Impact of Conversational Formality on the Quality and Formality of Written Summaries

Haiying Li[1(✉)] and Art C. Graesser[2]

[1] Department of Applied Research and Services, ACT, Inc., Iowa City, IA 52246, USA
haiying.li@act.org
[2] Department of Psychology, University of Memphis, Memphis, TN 38152, USA
art.graesser@gmail.com

Abstract. This study investigated the impact of conversational agent formality on the quality of summaries and formality of written summaries during the training session and on posttest in a trialog-based intelligent tutoring system (ITS). During training, participants learned summarization strategies with the guidance of conversational agents who spoke one of the following three styles of language: (1) a formal language for both the teacher agent and the student agent, (2) an informal language for both agents, and (3) a mixed language with a formal language for the teacher agent and the informal language for the student agent. Results showed that participants wrote better quality summaries during training than pretest and/or posttest in each condition. Results also showed that agent informal language caused participants to write more informal summaries during training than on pretest. Implications are discussed for the potential application of adaptive design of conversational agents in the ITS.

Keywords: Summary writing · Agent language · Formality

1 Introduction

How to design effective language for instruction and explanations is always a controversial topic for researchers who develop computer-assisted learning environments. The question "*which language better facilitates learning, formal language or informal language*" has been investigated for decades. Formal language and informal language are two opposite ends on a continuum of formality. Formal language is precise, cohesive, and articulate independent of the context and common ground, whereas informal language is conversational, personal, and narrative dependent on the context and common grounds [2, 9, 11–13]. Both formal and informal language could be either in print or oral.

The majority of studies on agent language used personal pronouns to distinguish formal language from informal language. These studies provided empirical evidence that agents' informal language (e.g., first- or second-person pronouns) enhanced learning, reduced perceived difficulties, and increased interests in varied domains such as science

[13, 15–19] and psychology [21, 22], as well as in diverse settings ranging from research labs [21] to massive online open course (MOOC) environments [22] and from intelligent tutoring systems (ITS) [13] to educational games [18]. No studies, to date, have examined the effect of agent language on learners' use of language in writing.

The Common Core State Standards for English Language Arts (CCSS-ELA) [3] require students to develop academic writing skills and to use an academic style in their writing. The National Assessment of Educational Progress [20], however, reports that it is still a daunting challenge for secondary students to meet these standards. Therefore, it is worthwhile to conduct more research on the effect of agent language on students' academic language use to better address this challenge. The present study aimed to investigate how agent language affected participants' learning of summary writing as well as the formality of language in their written summaries.

1.1 Agent Language and Learning

Increasingly, studies have investigated how agent language affects learning through different subject matters and in diverse computer-assisted learning environments. Moreno, Mayer, and colleagues [17–19] conducted a series of experiments to test the effect of informal language (e.g., first- and second-person pronouns) that the agent used to provide instructional explanations for science in an educational game. They found that agent informal language yielded better performance on retention tests and problem-solving transfer tests. These findings were further supported by a meta-analysis study that reviewed 74 empirical studies on agent language published from 1981 to 2012 [5]. A study from a science domain, however, showed an inconsistent finding: agent informal language enhanced retention performance but did not enhance transfer performance [15]. Inconsistent findings are likely due to the different learning environments in the experiments and the different languages in learning. Specifically, the former was in an educational game and the instructional language was in English, while the latter was in a multimedia lesson with a PowerPoint show and the instructional language was in Chinese.

These inconsistent findings were also found in the domain of psychology in different settings. Reichelt et al. conducted a study in the research lab and found that the use of informal language in learning materials yielded better retention performance than the formal language, but this effect was not found on transfer performance [21]. Riehemann and Jucks also used instructional material in psychology but conducted the study in a MOOC [22]. They found that the use of informal language enhanced transfer performance.

