the agents' informal discourse yielded higher performance, but elicited higher report of text difficulty and mind wandering. This discourse also caused longer response time and lower arousal. The implications are discussed.

Keywords: Agents · Arousal · Engagement · Formality · Mind wandering · Summary writing · Teacher language · Text difficulty · Valence

1 Introduction

The present study investigated the impact of conversational agents' formality on deep reading comprehension and engagement in a trialog-based intelligent tutoring system (ITS). Formality is defined as a language style on a continuum from informal discourse to formal discourse [1]. Formal discourse, either in print or pre-planned oratory, is precise, cohesive, articulate, and convincing to an educated audience. Informal discourse, at the opposite end of the continuum, is used in oral conversation, personal letters, and narratives, which are replete with pronouns, deictic references (e.g., these, those), and verbs with a reliance on common knowledge among speakers and listeners [2]. Mixed discourse is situated between informal and formal discourses, with moderate characteristics of both formal and informal discourses. Formality increases with grade level and informational texts, but decreases with narrative texts [1–3]. The rationale and significance of this study are elaborated below.

Language is one of the most powerful tools that teachers can use to organize and implement instructional activities and engage students in learning [4]. For example, the professional use of words and phrases engages students in active and interested learning

E. André et al. (Eds.): AIED 2017, LNAI 10331, pp. 188–200, 2017.
DOI: 10.1007/978-3-319-61425-0_16

[4]. Teacher language is correlated to student language [5] and reading comprehension [6]. Agent language affects science learning [7, 8]. No studies, to date, however, have studied teacher language as a unit at integrative levels of *vocabulary*, *sentence*, *discourse*, and *genre*. Our study is interested in the effect of teacher language at these multiple textual levels on deep reading comprehension and engagement.

1.1 Teacher Language and Formality

Recently, teacher language has been classified into academic versus conversational language and it has increasingly drawn researchers' interest [9]. The majority of studies concentrated on either the relationship between teacher language and student language, or between teacher language and learning performance. For example, researchers reported that students' vocabulary skills were positively correlated with teachers' use of sophisticated, academic vocabulary and complex syntax [5]. The teachers' use of sophisticated, academic vocabulary was correlated to students' reading comprehension performance [6]. Conversely, the experiments that manipulated the computer agent language in the ITS showed that the agent's conversational style (e.g., the 1st and 2nd personal pronoun) yielded better performance on deep learning than the formal style (e.g., the 3rd personal pronoun) [7, 8]. These conflicting findings likely result from inconsistent measures of language: one at the lexical and syntactic levels [6], and one using personal pronouns [7, 8]. Neither measure represented language style as a whole, but rather only one aspect of language. Therefore, a measure of teacher language that comprehensively represents the characteristics of language is needed to further investigate the effect of teacher language on learning.

Academic language and conversational language are at two extreme ends of the formality continuum, where academic language is at one end, namely, formal language and conversational language at the other, namely, informal language) [1]. Academic language and conversational language were measured using automated Coh-Metrix formality scores (cohmetrix.com) [1, 3]. Specifically, academic language or formal language increased with word abstractness, syntactic complexity, expository texts, high referential cohesion, and high deep cohesion. Conversational language or informal language increased with word concreteness, syntactic simplicity, narrative texts, low referential cohesion, and low deep cohesion. Formality was a standardized score ($M = 0$) [1, 3]. High numbers above 0 represented more formal discourse, whereas lower numbers below 0 represented more informal discourse.

Previous research on teacher language has been confined to correlational research [5, 6] due to the difficulty in consistently manipulating teacher language in the traditional classroom setting. Some researchers resorted to a computer-based system to manipulate the computer agent's speaking style, but the manipulation was restricted to personal pronouns (*I* and *you* versus third-person) [7, 8].

The present study designed a causal study to manipulate the language styles of the conversational, pedagogical agents via an ITS, called AutoTutor [10]. Conversational, pedagogical agents are on-screen computer characters that generate speech, facial expr -essions (e.g., eyebrow-raising, eye-moving), and some gestures and facilitate instruction to the learner [11]. AutoTutor helps improve learning by almost one letter grade [10].

