The relationship between theory of mind and multiple-document comprehension in university students: preliminary research findings

Christian Tarchi[1] · Lidia Casado-Ledesma[1] · Giulia Sanna[2] · Margherita Conti[2]

Abstract

The demands of learning in the twenty-first century require being skilled in the use and comprehension of multiple documents. Some individual factors such as the metacognitive skill of theory of mind (ToM) are related to this ability. This study investigated the relationship between university students' ability to comprehend multiple documents, measured through an argumentative essay task, and their ToM. The participants were 84 undergraduates. We assessed their ToM (using the strange stories task) and their mental state talk (using a think-aloud protocol during the reading of multiple documents—MST_R—and through the content of their subsequent written essays—MST_W). The relation between students' ToM and their MST_W was mediated through MST_R. In addition, students' ToM was significantly associated with the quality of the essays, through MST_R, and the relation between MST_R and the argumentative quality was mediated through MST_W. MST_W was the only significant predictor of the quality of the essays. The involvement of ToM in multiple-document comprehension suggests, from an educational perspective, the possibility to support multiple-document comprehension by scaffolding mental state talk.

Keywords Multiple-document comprehension · ToM · Mental state talk · University students

✉ Lidia Casado-Ledesma
lidia.casadoledesma@unifi.it

Christian Tarchi
christian.tarchi@unifi.it

Giulia Sanna
giulia.sanna5@stud.unifi.it

Margherita Conti
margherita.conti2@stud.unifi.it

1 Department of Education, Languages, Intercultures, Literatures, and Psychology, University of Florence, Via San Salvi, 12-50125 Florence, Italy

2 University of Florence, Via San Salvi, 12-50125 Florence, Italy

🖄 Springer

Introduction

The twenty-first century is characterized by increased access to information, which can be overwhelming in many ways. While on the one hand Web users are increasingly making use of pictures and videos (in social networks and on YouTube, for instance) to learn what they need to learn, reading documents still plays a fundamental role in learning. Indeed, from an educational perspective, multimedia material, and videos more in specific, seems to have a detrimental effect on the depth of learning (see *shallowing hypothesis*, Annisette & Lafreniere, 2017; see comparative studies on video versus text learning, e.g., Tarchi et al., 2021). Thus, investigating how people learn from digital texts is very relevant, considering also that text is still one of the principal modalities used to transmit knowledge in the digital world.

In this sociohistorical moment, it is especially necessary to perform a critical reading and be able to integrate content across multiple documents, which often illustrate diverse and even contradictory views on the same topic (Alexander, 2012). Web operators tend to use the information that they gather about the users' preferences to provide them with the contents or opinions they like best. This would feed an availability bias, that is, a tendency to assess as more reliable information that comes readily to mind and/or that is easy to retrieve. This bias, along with the myside bias (i.e., a tendency to assess as more reliable information that is aligned with our own prior beliefs), supports one-sided reasoning when learning about complex topics. Reflective citizens need to be equipped with the skills allowing them to represent a topic from multiple perspectives, in order to form rationale and informed decisions and participate in the public discourse.

Although the integration and evaluation of textual information are now considered essential skills in democratic societies, researchers and educators in many countries agree that they are not adequately developed through schooling, even at the university level. In fact, there is evidence that a large number of students have problems with the demands of processing more than a single text (Britt & Rouet, 2012). Understanding the factors that influence multiple-document comprehension is essential for designing interventions that promote this skill at all educational stages.

Several individual variables have been related to performance in multiple-document comprehension (e.g., prior topic beliefs (Richter & Maier, 2017), prior knowledge (Bråten et al., 2014a, b; Strømsø et al., 2010), and need for cognition (Dai & Wang, 2007; Kardash & Scholes, 1996)). However, other individual variables that may also be related to comprehension of single and multiple texts have received less attention, as they have traditionally been studied more from the field of developmental psychology than from the field of educational psychology; such is the case of the metacognitive variable of theory of mind (ToM). Theory of mind involves understanding that other people have mental states that guide their behavior, and it is an ability that develops across childhood (Wimmer & Perner, 1983). ToM is a skill that becomes more sophisticated over the years, helping to understand different dimensions of subjectivity (Wellman & Liu, 2004), which is a key element in comprehending information, sometimes contradictory, presented in different texts.

Given the important role of ToM in children's development, recent studies have empirically analyzed the relationship between this variable and reading comprehension of single and multiple documents in primary school students (Atkinson et al., 2017; Boerma et al., 2017; Florit et al., 2020; Kim, 2017). Nevertheless, there are no known studies that have related it to performance in multiple-document reading comprehension tasks in the adult population. There is evidence that ToM levels are not stable throughout the life course. In

fact, past research has shown a decline in ToM in old adults versus young adults (Cavallini et al., 2013; Maylor et al., 2010). Taking into account that multisource comprehension tasks are in high demand in university contexts, it is worth examining the extent to which ToM levels in this population help predict student performance. This study aims to address this gap in the literature.

For this purpose, we investigated the relationship between university students' ability to comprehend multiple documents, measured through an argumentative essay task, and their ToM. We distinguished between their ToM (assessed by the strange stories task) and the spontaneous use of ToM ability during the reading and writing processes (measured by the mental state talk index). In doing so, we took into account some control variables (prior topic beliefs, prior topic knowledge, and need for cognition) following literature indicating their association with multiple-document comprehension.

Multiple-document comprehension

The information society in which we live has generated a scenario in which readers, on many occasions, have to consult various documentary sources to form a global idea on numerous subjects (Rouet, 2006). This not only happens when we approach scientific topics, but the multiplicity of information is present even in everyday aspects that affect decision-making. Achieving a comprehensive understanding of a subject presented through different documents is, therefore, a necessary competence for readers that requires a high level of demand. First, when relevant information is presented in different documents rather than in a single text, it is necessary to infer relationships and establish coherence between documents (Rouet et al., 2017). Moreover, it is extremely likely that, in this process, the reader will come across information that contradicts his or her previous beliefs or contrary information on the topic (Richter & Maier, 2017; Stadtler & Bromme, 2007). Likewise, not all information comes from equally reliable sources, especially in a sociohistorical moment in which fake news abound (Miller & Bartlett, 2012). Many of these factors are not exclusive of multiple-document comprehension, but the challenges increase when we approach an area of knowledge from sources that present different perspectives. In these circumstances, synthesizing or integrating coherent, componential, or conflicting information is required (Bråten et al., 2014a, b).

In recent years, there has been a significant increase in research and theoretical models that address multiple-document comprehension. One of the theoretical frameworks that best explains how comprehension of multiple texts occurs is the Documents Model (Perfetti et al., 1999). According to this model, there are a series of cognitive, linguistic, and metacognitive factors that interact in the construction of different dimensions of representation: a situation model for each text, plus an intertext model and an integrated model (Britt & Rouet, 2012). Readers need to understand each text and compose a coherent representation of its content, taking into account the concepts and propositions presented in it and relating them to relevant knowledge about the topic (situation model). In addition, readers need to create an intertext model, which includes two types of connections: (1) between the text content and the source or the text's document information and (2) across texts using intertext predicates. Also, readers need to elaborate an internal representation that integrates content across documents, comparing consistencies and reconciling discrepancies among them (i.e., integrated mental model) (Perfetti et al., 1999).

The development of an integrated mental model about texts allows the reader to address multiple tasks that usually involve the generation of a piece of writing. Argumentative

essays have been used to assess comprehension of multiple documents in previous research (e.g., Anmarkrud et al., 2014; Barzilai et al., 2015; Florit et al., 2020; LaRusso et al., 2016). Of course, the use of argumentative writing has its own limitations. Writing is a complex process, and differences in writing skills within the sample may represent a potentially confounding variable. For instance, in a study by Weston-Sementelli et al. (2018), writing proficiency was associated with source-based essay quality in undergraduate students. Nevertheless, argumentative writing provides the opportunity to explore not only what participants comprehend when reading multiple texts but how participants use what they have read when reporting their perspective. Moreover, the argumentative essay provides a written protocol that allows to analyze the spontaneous use of ToM ability. In the present study, we assess multiple-document comprehension through the quality of the students' argumentative essays, understanding by quality the level of integration of the different perspectives showed in the source texts. As we were mainly interested in the ability to identify perspectives and supporting arguments, but wanted also to minimize the impact of confounding variables such as writing competence, we adopted a coding scheme targeting whether participants were able to present multiple viewpoint on a controversial topic and identify justifications for each viewpoint (Barzilai & Eshet-Alkalai, 2015).