We investigated the effect of agent language on summary writing in an ITS and found that the agent informal language enhanced better quality of summary writing [13], but this study is different from previous studies in four ways. First, this study designed a trialog rather than a dialogue or solo narrator. In the trialog, a human learner learned summarization strategies with two computer agents: one was the teacher agent, and another was the student agent. Second, the agent language was designed using three styles rather than two: the formal language for both agents, the informal language for both agents, and the mixed language by merging teacher agent formal language and student agent informal language. Third, the results that the agent informal language facilitated

summary writing was on the concentration of the main effect of agent language, namely, formal, informal, and mixed language, but did not consider the effect of agent language on posttest from pretest within each condition. Fourth, this study used multiple textual levels to measure agent language rather than merely personal pronouns as was the case in prior studies. The next section describes the multi-level measure of agent language in detail.

1.2 Measure of Agent Language

Many studies used personal pronouns to distinguish formal language from informal language. Specifically, third-person pronouns were used to produce the formal language, whereas first- and second-person pronouns were used to construct the informal language [17–19]. On one hand, the informal language with greater use of first- and second-person pronouns creates a more social environment that is more engaging for learners. Further, informal language is a familiar and everyday language, which requires less cognitive effort and is much easier to process and comprehend. Even though personal pronouns are an important indicator for formality, we could not ignore the essential roles of other language components that are used to differentiate the formal language from informal language [9, 11, 23].

We used the Coh-Metrix formality to measure agent language at multiple textual levels ranging from word, to syntax, to cohesion, and to genre [13]. Formality increases with the more use of abstract words (e.g., damage vs. hurricane), complex syntactic structures (e.g., subordinate sentences vs. simple sentences), referential cohesion (e.g., repetition of nouns vs. using pronouns to replace the repeated nouns), deep cohesion (e.g., more connectives vs. less/no connectives), and non-narrativity (e.g., third-person pronouns vs. first/second-person pronouns). This multi-level measure considers language in a holistic way rather than with one individual linguistic element. Thus, the agent language that was generated at multi-textual levels was more authentic and natural.

Our previous study examined the effect of agent language on learning. More studies are needed to explore whether agent language affects learners' use of language. The present study aimed to investigate the effect of agent language on both the quality of writing and formality of writing. Agent language was designed and developed into three styles: (1) a *formal* language for both the teacher agent and the student agent, (2) an *informal* language for both agents, and (3) a *mixed* language combining the teacher agent's formal language with the student agent's informal discourse. Specifically, this study addressed two research questions:

(1) Does agent language have an effect on the quality of participants' summary?
(2) Does agent language have an effect on formality in participants' written summaries?

This study advances research on agent language in the following two ways. First, this is the first study to unpack the effect of agent language on both learning outcomes and the use of language in writing. Findings could provide researchers with guidance on how to design language that adapts to the goals of instruction, namely, learning and/or academic writing. Second, this study reveals how agent language affects learning and use of language by comparing not only performance on pretest with posttest but also

with training. This method allows for scrutiny of the learning processes from pretest to training (i.e., learning with guidance), and to posttest (i.e., learning with guidance removed) associated with the appropriate and sufficient instructional time for effective learning and the use of academic language during and after the intervention.

2 Method

2.1 Participants

Participants were recruited from Amazon Mechanical Turk (AMT) with $30 compensation for a three-hour experiment [13]. Data collected through AMT are as trusted and reliable as those collected through traditional methods [1, 24]. Qualified participants met the criteria of being English learners and wanted to improve English summary writing. Participants were randomly assigned to three conditions: formal, informal, and mixed conditions (see the Manipulation section for details). Most participants were from India, so this study only used Indian participants to exclude the confounding of participants from different cultures. Ninety-three participants (66.4% male, $M_{Age} = 32.49$ with SD_{Age} $= 8.64$) were in three conditions: 29 in the formal condition, 29 in the informal condition, and 35 in the mixed condition. Participants had learned English for 16.39 years on average ($SD = 8.43$). They first took a demographic survey, then a pretest, training, and finally a posttest.

2.2 Materials

The reading materials were the same as Eight short expository English texts (195–399 words) in our previous study [13]. Four of them were comparison texts and four were causation texts. Two comparison texts (*Butterfly & Moth*, *Hurricane*) and two causation texts (*Floods*, *Job Market*) were randomly selected for pretest and posttest and the balanced 4 × 4 Latin-square designs were used to control for order effects. The training session used the remaining four passages (two comparison texts (*Walking and Running*, *Kobe and Jordan*) and two causation texts (*Effects of Exercising*, *Diabetes*)) and the same 4 × 4 balanced Latin-square design was applied. The causation texts displayed a causal relationship between ideas and concepts, whereas the comparison texts compared or contrasted ideas or persons and revealed their similarities and differences [20]. These texts were measured by the Coh-Metrix formality scores (.12–.64) and the Flesch-Kincaid grade level (Grade 8–12) and their text difficulties were equivalent to those for students from upper middle school to high school students.