The present study designed a trialog between a: teacher agent, student agent, and human learner. The learner in this study is both an active learner, not a vicarious observer who learns from observing how a student agent learns from a teacher agent and overhearing their ensuing dialogues [12].

1.2 Engagement

Engagement has been categorized into three types: emotional, behavioral, and cognitive [13]. Emotional engagement reflected affective states (e.g., mood, affect, interest) and was usually measured by self-reported affective states (valence and arousal) [14]. Behavioral engagement referred to learners' participation and involvement in a learning task (e.g., effort, persistence, attention) and was usually assessed by self-reported mind wandering [14, 15]. Cognitive engagement meant investment in the task (e.g., task management, material mastery) and was usually measured by reading time [15].

Most studies on engagement and reading focused on the impact of text difficulty and/or text preference [14, 15]. Previous research has shown conflicting findings. Specifically, increasing text difficulty was found to be either beneficial [14] or detrimental [15] to engagement and learning. Some findings showed that mind wandering occurred more frequently when students conducted easy rather than difficult tasks [14]. These findings posit the executive-resource hypothesis [16] because mind wandering employed more available resources for task-unrelated thoughts. Other studies have found that participants reported more mind wandering when they read difficult texts than easy texts [15] because mind wandering was the result of executive maintenance failures (control-failure hypothesis) [17]. One possible explanation for these conflicting findings is that studies used different reading materials and experimenter-paced reading. Researchers also found that learners spent more time reading difficult texts [14, 15], but only for texts that they preferred [14].

No studies to date, however, have investigated the impact of teacher language at multi-textual levels on learning and engagement. As teacher language is one of primary tools for teachers in daily instruction, it is worthwhile to understand how teacher language impacts learning and engagement. This understanding will allow for the development of guidance for teachers and researchers on how to use language during instruction.

This study advances research on teacher language in three ways. First, the present study adopts an automated measure of formality to comprehensively measure teacher language [1–3], ranging from lexical and syntactic levels to textbase (e.g., explicit propositions, referential cohesion), situation model (or mental model), discourse genre, and rhetorical structure (the type of discourse and its composition) [1]. This multilevel measure captures teacher language as a whole rather than at separate aspects of one level, such as vocabulary [5–8] or syntax [6]. Second, this study implements a causal design to manipulate teacher language in an ITS. Third, this study bridges the gap between research on teacher language and engagement so as to provide guidance and enhance the awareness of language for teachers and researchers when they design instruction in traditional classroom settings or in computer-assisted learning and assessment environments.

2 Method

2.1 Participants

Participants ($N = 240$) volunteered for monetary compensation ($30) on Amazon Mechanical Turk, a trusted and commonly used data collection service [18]. The requirement for participants was that they were English learners who aimed to improve English summary writing. The qualified participants were randomly assigned into one of three conditions (formal, informal, and mixed) and completed a 3-hour experiment. Finally, 164 participants completed the experiments due to technical issues. This led to an uneven number of participants in each condition: $N = 46$ (Age: $M = 33.17$, $SD = 8.77$), $N = 56$ (Age: $M = 33.70$, $SD = 8.92$), $N = 62$ (Age: $M = 33.47$, $SD = 8.76$) for formal, informal, and mixed, respectively. 57% were male and 82% obtained a bachelor's degree or above. 71% participants were Asian, 16% white or Caucasian, 7% African American, 5% Hispanic, 2% other. Non-English speakers (89%) had learned English for 14.71 years on average ($SD = 9.70$).

2.2 Materials

Text. Eight short English texts (195 to 399 words) were selected from the adult literacy repository of materials (http://csal.gsu.edu) with a slight modification, consisting of four comparison texts and four causation texts [19]. Two comparison texts and two causation texts were randomly selected for tests and the balanced 4 by 4 Latin-square designs were applied to control for order effects on pretest and posttest. The remaining four passages were used for training; the same 4 by 4 balanced Latin-square design was applied. The comparison text structure connected ideas by comparing or contrasting two things/ideas/persons or alternative perspectives on a topic and showing how they were similar or different [20]. The causation texts presented a causal or cause-effect relationship between ideas [20]. Text formality of these eight texts tended to be more formal ranging from .12 to .64 according to the Coh-Metrix formality scores. Based on the Flesch-Kincaid grade level, these texts were at the grade level of 8 to 12.