Individual variables and multiple-document comprehension

The role of individual difference factors is widely recognized and considered in the latest models and theoretical frameworks on the use of multiple texts (e.g., Barzilai & Strømsø, 2018; Bråten et al., 2011; List & Alexander, 2019). In this study, we focused on prior topic beliefs, prior topic knowledge, and need for cognition as they have been found as relevant dimensions of individual differences in studies on young adults when reading multiple documents. When reading about controversial topics, prior topic beliefs play an important role. According to the myside bias, people tend to favor one-sided reasoning (with the chosen side overlapping with what they already believe in) rather than considering multiple perspectives when thinking (Baron, 1995). Myside bias is active also when reading multiple documents, a phenomenon captured by the two-step validation model (Richter & Maier, 2017). According to it, prior topic beliefs are used by the reader as routine cognitive processes to monitor incoming information and verify whether it is consistent or not with active knowledge and beliefs. If readers are not highly engaged or motivated, they may use prior topic beliefs to validate (i.e., evaluate the plausibility of incoming information) on the basis of its consistency with preexisting beliefs (text-belief consistency effect).

While prior topic beliefs capture the position that people hold on a certain topic, prior topic knowledge refers to the amount of accurate knowledge that people have on a topic. The goal of reading comprehension is ultimately to create a coherent representation of textual content by embedding it within an existing cognitive schema. When reading, people need to possess relevant, accurate, specific, and coherent knowledge in order to draw inferences on the text content and create a coherent representation of the situation as described in the text (Kintsch, 1988; McCarthy & McNamara, 2021). While several studies have confirmed the relevance of prior topic knowledge on comprehension of single texts (see Bittermann et al., 2023), recently, its importance has been extended to multiple-text comprehension. There is evidence that learners' prior knowledge directly and indirectly influences comprehension of multiple texts (Bråten et al., 2014a, b) and that the effects of task instructions (e.g., writing summaries or argumentative essays) on comprehension and integration processes of multiple texts are moderated by learners' prior knowledge (Gil et al.,

2010). Besides helping to evaluate the plausibility of written information, prior knowledge supports the ability to make intertextual inferences across documents (e.g., Bråten et al., 2014a, b; Strømsø & Bråten, 2009). For instance, in a study conducted by Hagen et al. (2014), students' topic knowledge was positively related to students' deeper-level integration processing, evaluated through spontaneous note-taking and through self-reports of strategy use. In summary, and based on the relationship between prior knowledge and multiple-text comprehension processes, several studies have controlled the impact of this variable to analyze the contribution of other factors to multiple text comprehension (e.g., Bråten & Strømsø, 2010, 2011).

In relation to the need for cognition, there is empirical evidence about the positive relation between this variable and reading comprehension. Need for cognition is a thinking disposition to enjoy cognitively demanding activities (Cacioppo & Petty, 1982), which includes multiple-document comprehension. Need for cognition was indirectly associated with intertextual integration in a multiple-document comprehension task via the mediation of deeper-level strategies (Bråten et al., 2014a, b) and engagement and effort (Tarchi & Villalón, 2021). Moreover, a study conducted by Kardash and Scholes (1996) revealed that undergraduate readers' need for cognition predicted their ability to write conclusions integrating opposing perspectives on the controversial issue they had been reading about. Need for cognition was also found to be associated with multiple-document comprehension (Bråten et al., 2014a, b; Tarchi & Villalón, 2021).

ToM: a relevant individual factor in multiple-document comprehension?

Theory of mind (ToM) is defined as the ability to infer mental states in order to understand and predict others' behaviors (Wimmer & Perner, 1983). ToM develops across childhood. A large body of research has explored antecedents of individual differences in the developmental progression of ToM (Hughes & Devine, 2015). At the same time, individual differences in ToM development across childhood predict several outcomes such as social competence (Devine et al., 2016), academic achievement (Lecce et al., 2011), and reading comprehension.

According to Florit et al. (2020), the relationship between ToM and reading comprehension, both of narrative and informative texts, as well as single and multiple documents, can be explained on the basis of three assumptions. The first explanation would be that the greater the capacity to understand mental states, the greater the capacity to represent the perspective of the characters in narrative texts. The latter, in turn, would contribute to a better comprehension of texts. A second explanation, also valid for the comprehension of informative texts, would be that texts are representations of the world that can be interpreted in different ways. Therefore, understanding texts would also require the ability to reflect on one's own and others' cognitive processes and their relationship to behaviors, events, and other mental states. The third explanation would allude to the contribution of ToM in the emergence of a mature epistemic understanding of the nature of knowledge. ToM is closely linked to epistemic cognition, which concerns the beliefs we hold about knowledge and knowing (Hofer, 2002). However, studies that have traditionally addressed issues related to theory of mind have been carried out with children, whereas research on epistemic cognition has focused more on explicit beliefs about epistemological issues and their role in learning and multiple-text comprehension among older children, adolescents, and adults (Barzilai & Strømsø, 2018; Chinn et al., 2011; Greene et al., 2008).

In sum, these three explanations suggest that ToM can play a relevant role in the comprehension of narrative and informative texts as well as single and multiple texts. Indeed, several empirical studies have verified the contribution of ToM in multiple and single-text comprehension (see Atkinson et al., 2017; Boerma et al., 2017; Dyoniziak et al., 2023; Florit et al., 2020; Kim, 2017; LaRusso et al., 2016). Despite the large body of literature addressing the development and correlates of ToM in childhood, only recently researchers have begun to pay attention to ToM in older individuals. There are some direct and indirect evidence even young adults may vary in their ability to manipulate and predict others' mental states. Some studies have addressed the relation between ToM and cognitive functions in the adult population (Cavallini et al., 2013) or the positive impact of reading literary fiction on the improvement of ToM skills in adults (Kidd & Castano, 2013). Indeed, a study conducted by Dumontheil et al. (2010) suggests that the online usage of theory of mind (assessed through a perspective-taking task) improves between late adolescence and adulthood. The authors suggest that it is particularly the ability to inhibit prepotent responses that improves, a hypothesis further supported by Brown-Schmidt (2009). In Brown-Schmidt's study, the results suggested that people use perspective information when processing online language. Occasional insensitivity to perspective may partly depend on difficulties in inhibiting perspective-inappropriate interpretations. Valle et al. (2015) found that young adults still struggle in third-level theory of mind as assessed through recursive thinking ("I think that you think that he/she thinks that another person thinks…"). Apperly et al. (2010) tested adult subjects with a task in which they had to interpret instructions from a speaker with limited knowledge. They found that participants often failed to take lack of knowledge into account when performing the task, probably because of their limited ability to hold in mind the speaker' perspective and use it to guide their responses.

Nevertheless, to the best of our knowledge, there is no previous study analyzing the relationship between ToM and multiple-document comprehension in young adults. Considering that theory of mind does not remain stable throughout the life course (e.g., Cavallini et al., 2013; Maylor et al., 2010) and that there is evidence that even the adult population shows difficulties in putting it into practice (Keysar et al., 2003), it is worth analyzing its predictive role in the performance in multiple-text comprehension tasks.

Theory of mind tasks

ToM can be assessed employing several tasks at multiple levels of mental state understanding (e.g., the Sally & Ann task (Wimmer & Perner, 1983), MASC (Dziobek et al., 2006), RMET (Baron-Cohen et al., 2001), the director task (Wu & Keysar, 2007), or strange stories (Happè, 1994)). Most of the tests have been validated for children (e.g., the false belief task). However, some of these tasks are inappropriate for assessing ToM in an adult population, as the performance of a neurotypical adult would be maximal before completing the test, and therefore, the person's real potential in this ability could not be inferred. In this sense, it is necessary to apply assessment methods adjusted to the study population. For example, the strange stories task has often been used as a test of theory of mind for older individuals (Cavallini et al., 2013).