During training, two conversational agents [13] interactively presented a mini-lecture on the function of signal words in comparison and causation texts and lists signals frequently used in comparison texts (e.g., *same/similar* signifying similarity, *differ/but* signifying differences) and in causation texts (e.g., *as/because* signifying causes, *thus/so* signifying effects). Agents then interactively guided participants to read four passages and apply the summarization strategy that they learned. Learning was assessed through five multiple-choice (MC) questions for each passage. The first MC question required participants to identify the text structure of the passage, the second to identify the main

ideas, and the last three to distinguish important supporting information from unimportant minor information (see Fig. 1). After completing the MC questions, participants were required to write a summary for the passage they just read. Conversational agents provided real-time feedback and scaffolding for the MC questions, but not for the quality of written summaries due to the lack of accurate real-time automated assessment of summaries [14]. It took participants about one hour to complete the training session in this trialog-based ITS.
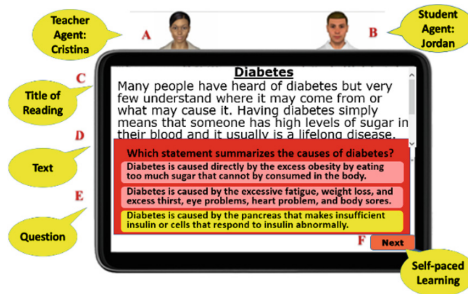

Fig. 1. Screenshot of interface during the training session.

The same procedure was applied to the pretest and posttest sessions, with the only difference in the exclusion of five MC questions along with the real-time feedback and scaffolding. One comparison and one causation texts were used on pretest and another one comparison and one causation texts were used on posttest.

2.3 Manipulation

The agents' conversations were generated by an expert at discourse processing, following a five-step tutoring frame and the expectation and misconception-tailored dialogue (EMT) [6–8]. These conversations were modified by another expert to make them more natural and authentic. Figure 2 presents this conversation mechanism: (1) the teacher agent first asked a main question, (2) the participant initiated an answer, (3) the teacher agent provides feedback and hints to help the participant seek the correct answer, (4) the agent evaluated learning by asking the question again and the participant took another try to answer the question, and (5) the agent wrapped up the question with assertion. This dialogue mechanism has been proven to enhance student learning and engagement [10]. Table 1 displays an example of the conversations that followed this mechanism. Agents delivered the content of their utterances via synthesized speech, but the participants clicked on or typed in their responses.

The agents' conversations involving mini-lecture, asking questions, providing hints, and wrapping-up questions were generated in informal and formal language styles at the multiple text levels of word usage (e.g., less vs. more frequently-used words), syntactic complexity (e.g., simple vs. complex sentences), referential cohesion (e.g., using pronouns to substitute the previous noun vs. repeating nouns), deep cohesion (e.g., less vs. more connectives to make meaning coherently), and genre (e.g., narrative vs. expository
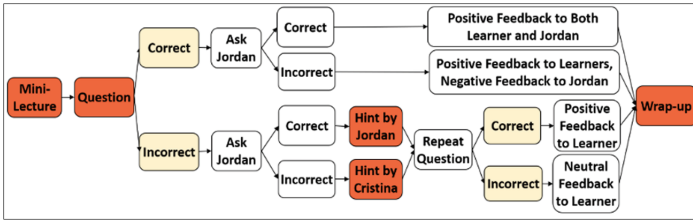
Fig. 2. Trialog moves during the training session.

Note. Conversations in red boxes were manipulated by agents' formality. Participants' responses were in yellow boxes. Jordan was the student agent. (Color figure online)

Table 1. An example of trialog during training.