Training. At the beginning of the training session, two conversational agents [11] interactively presented a mini-lecture on signal words that were frequently used in comparison texts (e.g., *similarly*, *likewise* for similarity and *differ*, *however* for differences) and causation texts (e.g., *because*, *since* for cause and *consequence*, *therefore* for effect). After participants read the passage and reported engagement (see the section of Independent Variables), agents interacted with participants and guided them to apply the summarization strategy to five multiple-choice questions. The application consisted of identifying: (1) a text structure (1 item), (2) the main ideas (1 item), and (3) the important supporting information (3 items). Thus, the summarization strategy was learned and assessed during a one-hour training session in this trialog-based ITS.

2.3 Manipulation

One expert at discourse processing generated agents' conversations in the formal and informal languages, following a five-step tutoring frame, and expectation and misconception-tailored dialogue (EMT) [11, 21]. Then another expert modified conversations based on the context. Table 1 lists an example of conversations that embodied a systematic conversational structure, which is described in Fig. 1. We annotated in brackets-with-italics some of the dialogue move categories. It should be noted that half of the Jordan responses were incorrect. Cristina always had the ground truth. Tim (the participant) needed to determine his answer based on two agents' suggestions. Therefore, the human was an active participant rather than being a merely vicarious observer. This dialogue structure improved student performance and student engagement in learning [22]. Agents delivered the content of their utterances via synthesized speech, whereas the participants clicked on or typed in their responses.

Table 1. An example of trialog.

Cristina: Tim [*Participant*], can you tell us the text structure of this text? [*Main question*]
Tim: (*Click*) Sequence. [*First trial: Wrong Answer*]
Cristina: Jordan, what do you think of this answer? [*Ask Jordan*]
Jordan: This answer might be correct. [*Jordan's incorrect response*]
Cristina: Signal words help tell the overall text structure. Sometimes, the text organization or even the title helps too. [*Hint*]
Cristina: The author uses the time sequence to talk about Kobe's and Jordan's careers. The author doesn't use sequence to organize the full text. [*Elaboration*]
Cristina: Try again. I will repeat the question. Tim, what is the text structure of this text? [*Repeat Question*]
Tim: (*Click*) Comparison. [*Second trial: Correct Answer*]
Cristina: Tim, you are absolutely right! Jordan, your answer is incorrect! [*Feedback*]
Cristina: The author first generally talks about how Kobe and Jordan are similar and different. Then it talks about them separately in each paragraph. [*Wrap-up*]
Jordan: You can see some signal words show similarities and differences, such as "two" and "different". So the correct answer is comparison. [*Wrap-up*]

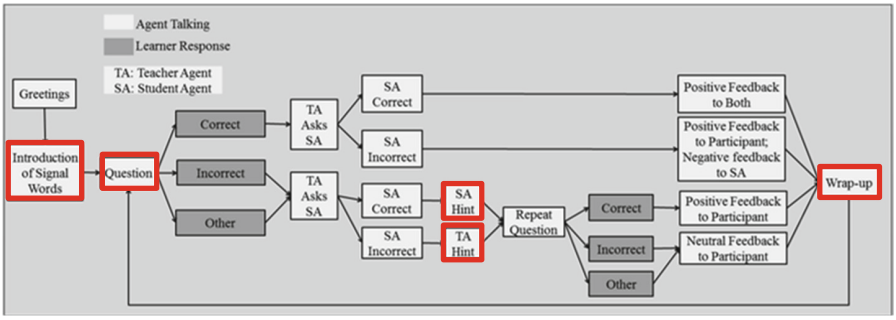

Fig. 1. Trialog moves in conversations. *Note*. Conversations in red box were manipulated. (Color figure online)