Another important issue, in addition to applying age-appropriate assessment methods to the population under study, is to adopt an ecological perspective when assessing theory of mind. There is evidence that many psychological mechanisms are not universal, but are affected by the culture in which individuals are embedded (Heine et al., 1999). This being so, tests used to assess theory of mind should not only assess this ability in a broad sense

but should be consistent with the everyday activities that people perform in a given historical moment and cultural context. In line with this idea, it is also possible to infer ToM by considering the mental state discourse, which can be defined as the set of words employed to attribute thoughts, feelings, emotions, and desires to people (Bretherton & Beegley, 1982; Lecce et al., 2019). Assessing ToM employing mental state talk has some benefits over more traditional methods such as "false belief" task. Two clear advantages are the ecological validity (mental state talk can be analyzed in naturally occurring language) and the possibility of considering a wider range of internal states (Hughes et al., 2010). Importantly, mental state talk has been validated as a ToM measure in the adult population (see Barnes et al., 2009; Hao et al., 2010; Lecce et al., 2019). Assessing ToM through mental state talk allows to infer the application of this skill in specific teaching–learning activities, such as multiple-text comprehension and writing tasks. Therefore, the use of not only an experimental test to assess ToM but also a more ecological and contextual method, such as the mental state talk, is another of the present study's contributions.

Research questions and hypotheses

In light of the current literature, the aim of the study was to investigate the role of university students' ToM in their performance in a multiple-document comprehension task. We distinguished between the level of ToM skill (assessed by the strange stories task) and the spontaneous use of ToM ability during the reading (MST_R) and writing processes (MST_W). To this aim, we implemented a think-aloud protocol to identify the thinking processes activated while reading and asked participants to write an argumentative essay, to identify the level of integration between perspectives. Past studies have used the think-aloud methodology to investigate multiple-document comprehension (see Ferguson et al., 2012). Although some authors may positively (Duke & Pearson, 2002) or negatively (Bowles & Leow, 2005) influence learning performances, recent evidence supports the notion that think-aloud does not influence argumentative reasoning when reading multiple documents (Tarchi, 2021). Both productions (think-aloud protocols and written outputs) were analyzed in terms of mental state talk. We used a reading–writing task about the controversial topic of flu vaccination.

We included the following control variables: prior topic knowledge, prior topic beliefs, and need for cognition. While this study focuses on the association between ToM and multiple-text comprehension, it is important to investigate such association after controlling for the effect of other variables. Indeed, low levels of prior knowledge, skewed prior beliefs, or low levels of engagement (i.e., need for cognition) may hinder the readers' ability to interpret the text through mental states. Thus, we included prior knowledge, prior beliefs, and need for cognition as control variables to investigate the additional contribution that ToM can offer to explain variance in multiple-text comprehension. Moreover, the inclusion of these control variables will contribute to our understanding of the role played by individual differences in multiple-text comprehension by replicating prior findings.

Three research questions (RQs) guided the study.

1. RQ1: are ToM, MST_R (assessed during the reading process), and MST_W (assessed during the writing process) associated? We hypothesized a positive relation between ToM, MST_R, and MST_W.
2. RQ2: are ToM and MST_{R-W} associated with multiple-document comprehension (inferred through argumentative quality) when the roles of prior topic knowledge, prior topic

beliefs, and need for cognition are taken into account? We hypothesized an association between ToM, MST_{R-W}, and multiple-document comprehension over all controls.

3. RQ3: are ToM, MST_R, MST_W, and argumentative quality directly or indirectly associated? This last research question is exploratory, as no prior literature can be used as a reference to make a specific research hypothesis.

Method

Participants

Participants were 84 undergraduate students from the central area of Italy ($M_{age} = 24.07 \pm 3.51$ years; 75% females). They were enrolled in several faculties (i.e., Psychology, Education, Social Sciences, Engineering, Law, and Business). Students were recruited by trained psychology students at the University of Florence (Italy) who presented the experiment during a class, in agreement with lecturers. Those who volunteered (per email) were contacted by research assistants. Written informed consent was obtained from all participants. All students spoke Italian as their primary language and were relatively homogeneous (i.e., middle class) regarding socioeconomic status. None of the study participants were diagnosed with learning disabilities or neurodevelopmental disorders at the time of the study. The study followed all the guidelines of the Declaration of Helsinki (World Medical Association, 2013) and was in line with the University Ethics Committee guidelines.

Materials

Students received six different digital documents, with the instruction of reading them and writing an argumentative essay. All six documents discussed the topic of flu vaccination, but differed from each other in terms of trustworthiness and position (two documents presented a pro-vaccine position, two documents presented an against vaccination stance, and two documents presented a neutral stance towards vaccination). The comprehension of these multiple texts requires second-order or higher-order recursive mentalistic reasoning. That is, the texts required the ability to reason about what a person believes or thinks about mental states held by a second person (Devine & Hughes, 2016). The documents have been used in prior studies (Tarchi, 2021). The texts were between 442 and 573 words long and were similar in readability (calculated by the Gulpease index, ranging between 34 and 48, which identifies texts that are readable by people with a high school degree, Lucisano & Piemontese, 1988). Trustworthiness was manipulated by including scientifically or socially validated sources (health ministry or encyclopedia) versus less accurate sources (blogs or magazine for laypeople). The six documents were all presented as if they were printed pages, in a pdf format.

Measures

Multiple-document comprehension

Students read the six different documents regarding the topic of flu vaccination, with the purpose of writing an argumentative essay reporting their opinion based on the information included in the texts. After the reading, they were asked to write an essay to answer the

following question: *would you recommend a person with heart problems to get a flu vaccination?* To answer this question, they had to rely on the information provided by the texts. While writing the essay, they had access to the texts.

Essays were coded for the argument about the advantages and disadvantages of flu vaccination. The arguments provided an index of deep comprehension, reflecting whether participants considered and integrated the different perspectives presented in the six source texts. The coding system used to assess the quality of the essays has been employed in previous studies (e.g., Tarchi & Villalón, 2021) and was based on the scheme proposed by Barzilai and Eshet-Alkalai (2015). Students' essays were scored at four levels. Table 1 shows a description of each level.

Two independent raters, who had received a specific training, coded all the essays, achieving a good score of agreement ($k = 0.91$).

Theory of mind (ToM)

An Italian version of the strange stories task (Happè, 1994) was administered as a measure of ToM. The material consists of 18 ToM stories and 6 control stories, which evaluate the skill to infer mental states by interpreting nonliteral statements. Students were asked to read the stories. They then wrote their answer to a mental state question to explain why a character in the story said something that was not literally true. Answers were scored according to the criteria defined in previous literature (Lecce et al., 2014). The score for each story ranged from 0 to 2, where 0 was given for absent or incorrect answers, 1 was given for partially correct answers (reference to physical or factual states), and 2 was given for fully correct answers (reference to mental states).

Two independent raters coded the stories. The overall agreement (Cohen's kappa) was 0.89.

Mental state talk (MST)

Students' mental state talk was assessed in students' think-aloud protocols while reading and also in argumentative essays. In both of them, we proceeded to identify terms and expressions referring to mental states. We identified terms describing seven different categories: physiological facts, world perception, positive and negative feelings or emotions, willingness to achieve something, thoughts, moral perspective, and sociorelational states. Each term corresponded to a code assigned to words or expression in participants' verbal

Table 1 Coding system to assess argumentative quality

Level	Description of the level
1	No claim is provided, or a claim is provided but not justified
2	Only one viewpoint is mentioned: a single position on the topic is provided and supported by a series of reasons
3	Both viewpoints are mentioned but only one is justified: participants take into account both pros and cons of vaccination and presented both viewpoints. However, only one position was justified, that is, supported by reasons
4	Both viewpoints are presented and justified: subjects present both pros and cons of vaccination and support claims of both sides with reasons

and written productions. To control for individual differences in verbosity, we calculated a total word score by summing all words used. The adjusted MST index was calculated according to the following formula (Lecce et al., 2019; Rosso et al., 2015):

$$\text{MST adjusted index} = \frac{N \text{ mental state term}}{\sqrt{N \text{ total word}}}$$

Transcripts were coded by two independent raters who received specific training. The interrater reliability was 0.93.