Cristina: Esther [*Participant*], which statement better summarizes the main idea of this text? [*Main question*]

Esther: (*Click*) Diabetes, a lifelong disease, is caused by too little insulin or resistance to insulin and its symptoms include fatigue, blurry vision, weight loss, and excess thirst. [*First trial: Wrong Answer*]

Cristina: Jordan, it is your turn. What is your answer? [*Ask Jordan*]

Jordan: This is the correct answer. [*Jordan's incorrect response*]

Cristina: The first part of this answer elaborates on the important causes of diabetes, whereas the second part illustrates many specific symptoms, which are inappropriate for the main idea. [*Elaboration*] The main idea in the causation texts should specify the causal relationships. For instance, what causes diabetes and how diabetes affects people's health? [*Hint*] Esther, take another try. I will repeat the question. Which statement better summarizes the main idea of this text? [*Repeat Question*]

Esther: (*Click*) Diabetes is indicated by high levels of sugar in the blood and it has two types: Type 1 happens at any age; Type 2 happens in adulthood. [*Second trial: Incorrect Answer*]

Cristina: Alright. [*Neutral Feedback*] The first part of this answer informs us that people with diabetes have high levels of blood sugar, whereas the second part states two types of diabetes. The second part doesn't demonstrate the causal relationships [*Elaboration*]

Jordan: I see. The text points out two reasons. One is the pancreas. Another is cells. The third answer sums up this information. Therefore, the correct answer should be the first one: Diabetes is caused by too little insulin or resistance to insulin and can cause harmful health complications over the years. [*Wrap-up*]

Cristina: Excellent, Jordan! [*Positive Feedback*] The correct answer should be the first statement: Diabetes is caused by too little insulin or resistance to insulin and can cause harmful health complications over the years. The first part of the statement specifies the causes of diabetes, whereas the second part states the consequences of diabetes. [*Wrap-up*]

---

style). These conversations were then evaluated by the measure of the Coh-Metrix 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. Then the teacher agent formal language and the student agent informal language were mixed, which generated the mixed language. Its formality score was 0.12. The agents' formality in three conditions represents three different levels of formality: informal,

medium, and formal [9]. Table 2 illustrates an example of the formal and informal discourse when agents explained why the answer was incorrect. In Table 1, agents used the formal language when they asked the participant the main question, elaborated why the selected answer was incorrect, and provided hints (e.g., *elaboration*, *hint*, *question*), whereas they used the informal language for socialization (e.g., *"Jordan, it is your turn"*) or to provide feedback (e.g., *"Excellent"*).

Table 2. Examples of explanations in the formal language and the informal language.

Cristina's Formal discourse:
The third statement specifies that the pancreas produces insufficient insulin and that cells do not respond to insulin normally, which are the right causes of diabetes. Therefore, this statement is correct.
Cristina's Informal Discourse:
The third answer shows how people get diabetes. We can find this information from the text. This answer is correct.

2.4 Measures

Participants were required to write a summary with 50–100 words after reading each passage. They were required to state the main idea and important information with a topic sentence when generating the summaries. They were also encouraged to use signal words to explicitly express their ideas. The summaries were graded based on the rubric used in the previous studies [4, 13] in terms of four components: (1) topic sentence, (2) content inclusion and exclusion, (3) grammar and mechanics, and (4) signal words of text structures [19]. Each component was given 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) graded summaries after three rounds of training. For each round of training, 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. Four raters graded 1,296 summaries in total.

We used the text analysis tool *Coh-Metrix* (3.0) to analyze participants' written summaries. Specifically, Coh-Metrix extracted the five primary Coh-Metrix components: word concreteness, syntactic simplicity, referential cohesion, deep cohesion, and narrativity [9, 13]. We reversed the first two and the last component scores and then computed the formality scores with an average of five scores. The higher scores represented more formal summaries.

3 Results and Discussion

We performed the mixed repeated ANOVA with agent language (informal, mixed, and formal) as a between-subjects factor, text structure (causation and comparison) and time

(pretest, training, and posttest) as within-subjects factors (i.e., a repeated measure), and participants' age and year of learning English as covariates. We aimed to examine how agent language affected the quality of summaries and language use from pretest to training and then to posttest. Thus, the design of analyses included three fixed factors: the main-effect of time, the two-way interaction between time and condition, and the three-way interaction among time, condition, and text structure.