The agents' conversations in the trialog were designed in formal and informal language styles that were then assigned to the teacher agent and student agent. The agents' conversations were evaluated by the measure of formality [1, 3]. The mean of agents' formal language was 1.02 and informal, −0.37, which was consistent with humans' perception of formality when they generated conversations. The mixed language was generated by combining the formal language of the teacher agent (Cristina) and the informal language of the student agent (Jordan), and its formality score was 0.12. Based on Graesser et al.'s study [1], the agents' formality in three conditions represents three different levels of formality, ranging from informal to medium to formal. Table 2 illustrates an example of conversations in each condition when agents introduced the functions of signal words. We did not design a mixed condition where an agent's language style changed from formal to informal when common ground increased between agent and learner. The reason was that this design would cause confounds with time. When a significant effect occurred, it would be unclear whether it was caused by language style or by length of time spent learning.

Table 2. Examples of conversations in the formal and informal conditions

Cristina's formal discourse:
The signal words enable readers to determine the text structure, and consequently enhance reading comprehension. Moreover, by using the signal words, the authors guide the readers in the direction that they want them to go. The comparison text consistently compares the similarities and differences of two things or two persons.
Cristina's informal discourse:
Yes, Jordan. The author uses the signal words to lead you in the reading. The signal words help identify the text structure. They help you understand the reading better. The comparison text usually compares how things or persons are similar or different.

Note. It consisted of (A) the teacher agent, Cristina (female), (B) the student agent, Jordan (male), (C) the instruction of the presented question, (D) the text presented with the scroll down button, (E) an input text-box for participants to enter and submit their summaries or choose the answers of multiple choice questions during training, and (F) the self-paced next button.

Fig. 2. Screenshot of Interface.

2.4 Procedure

Participants first took a demographic survey, a pretest, training, and a posttest. There were two passages in the pretest: one comparison and one causation. For each passage, participants first read the passage and then self-reported engagement. Participants then

wrote the summary for the passage with the text displayed to them (see Fig. 2). The same procedure was applied to training and posttest as well. However, the training session added instruction of summarization with four texts and accordingly four summaries were written. The summary was short, between 50 and 100 words. The summary required a topic sentence that stated the main idea and important information, and students were meant to use signal words to explicitly express their ideas.

2.5 Dependent Variables and Measures

Summary Writing. The summaries that participants wrote were graded based on the rubric used in the previous studies [23] with a slight modification. The rubric included four elements: (1) topic sentence, (2) content inclusion and exclusion, (3) grammar and mechanics, and (4) signal words of text structures [19]. Each element was assessed on a scale of 0–2 points, with 0 for the absence of target knowledge, 1 for the partial presence of knowledge, and 2 for the complete presence of knowledge.

Four experts whose native language was English (1 male and 3 females) participated in the training for summary grading. At the beginning of training, they discussed each element in the rubrics and then graded three summaries of good, medium, and poor quality. Participants then started three rounds of training. Each round, they graded 32 summaries that were randomly selected from eight texts and then discussed disagreements until an agreement was reached. The average interrater reliabilities for the three training sets reached the threshold (Cronbach $\alpha = .82$). After training, each rater graded summaries for two source texts. There were 1,296 summaries in total.

Engagement. Engagement in this study was measured with the same method that Fulmer et al. [14] adopted. Emotional engagement was measured by affective states that occurred during reading. The participants reported valence and arousal using a circomplex model of affect, called Affect Grid [24]. Figure 3 shows the image of the 9×9 Affect Grid along two dimensions of valence $\times$ arousal. The valence dimension ranges from unpleasant feelings to pleasant feelings (1–9), whereas the arousal dimension ranges from low arousal (i.e., sleepiness) to high arousal (1–9). These two dimensions compressively represent the variations of affective states from positive

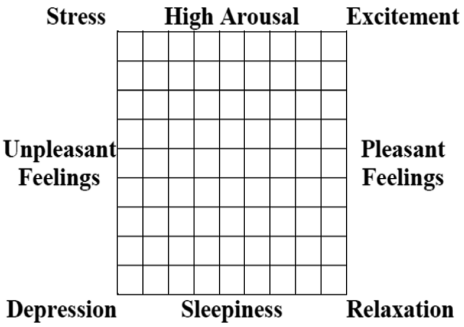

Fig. 3. Affect Grid [14, 24]

(e.g., excitement) to negative (e.g., sadness) valence, and from activating (e.g., excitement) to deactivating (e.g., relaxation) arousal [14].