Prior topic beliefs (control variable)

The task was adapted from Maier and Richter (2013) and assessed prior topic beliefs about vaccines through a 10-item self-report instrument on a 6-point Likert scale (1 = completely disagree; 6 = completely agree). An example of item is "I think that vaccines are the most effective way against infectious diseases." The reliability was $\alpha = 0.89$. Against-vaccine items were reversed, and sum scores were calculated. Thus, a high score (near to 60) would represent a strong pro-vaccination attitude, whereas a low score (near to 10) would represent a strong against-vaccination attitude. Scores between 30 and 40 would correspond to a neutral attitude towards the topic.

Prior topic knowledge (control variable)

Twelve multiple-choice questions were devised for assessing prior topic knowledge about flu and vaccination. An example of item is "Vaccines are available for these diseases, except for: A. Tetanus; B. Hepatitis A; C. Flu." The reliability was $\alpha = 0.67$. Right answers were summed up, and students' scores could range between 0 and 12.

Need for cognition (control variable)

Need for cognition was measured through the rational-experiential inventory (Epstein et al., 1996). It consists of two unipolar scales, one measuring rational thinking (derived by the need for cognition scale, 20 items; e.g., "I would prefer complex to simple problems") and the other measuring experiential thinking (faith in intuition, 20 items; e.g., "My initial impressions of people are almost always right"). In this study, we only considered the rational thinking scale. Participants scored each item on a 5-point Likert scale (from 1 = completely false to 5 = completely true). The instrument was translated from English to Italian and then back-translated for control. Scores could range between 20 and 100. The reliability of the instrument was $\alpha = 0.83$ for the rational thinking scale.

Procedure

The study was conducted at the school psychology laboratory of the University of Florence (Italy). The measures were administered in two sessions. The first session was used to assess the three control variables (the session lasted for around 20 min and was collectively administered). In the second session, strange stories were administered, as well as texts and essays on the topic of flu vaccination. Before performing the reading and writing activity, the students received the full set of task instructions. Being so, they

were informed that they had to read six different documents regarding the topic of flu vaccination, with the final aim of writing an argumentative essay. In this argumentative essay, they had to answer the following question: would you recommend a person with heart problems to get a flu vaccination? To answer this question and report their personal opinion, they had to rely on the information provided by the texts. While writing the essay, they had access to the texts. While they were reading, they were asked to think-aloud, following the same procedure as in Tarchi (2021). This second session lasted for around 1 h and was individually administered.

Data analysis

To investigate the first research question (are ToM's performances associated across tasks), we first proceeded to calculate the MST indicator related to the reading phase (MST_R) and the MST indicator related to the writing phase (MST_W). The calculation of these two indicators was performed by identifying and adding the frequencies obtained by the participants in each mental state category, in order to generate an overall score. In addition, the total number of words used during the think-aloud protocol was counted, in order to create ratios and standardize participants' performances and check for the potentially confounding effect of speech production. Then, the relationship between ToM and MST_R and MST_W was explored through correlational and mediation analysis.

To investigate the second research question (is there an association between ToM, MST_{R-W}, and multiple-document comprehension), we conducted hierarchical multiple regression analyses, prior topic knowledge, prior topic beliefs, and need for cognition included as control variables.

To investigate the third and final research question (are ToM and multiple-document comprehension directly or indirectly associated), we computed correlation and mediation analysis.

Results

Descriptive statistics and correlations

The basic descriptive statistics were carried out. The analysis of skewness and kurtosis indices suggested that all variables were normally distributed, except for prior topic knowledge, argumentative quality, essay words, and ToM. Missing data ranged between 6 and 16%. Five subjects were identified as outliers and removed from the sample for the analysis.

Table 2 presents descriptive statistics and Table 3 correlations among the variables (Spearman correlation coefficient was used, taking into account the nonnormal distribution of some of the variables).

Participants showed a positive attitude towards flu vaccination considering their previous topic beliefs. However, they showed a low level of topic knowledge. In relation to the quality of the argumentative essays, it should be noted that the students wrote texts of very low quality and of very short length.

Table 2 Mean (standard deviations) and range of variables investigated

	M (SD)	Missing[*]	Range	Skewness	Kurtosis
Session I: covariates					
Prior topic beliefs	46.3 (8.18)	11%	21–60	−0.57	0.17
Prior topic knowledge	4.25 (1.47)	16%	1–8	0.48[a]	0.23[a]
Need for cognition	63.23 (8.12)	11%	48–80	0.15	−0.84
Session II: ToM and MD task					
Argumentative quality	1.31 (0.31)	6%	1–3	0.24[a]	−1.59[a]
Essay words	99.79 (63.44)	6%	17–310	1.1[a]	1[a]
ToM	11.63 (2.39)	8%	5–16	−0.64[a]	−0.27[a]
MST_R	3.53 (1.08)	2%	1.32–6.18	−0.07	−0.26
MST_W	1.23 (0.6)	6%	0–3.21	0.6	0.71

[a]Skewness or kurtosis outside acceptable range

[*]For session I, missing data represent the percentages of participants who did not complete the tasks out of the contacted people; for session II, missing data represent the drop rate from session I; in session II, some data were excluded because they were incomplete

Table 3 Correlations (Spearman's rho, r_s) between the examined variables

		1	2	3	4	5	6	7	8
1	Prior topic beliefs	–							
2	Prior topic knowledge	−0.08	–						
3	Need for cognition	0.29**	0.08	–					
4	AQ	0.06	0.11	0.19	–				
5	ToM	0.06	0.01	0.06	0.09	–			
6	MST_R	0.09	0.04	0.13	0.31**	0.34**	–		
7	MST_W	0.08	−0.1	0.19	0.59**	0.1	0.51**	–	
8	Essay words	−0.005	−0.11	0.19	0.71**	0.18	0.34**	0.71**	–

AQ argumentative quality, *ToM* theory of mind, MST_R mental state talk during reading, MST_W mental state talk in writing

*$p < 0.05$

**$p < 0.01$

Relation between ToM, MSTR, and MSTW

MST_R and MST_W were positively associated with a large effect size (r_s (81) = 0.51, $p < 0.01$). Also, we found an association between ToM and MST_R with a medium effect size (r_s (79) = 0.34, $p < 0.01$). However, ToM and MST_W were not correlated (r_s (77) = 0.1, $p = 0.38$). In view of these results, a mediation analysis was computed to examine a possible indirect effect of ToM (predictor variable [X]) on MST_W (criterion variable [Y]) through MST_R (proposed mediator [M]). The indirect effect was tested using PROCESS (model 4), a conditional process modeling program that tests direct and indirect effects using an ordinary least squares-based (OLS) path analytical framework (Preacher & Hayes, 2004). Bias-corrected bootstrapping ($k = 5000$) was used to

generate a 95% confidence interval (CI) to test the significance of the indirect effect (Preacher & Hayes, 2004). See Fig. 1 for a depiction of the mediation analysis.

Results indicated that ToM, assessed through strange stories, was significantly associated with MST during the reading of multiple documents (path a; $t = 3.37$, $p < 0.05$; $R^2 = 0.13$). In addition, higher levels of MST during the reading process were significantly associated with higher MST showed in the argumentative essays (path b; $t = 3.7$, $p < 0.05$; $R^2 = 0.18$). However, a direct effect of ToM on MST_W was not found (path c'; $t = 0.28$, $p = 0.78$; $R^2 = 0.18$). In terms of the indirect effect, the level of students' ToM was significantly associated with the MST that they manifested in their essays, through the MST showed during the reading process (indirect effect = 0.15; SE = 0.05; 95% CI [0.06, 0.25].