Two dependent variables were the quality of written summaries and formality of written summaries. All significance testing was conducted with an alpha level of .05 with Bonferroni correction for multiple analyses. Cohen's d was computed as an appropriate effect size. Table 3 displays the means and standard deviations of summary scores and formality scores of written summaries within each condition in each text structure on pretest, during training, and on posttest.

Table 3. Means and standard deviations of quality and formality of summaries.

Condition	Text structure	Quality of written summaries			Formality of written summaries		
		Pretest	Training	Posttest	Pretest	Training	Posttest
Formal	Causation	3.86(1.35)	4.89(1.55)	3.84(1.45)	0.57(0.58)	0.34(0.51)	0.41(0.51)
	Comparison	3.99(2.01)	4.63(1.71)	4.28(2.05)	0.16(0.47)	0.08(0.43)	0.21(0.58)
	Total	*3.93(1.70)*	*4.76(1.63)*	*4.06(1.77)*	*0.36(0.56)*	*0.21(0.49)*	*0.31(0.55)*
Informal	Causation	4.14(1.38)	5.64(1.54)	4.45(1.38)	0.55(0.52)	0.38(0.45)	0.51(0.44)
	Comparison	4.21(1.70)	4.95(1.49)	4.86(1.85)	0.30(0.56)	0.08(0.35)	0.15(0.48)
	Total	*4.17(1.53)*	*5.29(1.55)*	*4.66(1.63)*	*0.43(0.55)*	*0.23(0.43)*	*0.33(0.49)*
Mixed	Causation	3.69(1.59)	4.71(1.74)	3.79(1.84)	0.49(0.49)	0.35(0.49)	0.32(0.44)
	Comparison	3.54(2.08)	4.76(1.58)	3.97(1.74)	0.17(0.62)	0.05(0.46)	0.26(0.61)
	Total	*3.61(1.84)*	*4.74(1.65)*	*3.88(1.78)*	*0.33(0.57)*	*0.20(0.49)*	*0.29(0.53)*
Total		**3.89(1.71)**	**4.92(1.63)**	**4.18(1.75)**	**0.37(0.56)**	**0.21(0.47)**	**0.31(0.52)**

3.1 Quality of Written Summaries

The mixed repeated ANOVA analysis on the quality of summaries exhibited a significant main effect of time, $F(2, 720) = 26.64$, $p < 0.001$. Pairwise analyses (see the Quality column and the Total row in Bold in Table 3) indicated that participants wrote better quality summaries during the training session when they read the texts with the guidance of conversational agents than when they read by themselves without the guidance of conversational agents on pretest ($p < 0.001$, Cohen's $d = 0.50$) and on posttest ($p < 0.001$, Cohen's $d = 0.35$).

Analyses also displayed a significant two-way interaction for the quality of summaries, $F(6, 720) = 2.91$, $p = 0.008$. Further pairwise analyses showed that participants wrote significantly better quality summaries in each agent formality condition: $F(2, 720) = 6.33$, $p = 0.002$ for the formal condition, $F(2, 720) = 8.88$, $p < 0.001$ for the informal

condition, $F(2, 720) = 12.60$, $p < 0.001$ for the mixed condition. Further analyses (see the Total Row within each condition in Italic in Table 3) revealed that participants wrote significantly better summaries during the training session than on pretest and posttest in both formal condition and mixed condition: $p = 0.005$, Cohen's $d = .41$ for pretest and training and $p = 0.028$, Cohen's $d = .34$ for posttest and training in the formal condition; and $p < 0.001$, Cohen's $d = .52$ for pretest and training and $p = 0.003$, Cohen's $d = .40$ for posttest and training in the mixed condition. In the informal condition, participants wrote significantly better quality summaries during the training session than on pretest, $p < 0.001$, Cohen's $d = .5$. Results did not show a significant three-way interaction.