Behavioral engagement was measured by mind wandering. Participants were given the definition [16]: "At some point during reading, you may realize that you have no idea what you just read. Not only were you not thinking about the text, you were thinking about something else altogether. This is called 'zoning out.'" Participants reported mind wandering once they finished reading by indicating the extent they were conducting off-task behavior during reading. This was reported on a 7-point scale with 1 as mind wandering never occurs and 7 as mind-wandering always occurs.

Cognitive engagement was measured by reading time and summary writing time. Reading time was recorded from displaying the text page to going to next page. Summary writing time was recorded from displaying the summary writing page to the submission of the summary. Both reading time and writing time were self-paced. As previous research has studied the effect of text difficulty on learning and engagement, the present study also included the perception of text difficulty that participants reported with a 6-point scale from very easy (1) to very difficult (6).

The primary independent variable (IV) was agents' formality (formal, informal, and mixed). This study consisted of two types of text structures, comparison and cause-effect, so text structure was also used as an IV. As participants consecutively wrote eight summaries, time phase was used as a repeated measure. We performed the mixed repeated ANOVA with agents' formality as a between-subjects factor, and text structure and time as within-subjects factors. All significance testing was conducted with an alpha level of .05 with Bonferroni correction for multiple analyses.

3 Results and Discussion

Table 3 displays the estimated means (standard errors) of dependent variables in the three conditions. Results showed that participants' summaries were at the medium level, but participants were highly engaged in reading and summary writing. Engagement was represented by moderate valence and arousal, and low mind wandering and text difficulty in all three conditions. Reading time was almost 2 min on average, whereas summary writing time was approximately 7 min on average.

Table 3. Estimated means and standard errors

	Summary	TD (1–6)	Valence (1−9)	Arousal (1−9)	MW (1−7)	RT (Second)	WT (Second)
Formal	4.71(.09)	2.15(.06)	5.84(.12)	6.57(.10)	1.81(.06)	116.31(4.61)	418.05(10.68)
Mixed	4.86(.08)	2.45(.05)	5.73(.10)	6.01(.08)	1.92(.06)	95.63(3.93)	390.52(9.11)
Informal	5.09(.08)	2.42(.05)	5.56(.11)	6.22(.09)	2.08(.06)	113.47(4.18)	441.71(9.68)

Note. TD = Text Difficulty. MW = Mind Wandering. RT = Reading Time. WT = Summary Writing Time. Summary = Summary Writing Scores (0–8 points).

Pearson correlations among dependent variables were performed to examine the relationships between summary writing and engagement after reading but before summary writing. Results displayed that summary scores were significantly but negatively correlated with the perception of text difficulty in three conditions, $r = -.15$, $r = -.19$, and $r = -.11$ for formal, informal, and mixed conditions ($p < .01$), respectively. Participants wrote better summaries for easy texts, which was consistent across the three conditions. Also, summary writing was significantly but negatively correlated with mind wandering in the informal ($r = -.13$, $p < .01$) and mixed conditions ($r = -.15$, $p < .01$), but not in the formal condition. Findings support the claim that mind wandering impaired learning when tasks for the informal and mixed discourses were easier to understand relative to the formal discourse. This finding is inconsistent with previous findings that mind wandering impairs learning when tasks are more difficult [15]. Valence was significantly and positively correlated with learning in the informal condition ($r = .15$, $p < .01$). Reading time before summary writing was significantly and positively correlated with summary scores in the mixed condition ($r = .12$, $p < .01$).