Relation between ToM, MSTR, MSTW, and AQ

According to the results of the correlational analysis shown in Table 2, we found an association between MST_R and AQ with a medium effect size (r_s (81) = 0.31, $p < 0.01$), as well as between MST_W and AQ with a large effect size (r_s (82) = 0.59, $p < 0.01$). However, ToM and AQ were not correlated (r_s (77) = 0.09, $p = 0.4$). A mediation analysis was conducted to test the indirect effect of ToM (predictor variable [X]) on AQ (criterion variable [Y]) through MST_R. The indirect effect was explored using PROCESS (model 4). Bias-corrected bootstrapping ($k = 5000$) was used to generate a 95% confidence interval (CI) to test the significance of the indirect effect (Preacher & Hayes, 2004). See Fig. 2 for a depiction of the mediation analysis.

Results indicated that ToM, assessed through strange stories, was significantly associated with MST during the reading of multiple documents (path a; $t = 3.37$, $p < 0.05$; $R^2 = 0.13$). In addition, higher levels of MST during the reading process were significantly associated with the quality of the argumentative essays (path b; $t = 2.15$, $p < 0.05$; $R^2 = 0.08$). However, a direct effect of ToM on AQ was not found (path c'; $t = 0.49$, $p = 0.62$; $R^2 = 0.08$). In terms of the indirect effect, the level of students' ToM was significantly associated with the argumentative quality of the essays, through the MST showed during the reading process (indirect effect = 0.09; SE = 0.05; 95% CI [0.006, 0.19].

On the other hand, the results of the correlational analysis revealed an association between MST_R and MST_W with a large effect size (r_s (81) = 0.51, $p < 0.01$), between MST_W and AQ (r_s (82) = 0.59, $p < 0.01$) with a large effect size, and between MST_R and AQ with a medium effect size (r_s (81) = 0.31, $p < 0.01$). Thus, we decided to conduct a mediation analysis in order to test the indirect effect of MST_R (predictor variable [X]) on AQ (criterion variable [Y]) through MST_W. The indirect effect was explored using PROCESS (model 4). Bias-corrected bootstrapping ($k = 5000$) was used to generate a 95% confidence interval

Fig. 1 Diagram of simple mediation analysis of ToM on MST_W trough MST_R. Note: dotted lines indicate nonsignificant relations. *$p < 0.05$

a = .36 *   b = .41*

MST_R

ToM

MST_W

c'= .03
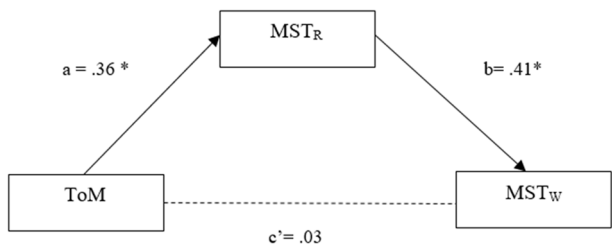
Fig. 2 Diagram of simple mediation analysis of ToM on AQ trough MST_R. Note: dotted lines indicate nonsignificant relations. *$p < 0.05$
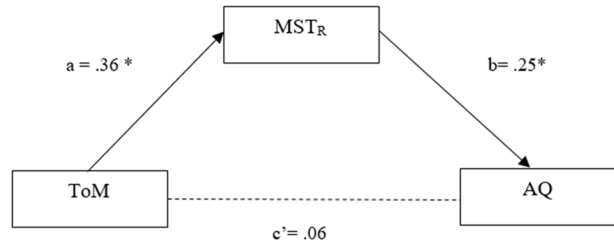
(CI) to test the significance of the indirect effect (Preacher & Hayes, 2004). See Fig. 3 for a depiction of the mediation analysis.

Results showed that MST_R was significantly associated with MST_W (path a; $t = 4.65$, $p < 0.05$; $R^2 = 0.21$). In addition, higher levels of MST_W were significantly associated with the quality of the argumentative essays (path b; $t = 4.93$, $p < 0.05$; $R^2 = 0.3$). However, a direct effect of MST_R on AQ was not found (path c'; $t = 0.55$, $p = 0.58$; $R^2 = 0.3$). In terms of the indirect effect, MST_R was significantly associated with the argumentative quality of the essays, through the MST_W (indirect effect $= 0.24$; SE $= 0.06$; 95% CI [0.13, 0.36]).

Hierarchical multiple regression analyses

To explore the patterns of association between ToM, MST_{R-W}, and AQ, we also conducted a hierarchical regression analysis, with AQ as the dependent variable. Control variables were considered at the first step, namely, prior topic beliefs, prior topic knowledge, and need for cognition. ToM, MST_R, and MST_W were introduced at the second step.

The average VIF (variance inflation factor) was 1.11, and tolerance ranged from 0.71 to 0.92; thus, we concluded that the regression model was not biased by multicollinearity. Results are reported in Table 4.

At Step 1, the overall equation was not significant, $F (3, 64) = 1.06$, $p = 0.37$, and explained the 4.7% of the variance in AQ, Cohen's $f^2 = 0.05$, which is a small effect size. None of the control variables introduced in this model was significantly associated with the argumentative quality of the essays. More interesting, the final model was significant, $F (6, 61) = 4.36$, $p < 0.01$, and explained the 30% of the variance in AQ, Cohen's $f^2 = 0.43$, which represents a large effect size. Notably, the introduction of MST_R, MST_W, and ToM into the model significantly increased the amount of variance explained. $\Delta R^2 = 0.25$, $F (3, 61) = 7.35$, $p < 0.01$. However, beta coefficients revealed that MST_W was the only variable that was associated with AQ.

To investigate which category of mental state talk was more strongly associated with argumentative quality, we conducted a series of nonparametric correlation (Spearman's

Fig. 3 Diagram of simple mediation analysis of MST_R on AQ trough MST_W. Note: dotted lines indicate nonsignificant relations. *$p < 0.05$
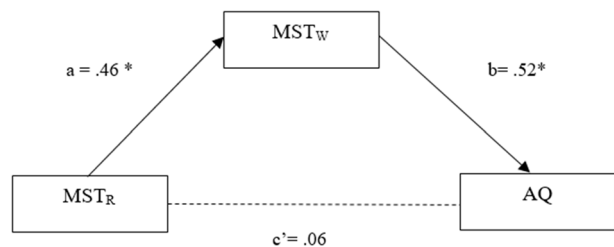
Table 4 Results of the regression analysis with AQ as dependent variable

		B	SE	β	95% CI	R^2
Step 1	Prior topic beliefs	0.001	0.005	0.027	[−0.009, 0.011]	0.047
	Prior topic knowledge	0.031	0.027	0.145	[−0.023, 0.086]	
	Need for cognition	0.006	0.005	0.151	[−0.004, 0.015]	
Step 2	MST_W	0.246	0.06	0.493**	[0.126, 0.367]	0.30**
	MST_R	0.007	0.041	0.023	[−0.075, 0.089]	
	ToM	0.006	0.015	0.05	[−0.024, 0.037]	

$*p < 0.05$

$**p < 0.01$

rho). Among mental state talk categories when reading, only cognitive states (rho = 0.32, $p < 0.01$) and perceptual states (rho = 0.23, $p < 0.05$) were significantly associated with argumentative quality. Among mental state talk categories when writing, only cognitive states (rho = 0.47, $p < 0.01$) and moral states (rho = 0.44, $p < 0.01$) were significantly associated with argumentative quality.

Discussion

The present study investigated the relationship between university students' ability to comprehend multiple documents, measured through an argumentative essay task, and their ToM. We distinguished between their ToM (assessed by the strange stories task) and the spontaneous use of ToM during the reading and writing processes (measured by the mental state talk index). In doing so, we took into account some control variables (prior topic beliefs, prior topic knowledge, and need for cognition).