These findings indicated that conversational agents could facilitate participants writing better summaries when agents guided them to read the texts during training than on pretest and posttest when the guidance was removed. Participants benefited from the guidance provided by agents, no matter what language formality that agents used, formal, informal, or mixed. Unfortunately, when agents' guidance was removed on posttest, participants read and processed texts and then wrote summaries independently, and the quality of their written summaries was not significantly different from that on pretest. These findings imply that intervention on text structure could effectively facilitate summary writing, but would likely be insufficient for participants to master these skills and apply them towards completing summary writing tasks independently. Participants probably need more assistance to complete summary writing tasks successfully.

3.2 Formality of Written Summaries

The analysis on the formality of summaries that participants constructed showed a significant main effect of time, $F(2, 720) = 7.32$, $p < 0.001$. Pairwise analyses (see the Formality column and the Total row in Bold in Table 3) showed that participants wrote more informal summaries during the training session than on pretest ($p = 0.001$, Cohen's $d = 0.25$). This pattern was not found on posttest and training. Further analyses for each condition (see the Total Row within each condition in Italic in Table 3) showed that this pattern was only found in the informal condition, $F(2, 720) = 3.23$, $p = .040$. Participants wrote more informal summaries during training than on pretest ($p = .038$, Cohen's $d = 0.32$). Results indicated a significant three-way interaction, $F(9, 720) = 6.73$, $p < .001$, but further pairwise analyses showed no significant effects.

These findings suggest that the agent informal language more easily influenced participants' use of informal language. Specifically, participants tended to imitate agents' informal discourse when they generated their summaries, possibly because the informal language is more familiar and much easier as it requires less cognitive effort. Fortunately, when participants wrote summaries without agents' guidance, the effect disappeared. We did not find that the agent formal or mixed language affected participants' language use in their written summaries. One explanation is that the training session focused on the instruction of text structures, not the use of formal language in writing. Another explanation is that the directions for summary writing did not require participants to write summaries in the formal language. The third explanation is that the formal language is more complex and using the formal discourse would take more effort and time. Thus, participants tended to use the more familiar language to save effort and time in the situation where they were not explicitly required to write formally.

4   Conclusions and Implications

In this study, we investigated the impact of conversational agent formality on learning of text structures, concentrating on the quality of participants' written summaries and formality of their written summaries during learning processes that included independent reading and writing on pretest, reading with guidance during training, and independent reading and writing on posttest in a trialog-based ITS. During the training session, participants learned with the guidance of conversational agents who spoke one of three styles of language: a formal language for both the teacher agent and the student agent, an informal language for both agents, and a mixed language with the formal language for the teacher agent and the informal language for the student agent.

We found that when participants were guided to read texts, they wrote better quality summaries. When the guidance was removed, they wrote summaries as well as on the pretest. Our findings were inconsistent with prior studies: agents' informal language facilitated learning. The reason for the ineffective intervention on posttest is likely that summary writing involves complex cognitive processes, which may take participants, particularly language learners a longer time to master the skills and successfully apply them to a new task independently. The reason for the effective intervention during training is that even if summary writing is a difficult and complex task, participants could benefit from the intervention no matter what language the agents use. Our findings imply that one-hour intervention is insufficient for summary writing. Further studies are needed to investigate how long participants, including English-native speakers and language learners, need agents' guidance and what is the best time to remove the guidance so that they could successfully complete summary writing tasks independently.

Moreover, we found that participants' use of language was affected by agents' use of language only when agents spoke with the informal language. This finding implies that participants' use of language is potentially affected by the agents' use of language. This study provides evidence that we might facilitate participants' use of language through the use of similar language by agents. The reason for the ineffectiveness of the agents' formal or mixed language on participants' use of formal language is possibly that we did not specify the use of formal language in writing or provide explicit instruction for formal language. Further studies are needed to examine whether agent formal discourse could elicit learners' use of formal language through explicit instruction and requirement in the writing prompts.

This study has some limitations, which could be addressed in future studies, as aforementioned. Another restriction, which has not been mentioned, is the measure of agent language. Previous studies on agent language used personal pronouns, which are easier to manipulate. The present study used multiple textual-level measures to measure agent language, including word abstractness, syntactic complexity, referential cohesion, deep cohesion, and non-narrativity, which are implicit and complex. Future studies will focus on certain specific language and discourse features at each level, provide instruction on the use of these features in academic writing, and investigate the effectiveness of the intervention.