Results also showed that perceived text difficulty was significantly but negatively correlated with arousal ($r = -.19 \sim -.23$) and valence ($r = -.11 \sim -.29$), but positively correlated with mind wandering ($r = .38 \sim .44$) in three conditions with $p < .01$. Findings indicated that difficult texts reduced engagement because the more difficult texts were, the lower arousal and valence were, but the more mind wandering. These findings are consistent with the report that mind wandering occurs with an increase in text difficulty [15]. One possible explanation is that engagement is reduced when readers have difficulty constructing a situation model from the difficult text [15]. These results demonstrated a consistent pattern of engagement in different conditions, but an inconsistent relationship between learning and engagement. The correlation coefficients between summary writing and engagement were small because the engagement was measured before, but not after summary writing.

Mixed repeated ANOVA showed no significant two-way or three-way interactions for learning and engagement. However, there was a significant main effect of agents' formality on summary scores, $F(2, 1248) = 5.25$, $p = 0.005$. Pairwise analyses showed that the participants wrote better summaries when they interacted with agents who spoke the informal discourse than with agents who spoke the formal discourse, Cohen's $d = .63$, $p = 0.004$. This finding is consistent with previous study [7, 8] and suggests that informal discourse is easier to process than formal discourse. The informal style facilitates learners to better understand the instructional content and more successfully apply the newly-learned summarization strategy to summary writing.

Results also demonstrated a significant main effect of agents' formality on text difficulty, $F(2, 1246) = 9.09$, $p < 0.001$. Pairwise analyses showed that participants reported lower text difficulty in the formal condition than in the informal (Cohen's $d = .69$, $p = 0.001$) and mixed conditions (Cohen's $d = .77$, $p < 0.001$). This finding signifies that the agents' formal discourse is more complex and hard to process so as to cause participants to perceive that reading texts are much easier to process relative to listening to agents. Conversely, the informal and mixed discourses are simpler and easier to process, which causes participants to feel that texts are more difficult to read.

Results did not show a significant main effect of agents' formality on valence. Agents' formality, however significantly affected arousal, $F(2, 1246) = 9.66$, $p < 0.001$; mind wandering, $F(2, 1246) = 5.08$, $p = 0.006$; reading time, $F(2, 1248) = 7.45$, $p = 0.001$; and writing time, $F(2, 1248) = 7.45$, $p = 0.001$. Pairwise analyses showed that participants in the formal condition reported higher arousal than in the informal (Cohen's $d = .53$, $p = 0.024$) and mixed (Cohen's $d = .86$, $p < 0.001$) conditions. They reported lower mind wandering in the formal condition than in the informal condition, Cohen's $d = .61$, $p < 0.001$. They spent less time reading text in the mixed condition than in the formal (Cohen's $d = .67$, $p = 0.001$) and informal conditions (Cohen's $d = .57$, $p = 0.001$). They also used less time to write summary in the mixed condition than in the informal condition, Cohen's $d = .71$, $p = 0.001$.

To sum up, participants reported moderate valence and arousal, but low mind wandering and text difficulty, which represented high engagement in three conditions. Mind wandering in the informal condition, however, was higher relative to the formal condition. Interestingly, the time that participants spent reading and writing in these two conditions was not significantly different. One possible explanation, supported by the executive-resource hypothesis, is that informal discourse was easy to understand so after the first time learning summarization strategy, its execution had been automated due to unused executive resources from the primary task [15]. Consequently, mind wandering increased with the simple discourse. Furthermore, reading time and writing time were longer in the informal condition than in the mixed condition. The self-reported affective and behavioral engagement indicated that the agents' informal discourse caused higher mind wandering, which caused longer time on the task [25]. Oppositely, the cognitive engagement measured by reading and writing time showed that longer reaction times often reflected active engagement in tasks [26] due to increased efforts and persistence [27], especially when the task was a high-level processing task of reading [15]. These conflicting findings revealed that the agents' informal discourse helped learners with deeper reading comprehension than the agents' formal discourse. It is likely that participants in the informal condition reported higher mind wandering due to the fast mastery of summarization strategy.