The study showed that students' ToM, assessed by the strange stories task, was directly associated with their mental state talk (MST_R) during the reading of the multiple documents and indirectly associated with their mental state talk during the writing of their essays (MST_W). Thus, the first contribution of this study is to show the interest to assess TOM using a think-aloud task through the identification of mental states. Finally, the results of the current study revealed that students' ToM was significantly associated with the argumentative quality of their essays, through the MST showed during the reading process, and that the relation between their MST_R and the argumentative quality was mediated through MST_W (Q3). These findings and their theoretical and educational implications are discussed below.

Relation between ToM, MSTR, and MSTW

The results from the analyses partially corroborated the assumptions presented in our first hypothesis. We expected a positive relation between ToM, MST_R, and MST_W. However, the results revealed a significant positive relation between ToM and MST_R, but not between ToM and MST_W. The association between MST_R and MST_W was positive and significant and the mediation analysis showed an indirect effect of ToM on MST_W, through MST_R. The absence of a direct relation between ToM and MST_W may be explained according to

the following reasons. The association between ToM and MST_W may disappear because during the writing process students need to be focused on a demanding activity such as text production. Students may focus more on the content and concepts there were included in the documents while they are writing and reflect less on the mental states that are associated with them.

Another explanation may be that ToM, MST_R, and MST_W are distinct constructs and that having a high level of ToM, as what emerged from a high score on an experimental ToM task, does not necessarily imply a propensity to employ that skill in certain activities. Mediation analyses allow us to support this assumption. According to our results, the relationship between ToM and MST_W is mediated by MST_R. We found a strong relationship between ToM and MST_R, meaning that certain tasks, such as reading multiple documents, elicit the use of theory of mind more directly than argumentative essay writing. This result would be aligned with the idea that manifestation of theory of mind is dependent on the type of activity demanded and that it could be a domain-specific rather than a domain-general competence. In fact, although self-other control is assumed to be a commonly assessed process in all ToM tasks, egocentrism and other self-interference effects are instead a broad spectrum of phenomena that may manifest themselves in different ways depending on the task and situations (Qureshi et al., 2020).

Relation between ToM, MSTR, MSTW, and AQ

The results from the analyses partially corroborated the assumptions presented in the second hypothesis. We hypothesized an association between ToM, MST_{R-W}, and multiple-document comprehension over all control variables, i.e., prior knowledge, prior beliefs, and need for cognition. However, the hierarchical regression analysis revealed that these individual variables were not related to multiple-document comprehension in the current study (see Step 1 in Table 4). In fact, the model that included them only explained the 4.7% of the variance in the argumentative quality of the essays. These data are not aligned with the results of previous studies that have shown the role of prior topic beliefs, prior topic knowledge, and need for cognition on multiple-document comprehension (Bråten et al., 2014a, b; Dai &Wang, 2007; Kardash & Scholes, 1996; Richter & Maier, 2017; Strømsø et al., 2010). The lack of association between control variables and multiple-document comprehension in our research could be due to the characteristics of the essays that the students produced. The students wrote texts of very short length and low quality (see Table 2). A tentative interpretation would be that the features of the products used to infer the level of comprehension of multiple documents may have interfered with the relationship between the control variables and the outcome. Another possible explanation could be that the instruments used to evaluate these individual variables are less effective on the specific topic discussed in the texts.

On the other hand, the hierarchical regression analysis also illustrated how introducing the variables MST_R, MST_W, and ToM into the model significantly increased the amount of variance explained of the outcome. Further exploration of the beta coefficients revealed that the MST_W variable was the only variable significantly associated with the argumentative quality of the essays (see Table 4). This data suggests that not all manifestations of ToM would be associated with the level of performance in comprehending multiple documents. Previous research has shown the contribution of ToM in multiple- and single-text comprehension (see Atkinson et al., 2017; Boerma et al., 2017; Florit et al., 2020; Kim, 2017). However, in all these studies, ToM was assessed by false belief tasks and, moreover, they were conducted with

a child population. Our data illustrate that, in university students, multiple-document comprehension is not associated with ToM measures obtained from experimental tasks, but with more ecological ToM measures such as mental state talk. In addition, our results revealed that not all measures of MST would be related to the comprehension of multiple documents, but only those which occur during the activity that will be used to assess this competency, i.e., during the writing of the essays from which we infer comprehension of multiple sources. Therefore, our results allow us to further discriminate the validity of MST measures when they are taken during the performance of different cognitive processes (e.g., reading and writing processes). In the present study, students who showed higher rates of MST in the writing process also produced higher quality essays, which means that they were more proficient in understanding different sources. One possible interpretation would be that the use of nouns/verbs/adjectives during writing that refer to mental states (feelings, desires, thoughts…) may contribute to the understanding of the sources and the integration of the different perspectives presented by the authors. Thus, the greater the capacity to understand and produce mental states during the writing process, the greater the capacity to represent and integrate text perspectives (Florit et al., 2020). Of notice, among mental state talk, it appears that cognitive states are particularly relevant for argumentative quality, as this category was significantly associated with the outcome measure both when measured in the reading and writing stages. Moral states when writing also are quite strongly associated with argumentative quality. Overall, it seems that reflecting on the cognitive and moral states behind people's actions or ideas helps to achieve a two-sided reflection on a controversial topic.

Regarding the question on the association between ToM, MST_R, MST_W, and argumentative quality (AQ), we expected a positive relationship between these four variables. The results of the mediation analysis allow us to partially confirm this hypothesis. Of particular interest are the indirect effects found. We observed (1) an indirect effect of ToM on AQ, through MST_R, and (2) an indirect effect of MST_R on AQ, through MST_W. These results are aligned with the idea that ToM and MST are different constructs and that manifesting a high level of ToM in an experimental task may not imply the propensity to use it in a specific activity. Furthermore, these results highlight the complex interplay of ToM capabilities.

Surprisingly, control variables (prior beliefs, prior knowledge, or need for cognition) were not associated with neither argumentative quality nor with essay length. This result is not in line with prior findings (Bråten et al., 2014a, b; Richter & Maier, 2017; Tarchi & Villalón, 2021). On a speculative level, prior knowledge and prior beliefs may have not been associated with multiple-text comprehension because of the specific topic chosen. Individuals displayed overall low levels of prior knowledge and moderately positive attitudes towards the topic of vaccination. Probably, topics with more skewed prior beliefs or with more variance in prior knowledge (*e.g.,* climate change) would require a higher involvement of these individual differences in constructing a mental representation of the texts' content. What concerns need for cognition, either the texts were considered easy to read by the subjects or the think-aloud helped to produce an overall higher engagement. Unfortunately, the present research design does not allow to provide an evidence-based interpretation of these results and further investigation is warranted.

Limitations

The present study has some limitations that should be pointed out. Firstly, we only evaluated some individual variables of the participants, specifically prior topic beliefs, prior topic knowledge, and need for cognition. Research has revealed the association

of other intrasubject factors and multiple-document comprehension that we have not considered in the present study. Importantly, initial reading comprehension competence was not assessed in this study. While it is true that none of the participants in the study had a diagnosis of learning difficulties related to reading and writing (e.g., dyslexia), it is possible that some students may have had suboptimal reading competence levels. Considering that low reading skills are a risk factor for students in higher education, an initial assessment in this regard should be carried out in future studies (Rasinski et al., 2017). Since the strange stories task and the multiple-document comprehension task both rely on reading comprehension skills, there may be an overlap in processes involved, causing a risk of possible overdetermination through factors that were not controlled in the experiment. Of notice, the two tasks differ in complexity at several levels: the multiple-document comprehension task is longer and more complex and requires more prior knowledge than the strange stories task does. Indeed, we found only modest correlations between the outcomes measures derived from these two tasks. Nevertheless, performances in these two tasks may be supported by several reading-related processes (*e.g*., vocabulary, reading fluency, working memory, and comprehension monitoring). Unfortunately, in the present research design, we did not include a reading comprehension measure and more studies are warranted to determine if this variable may moderate the results of this study.