This study contributes to research on agent language in the following three ways. First, this fine-grained analysis unpacked learning processes and informed researchers and educators of the potential for intervention for complex learning tasks. Second, the

findings of effective learning during training in all conditions suggest that further studies on agent language need to consider the design of language holistically and comprehensively. Third, this study is the first one to lay a foundation that agent language affects participants' use of language and encourage researchers to design more learning environments to facilitate academic writing.

References

1. 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). https://doi.org/10.1177/1745691610393980

2. Clark, H.H.: Using Language. Cambridge, New York (1996). https://doi.org/10.1017/cbo9780511620539

3. Common Core State Standards Initiative: Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects (2010). http://www.corestandards.org/ELA-Literacy/

4. Friend, R.: Effects of strategy instruction on summary writing of college students. Contemp. Educ. Psychol. **26**, 3–24 (2001). https://doi.org/10.1006/ceps.1999.1022

5. Ginns, P., Martin, A.J., Marsh, H.W.: Designing instructional text in a conversational style: a meta-analysis. Educ. Psychol. Rev. **25**, 445–472 (2013). https://doi.org/10.1007/s.10648-013-9228-0

6. Graesser, A.C., Chipman, P., Haynes, B.C., Olney, A.: AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. IEEE Trans. Educ. **48**, 612–618 (2005). https://doi.org/10.1109/TE.2005.856149

7. 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, pp. 491–509. Oxford, New York (2014)

8. Graesser, A.C., Li, H., Forsyth, C.: Learning by communicating in natural language with conversational agents. Curr. Dir. Psychol. Sci. **23**, 374–380 (2014). https://doi.org/10.1177/0963721414540680

9. 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. Sch. J. **115**, 210–229 (2014). https://doi.org/10.1086/678293

10. 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). https://doi.org/10.1007/978-3-319-19773-9_94

11. 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)

12. 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). https://doi.org/10.1080/0163853X.2015.1010191

13. Li, H., Graesser, A.: Impact of pedagogical agents' conversational formality on learning and engagement. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 188–200. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61425-0_16

14. 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)

15. Lin, L., Ginns, P., Wang, T., Zhang, P.: Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain? Comput. Educ. **143**, 103658 (2020). https://doi.org/10.1016/j.compedu.2019.103658

16. 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, New York (2005)

17. Mayer, R.E., Fennell, S., Farmer, L., Campbell, J.: A personalization effect in multimedia learning: students learn better when words are in conversational style rather than formal style. J. Educ. Psychol. **96**, 389–395 (2004). https://doi.org/10.1037/0022-0663.96.2.389

18. Moreno, R., Mayer, R.E.: Personalized messages that promote science learning in virtual environments. J. Educ. Psychol. **96**, 165–173 (2004). https://doi.org/10.1037/0022-0663.96.1.165

19. Moreno, R., Mayer, R.E.: Engaging students in active learning: the case for personalized multimedia messages. J. Educ. Psychol. **92**, 724–733 (2000). https://doi.org/10.1037/0022-0663.92.4.724

20. NAEP: 2015 Reading Assessment [Data file] (2015). http://nces.ed.gov/nationsreportcard/subject/publications/stt2015/pdf/2016008AZ4.pdf

21. Reichelt, M., Kämmerer, F., Niegemann, H.M., Zander, S.: Talk to me personally: personalization of language style in computer-based learning. Comput. Hum. Behav. **35**, 199–210 (2014). https://doi.org/10.1016/j.chb.2014.03.005

22. Riehemann, J., Jucks, R.: "Address me personally!": On the role of language styles in a MOOC. J. Comput. Assist. Learn. **34**, 713–719 (2018). https://doi.org/10.1111/jcal.12278

23. Snow, C.E., Uccelli, P.: The challenge of academic language. In: Olson, D.R., Torrance, N. (eds.) The Cambridge Handbook of Literacy, pp. 112–133. Cambridge, New York (2009)

24. Snow, R., O'Connor, B., Jurafsky, D., Ng, A.Y.: Cheap and fast–but is it good? Evaluating non-expert annotations for natural language tasks. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pp. 254–263. ACM, New York (2008)