Participants reported higher engagement in the formal condition than in the mixed condition, as indicated by low text difficulty, higher arousal, and longer time spent reading. However, this difference did not occur in summary writing. This finding implies that mind wandering was essential to successful learning. Participants spent longer time reading and writing in the informal condition than in the mixed condition, but their summary writing scores were not significantly different. This finding further demonstrates that even though the time devoted was different, learning was not affected if mind-wandering did not occur.

4 Implications and Future Directions

The present study investigated the impact of agent formality on deep reading comprehension measured by summary writing and engagement in an authentic reading and writing environment. Namely, learners can read and write in their own pace without the constraints to experimenter-paced presentations of text. This self-paced reading will not

impact mind wandering during the task [15]. Therefore, the findings more authentically reflect learners' engagement and learning, which provide implications for teachers and researchers. For example, teachers and researchers need to consider the function of teacher language during instruction and the importance of design of teacher language to foster students' deep learning and engagement. The findings can be applied to ITS as more systems have begun using natural language. To sum up, informal discourse may yield more accurate deep learning because it causes high engagement (relatively more effort represented by more response time), even though it leads to lower arousal, higher mind wandering, and higher text difficulty relative to formal condition. The relative mind wandering may elicit more effort and persistence on the high-level cognitive tasks, such as summary writing.

One limitation of the study was that we did not investigate the effect of text difficulty, text interest, or other text characteristics, such as domain-specific versus domain-general texts. These factors may affect learning and engagement along with agents' formality. Another concern was that the experiment lasted more than three hours and participants wrote eight summaries. The long-term studying may have impacted learning and engagement. In the future, the tasks may be allotted into different periods to see whether the same pattern occurs. Moreover, a future study may devise one agent that uses a mixed discourse whose formality falls between formal and informal discourse, as opposed to having the two discourses used by two distinct agents.

Acknowledgement. This work was funded by the Institute of Education Sciences (Grant No. R305C120001). Any opinions are those of the authors and do not necessarily reflect the views of these funding agencies, cooperating institutions, or other individuals.

References

1. Graesser, A.C., McNamara, D.S., Cai, Z., Conley, M., Li, H., Pennebaker, J.: Coh-Metrix measures text characteristics at multiple levels of language and discourse. Elem. School J. **115**, 210–229 (2014). doi:10.1086/678293
2. Li, H., Graesser, A.C., Conley, M., Cai, Z., Pavlik, P., Pennebaker, J.W.: A new measure of text formality: an analysis of discourse of Mao Zedong. Discourse Process. **53**, 205–232 (2016). doi:10.1080/0163853X.2015.1010191
3. Li, H., Graesser, A.C., Cai, Z.: Comparing two measures of formality. In: Boonthum-Denecke, C., Youngblood, G.M. (eds.) 2013 FlAIRS, pp. 220–225. AAAI Press, Palo Alto (2013)
4. Denton, P.: The power of our words: teacher language that helps children learn. Center for Responsive Schools Inc., Turners Falls (2013)
5. Gámez, P.B., Lesaux, N.K.: The relation between exposure to sophisticated and complex language and early—adolescent English—only and language minority learners' vocabulary. Child Dev. **83**, 1316–1331 (2012). doi:10.1111/j.1467-8624.2012.01776.x
6. Gámez, P.B., Lesaux, N.K.: Early-adolescents' reading comprehension and the stability of the middle school classroom-language environment. Dev. Psychol. **51**, 447–458 (2015). doi:10.1037/a0038868