Secondly, the essays elaborated by the students had, in general, a low argumentative quality. This may depend on a series of factors, including low writing competence, lack of exposure to the argumentative genre, low perceived task relevance, task model, or that the question guiding the argumentative essay has not elicited the type of writing expected. The choice of topic may have had an influence too, as the use of other controversial topics could have revealed another pattern of results taking into account the study variables. The authors' measure of essay quality highly correlated with essay length, suggesting that the effects may be related to participants' varying levels of motivation and engagement in the task or writing-related factors, such as verbosity. This fact makes it difficult to interpret the results of the study and the association between students' ToM and comprehension of multiple documents. Therefore, it would be desirable to replicate the study in a sample of students showing greater variability in performance on the writing task by acting on one of the aforementioned factors. Moreover, we did not evaluate the participants' perceived argumentative position of the texts. The neutral texts may have been perceived as either pro or con texts, generating an imbalance between the positions of the sources.

Finally, what concerns the mental state talk measures, what participants said during reading could have primed the mental state talk expressed in their writing. In addition, the same content-analysis framework was used to measure references to mental states in the think-aloud protocols and in the essays. However, reading and writing are partially overlapping processes (the correlations may thus be overdetermined), and in addition, while thinking aloud is based on spontaneous talk, writing should adhere to conventional structures. The results confirmed that MST in reading and writing were highly correlated, but also that ToM ability was directly associated with ToM use when reading but only indirectly associated with ToM use when writing. This pattern of results seems to suggest that the two mental state talk measures are partially overlapping, but they are also differently influenced by ToM ability. Since specific control measures to determine the exact contribution of ToM in reading and writing tasks are missing, we interpreted the results with caution.

Conclusions

Despite the limitations, the present study contributes to the literature on multiple-document comprehension in several ways. First, it relates this competence to a metacognitive variable of the subjects such as ToM, which has received little attention compared to other individual variables in performance on multiple-document comprehension tasks. Second, the present study focuses on adults, in contrast to previous research that analyzed the impact of ToM on single and multiple-text comprehension in a child population. Also, this study investigates the relationship between multiple-document comprehension performance, measured through an argumentative essay task, and ToM, measured through (1) the strange stories task and (2) the mental state talk during two different processes (reading and writing). These two ways of assessing ToM have allowed us to make a distinction between the level of ToM, according to the results in experimental tasks, and the propensity to use this skill in contextual activities. The results of the present study shed light on the importance of employing more ecological ToM measures, such as mental state talk, when we seek to analyze the impact of ToM on multiple-document comprehension. Likewise, our results reveal that mental state talk, depending on the cognitive activity in which it is applied (reading or writing), is associated with performance in multiple-document comprehension tasks to a greater or lesser extent. Moreover, the present study reveals that the think-aloud protocol is a valid method to access mental state talk during the development of different learning activities.

The involvement of ToM in multiple-document comprehension through both ToM ability and ToM use suggests, from an educational perspective, that the former represents an important individual difference that may influence the way readers approach the task, whereas the latter one suggests the possibility to support multiple-document comprehension by scaffolding mental state talk. There are at least two important implications.

Firstly, although most of the intervention studies have been conducted with preschoolers, there is mounting evidence that ToM can be trained in older ages (see Lecce et al., 2014, 2015). Thus, we could expect that a ToM training could be also beneficial on students' ability to integrate information as represented by multiple perspectives discussed in texts. This represents also one more reason to dedicate attention on ToM development even on older ages; thus, educators should be provided with ecologically valid measures of ToM (such as mental state talk) to monitor its development throughout the life cycle.

Secondly, we can hypothesize that making readers/writers more aware of mental state talk when reflecting while reading or writing after reading may increase their multiple-document comprehension. For example, an educator could support the comprehension of multiple texts and the generation of an integrated representation of their content if he/she introduces during the reading task a list of prompts to direct the readers' attention to the mental states associated to each perspective that is discussed in the texts. This being the case, students could be guided in the process of reflecting on mental states, especially cognitive and moral ones, of the people defending the different perspectives on the controversy addressed in the texts.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

Ethics approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional research committee of the University of Florence.

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

Alexander, P. A. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist, 47*, 259–280. https://doi.org/10.1080/00461520.2012.722511

Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences, 30*, 64–76. https://doi.org/10.1016/j.lindif.2013.01.007

Annisette, L. E., & Lafreniere, K. D. (2017). Social media, texting, and personality: A test of the shallowing hypothesis. *Personality and Individual Differences, 115*, 154–158. https://doi.org/10.1016/j.paid.2016.02.043

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology, 63*, 1201–1217. https://doi.org/10.1080/17470210903281582

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology, 164*, 225–238. https://doi.org/10.1016/j.jecp.2017.04.007

Barnes, J. L., Lombardo, M. V., Wheelwright, S., & Baron-Cohen, S. (2009). Moral dilemmas film task: A study of spontaneous narratives by individuals with autism spectrum conditions. *Autism Research, 2*, 148–156. https://doi.org/10.1002/aur.79

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 1*, 221–235. https://doi.org/10.1080/13546789508256909

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241–251. https://doi.org/10.1111/1469-7610.00715

Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learning and Instruction, 36*, 86–103. https://doi.org/10.1016/j.learninstruc.2014.12.003

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 99–116). Routledge.

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation. *Instructional Science, 43*, 737–766. https://doi.org/10.1007/s11251-015-9359-4

Bittermann, A., McNamara, D., Simonsmeier, B. A., & Schneider, M. (2023). The landscape of research on prior knowledge and learning: A bibliometric analysis. *Educational Psychology Review, 35.* https://doi.org/10.1007/s10648-023-09775-9

Boerma, I. E., Mol, S. E., & Jolles, J. (2017). The role of home literacy environment, mentalizing, expressive verbal ability, and print exposure in third and fourth graders' reading comprehension. *Scientific Studies of Reading, 21*, 179–193. https://doi.org/10.1080/10888438.2016.1277727

Bowles, M. A., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology. Expanding the scope of investigation. *Studies in Second Language Acquisition, 27*, 415–440. https://doi.org/10.1017/S0272263105050187

Bråten, I., & Strømsø, H. I. (2010). When law students read multiple documents about global warming: Examining the role of topic-specific beliefs about the nature of knowledge and knowing. *Instructional Science, 38*, 635–657. https://doi.org/10.1007/s11251-008-9091-4

Bråten, I., & Strømsø, H. I. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning, 6*, 111–130. https://doi.org/10.1007/s11409-011-9075-7

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*, 48–70. https://doi.org/10.1080/00461520.2011.538647

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2014a). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology, 84*, 58–85. https://doi.org/10.1111/bjep.12005

Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømsø, H. I. (2014b). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction, 30*, 9–24. https://doi.org/10.1016/j.learninstruc.2013.11.002

Bretherton, I., & Beegley, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology, 18*, 906–921. https://doi.org/10.1037/0012-1649.18.6.906

Britt, M. A., & Rouet, J. F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276–314). Cambridge University Press.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychological Bullettin & Review, 16*, 893–900. https://doi.org/10.3758/PBR.16.5.893

Cavallini, E., Lecce, S., Bottiroli, S., Palladino, P., & Pagnin, A. (2013). Beyond false belief: Theory of mind in young, young-old, and old-old adults. *The International Journal of Aging and Human Development, 76*, 181–198. https://doi.org/10.2190/AG.76.3.a

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*, 141–167. https://doi.org/10.1080/00461520.2011.587722

Dai, D. Y., & Wang, X. (2007). The role of need for cognition and reader beliefs in text comprehension and interest development. *Contemporary Educational Psychology, 32*, 332–347. https://doi.org/10.1016/j.cedpsych.2006.05.002

Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology, 149*, 23–40. https://doi.org/10.1016/j.jecp.2015.07.011

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52*, 758–771. https://doi.org/10.1037/dev0000105

Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 205–242). International Reading Association, Inc. https://doi.org/10.1598/0872071774.10

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13*, 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Dyoniziak, Y., Potocki, A., & Rouet, J. F. (2023). Role of advanced theory of mind in teenagers' evaluation of source information. *Discourse Processes, 60*, 363–377. https://doi.org/10.1080/0163853x.2023.2197691