7. Moreno, R., Mayer, R.E.: Personalized messages that promote science learning in virtual environments. J. Educ. Psychol. **96**, 165–173 (2004). doi:10.1037/0022-0663.96.1.165
8. Mayer, R.E.: Principles based on social cues: personalization, voice, and presence principles. In: Mayer, R.E. (ed.) Cambridge Handbook of Multimedia Learning, pp. 201–212. Cambridge University Press, New York (2005)
9. Snow, C.E., Uccelli, P.: The challenge of academic language. In: Olson, D.R., Torrance, N. (eds.) The Cambridge Handbook of Literacy, Cambridge, New York, pp. 112–133 (2009)
10. Graesser, A.C., Chipman, P., Haynes, B.C., Olney, A.: AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. IEEE Trans. Edu. **48**, 612–618 (2005). doi:10.1109/TE.2005.856149
11. Graesser, A.C., Li, H., Forsyth, C.: Learning by communicating in natural language with conversational agents. Curr. Dir. Psychol. Sci. **23**, 374–380 (2014). doi:10.1177/0963721414540680
12. Chi, M.T.H., Roy, M., Hausmann, R.G.M.: Observing tutoring collaboratively: Insights about tutoring effectiveness from vicarious learning. Cog. Sci. **32**, 301–341 (2008). doi:10.1080/03640210701863396
13. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School engagement: potential of the concept, state of the evidence. Rev. Educ. Res. **74**, 59–109 (2004). doi:10.3102/00346543074001059
14. Fulmer, S.M., D'Mello, S.K., Strain, A., Graesser, A.C.: Interest-based text preference moderates the effect of text difficulty on engagement and learning. Contemp. Educ. Psychol. **41**, 98–110 (2015). doi:10.1016/j.cedpsych.2014.12.005
15. Feng, S., D'Mello, S., Graesser, A.C.: Mind wandering while reading easy and difficult texts. Psychon. B. Rev. **20**, 586–592 (2013). doi:10.3758/s13423-012-0367-y
16. Smallwood, J.M., Schooler, J.W.: The restless mind. Psychol. Bull. **132**, 946–958 (2006). doi:10.1037/0033-2909.132.6.946
17. McVay, J.C., Kane, M.J.: Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). Psychol. Bull. **136**, 188–197 (2010). doi:10.1037/a0018298
18. Buhrmester, M., Kwang, T., Gosling, S.D.: Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. **6**, 3–5 (2011). doi:10.1177/1745691610393980
19. Li, H., Cai, Z., Graesser, A.C.: How good is popularity? Summary grading in crowdsourcing. In: Barnes, T., Chi, M., Feng, M. (eds.) 2016 EDM, pp. 430–435. EDM Society, Raleigh (2016)
20. Meyer, B.J.F.: Text coherence and readability. Top. Lang. Disord. **23**, 204–224 (2003). doi:10.1097/00011363-200307000-00007
21. Graesser, A.C., Keshtkar, F., Li, H.: The role of natural language and discourse processing in advanced tutoring systems. In: Holtgraves, T. (ed.) The Oxford handbooks of language and social psychology, Oxford, New York, pp. 491–509 (2014)
22. Li, H., Cheng, Q., Yu, Q., Graesser, A.C.: The role of peer agent's learning competency in trialogue-based reading intelligent systems. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M. (eds.) AIED 2015. LNCS (LNAI), vol. 9112, pp. 694–697. Springer, Cham (2015). doi:10.1007/978-3-319-19773-9_94
23. Friend, R.: Effects of strategy instruction on summary writing of college students. Contemp. Edu. Psychol. **26**, 3–24 (2001). doi:10.1006/ceps.1999.1022
24. Russell, J.A., Weiss, A., Mendelsohn, G.A.: Affect Grid: a single-item scale of pleasure and arousal. J. Pers. Soc. Psychol. **57**, 493–502 (1989). doi:10.1037/0022-3514.57.3.493

25. Smallwood, J., Davies, J.B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., Obonsawin, M.: Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention. Conscious. Cogn. **13**, 657–690 (2004). doi:10.1016/j.concog.2004.06.003
26. Smallwood, J.M., Baracaia, S.F., Lowe, M., Obonsawin, M.: Task unrelated thought whilst encoding information. Conscious. Cogn. **12**, 452–484 (2003). doi:10.1016/S1053-8100(03)00018-7
27. Clifford, M.: Students need challenge, not easy success. Edu. Leadership **48**, 22–26 (1990)