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36*, 623–636. https://doi.org/10.1007/s10803-006-0107-0

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology, 71*, 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction, 22*, 103–120. https://doi.org/10.1016/j.learninstruc.2011.08.002

Florit, E., De Carli, P., Giunti, G., & Mason, L. (2020). Advanced theory of mind uniquely contributes to children's multiple-text comprehension. *Journal of Experimental Child Psychology, 189*, 104708. https://doi.org/10.1016/j.jecp.2019.104708

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*, 157–173. https://doi.org/10.1016/j.cedpsych.2009.11.002

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist, 43*, 142–160. https://doi.org/10.1080/00461520802178458

Hagen, Å. M., Braasch, J. L. G., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies, and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading, 37*, 141–157. https://doi.org/10.1111/j.1467-9817.2012.01536.x

Hao, J., Su, Y., & Chan, R. C. K. (2010). Do deaf adults with limited language have advanced theory of mind? *Research in Developmental Disabilities, 31*, 1491–1501. https://doi.org/10.1016/j.ridd.2010.06.008

Happè, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Development, 66*, 843–855. https://doi.org/10.1007/BF02172093

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*, 766–794. https://doi.org/10.1037/0033-295X.106.4.766

Hofer, B. K. (2002). Personal epistemology as a psychological and educational construct: An introduction. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 3–14). Erlbaum.

Hughes, C., & Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives, 9*, 149–153. https://doi.org/10.1111/cdep.12124

Hughes, C., Marks, A., Ensor, R., & Lecce, S. (2010). A longitudinal study of conflict and inner state talk in children's conversations with mothers and younger siblings. *Social Development, 19*, 822–837. https://doi.org/10.1111/j.1467-9507.2009.00561.x

Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology, 88*, 260–271. https://doi.org/10.1037/0022-0663.88.2.260

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41. https://doi.org/10.1016/s0010-0277(03)00064-7

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*, 377–380. https://doi.org/10.1126/science.1239918

Kim, Y. G. (2017). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141*, 101–120. https://doi.org/10.1016/j.jecp.2015.08.003

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95e*, 163–182. https://doi.org/10.1037/0033-295x.95.2.163

LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., Donovan, S., & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness, 9*, 201–222. https://doi.org/10.1080/19345747.2015.1116035

Lecce, S., Caputi, M., & Hughes, C. (2011). Does sensitivity to criticism mediate the relationship between theory of mind and academic competence? *Journal of Experimental Child Psychology, 110*, 313–331. https://doi.org/10.1016/j.jecp.2011.04.011

Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: A training program. *Journal of Experimental Child Psychology, 126*, 52–67. https://doi.org/10.1016/j.jecp.2014.03.002

Lecce, S., Bottiroli, S., Bianco, F., Rosi, A., & Cavallini, E. (2015). Training older adults on theory of mind (ToM): Transfer on metamemory. *Archives of Gerontology and Geriatrics, 60*, 217–226. https://doi.org/10.1016/j.archger.2014.10.001

Lecce, S., Ceccato, I., & Cavallini, E. (2019). Theory of mind, mental state talk and social relationships in aging: The case of friendship. *Aging & Mental Health, 23*, 1105–1112. https://doi.org/10.1080/13607863.2018.1479832

List, A., & Alexander, P. A. (2019). Toward an integrated framework of multiple text use. *Educational Psychologist, 54*, 20–39. https://doi.org/10.1080/00461520.2018.1505514

Lucisano, P., & Piemontese, M. E. (1988). GULPEASE: Una formula per la predizione della difficoltà dei testi in lingua italiana (en. tr. GULPEASE: A formula to predict the difficulty of texts in Italian). *Scuola e Città, 3*, 110–124.

Maier, J., & Richter, T. (2013). Text belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction, 31*, 151–175. https://doi.org/10.1080/07370008.2013.769997

Maylor, E. A., Moulson, J. M., Muncer, A. M., & Taylor, L. A. (2010). Does performance on theory of mind tasks decline in old age? *British Journal of Pyschology, 93*, 465–485. https://doi.org/10.1348/000712602761381358

McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist, 56*, 196–214. https://doi.org/10.1080/00461520.2021.1872379

Miller, C., & Bartlett, J. (2012). 'Digital fluency': towards young people's critical use of the internet. *Journal of Information Literacy, 6*, 35–55. https://doi.org/10.11645/6.2.1714

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. Van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representation during reading* (pp. 99–122). Erlbaum.

Preacher, K., & Hayes, J. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717–731. https://doi.org/10.3758/BF03206553

Qureshi, A. W., Monk, R. L., Samson, D., & Apperly, I. A. (2020). Does interference between self and other perspectives in theory of mind tasks reflect a common underlying process? Evidence from individual differences in theory of mind and inhibitory control. *Psychonomic Bulletin & Review, 27*, 178–190. https://doi.org/10.3758/s13423-019-01656-z

Rasinski, T. V., Chang, S. C., Edmondson, E., Nageldinger, J., Nigh, J., Remark, L., Kenney, K. S., Walsh-Moorman, E., Yildirim, K., Nichols, W. D., Paige, D. D., & Rupley, W. H. (2017). Reading fluency and college readiness. *Journal of Adolescent & Adult Literacy, 60*, 453–460. https://doi.org/10.1002/jaal.559

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist, 52*, 1–19. https://doi.org/10.1080/00461520.2017.1322968

Rosso, A. M., Viterbori, P., & Scopesi, A. M. (2015). Are maternal reflective functioning and attachment security associated with preadolescent mentalization? *Frontiers in Psychology, 6*, 1–12. https://doi.org/10.3389/fpsyg.2015.01134

Rouet, J. F. (2006). *The skills of document use: From text comprehension to Web-based learning.* Lawrence Erlbaum Associates.

Rouet, J. F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist, 52*, 200–215. https://doi.org/10.1080/00461520.2017.1329015

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning, 2*, 191–210. https://doi.org/10.1007/s11412-007-9015-3

Strømsø, H. I., & Bråten, I. (2009). Beliefs about knowledge and knowing and multiple-text comprehension among upper secondary students. *Educational Psychology, 29*, 425–445. https://doi.org/10.1080/01443410903046864

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20*, 192–204. https://doi.org/10.1016/j.learninstruc.2009.02.001

Tarchi, C. (2021). Effects of think-aloud on students' multiple-documents comprehension. *Applied Cognitive Psychology, 35*, 526–537. https://doi.org/10.1002/acp.3782

Tarchi, C., & Villalón, R. (2021). The influence of thinking dispositions on integration and recall of multiple texts. *British Journal of Educational Psychology, 91*, 1498–1516. https://doi.org/10.1111/bjep.12432

Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers & Education, 160*, 104034. https://doi.org/10.1016/j.compedu.2020.104034

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of mind development in adolescence and early adulthood: The growing complexity of recursive thinking ability. *Europe's Journal of Psychology, 11*, 112–124. https://doi.org/10.5964/ejop.v11i1.829

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Weston-Sementelli, J. L., Allen, L. K., & McNamara, D. S. (2018). Comprehension and writing strategy training improves performance on content-specific source-based writing tasks. *International Journal of Artificial Intelligence in Education, 28*, 106–137. https://doi.org/10.1007/s40593-016-0127-7

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

World Medical Association. (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*, 2191–2194. https://doi.org/10.1001/jama.2013.281053

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18*, 600–606. https://doi.org/10.1111/j.1467-9280.2007.01946.x

Christian Tarchi. Department of Education, Languages, Intercultures, Literatures, and Psychology, University of Florence, via San Salvi, 12–50125 Florence, Italy. E-mail: christian.tarchi@unifi.it

Lidia Casado-Ledesma. Department of Education, Languages, Intercultures, Literatures, and Psychology, University of Florence, via San Salvi, 12–50125 Florence, Italy. E-mail: lidia.casadoledesma@unifi.it

Giulia Sanna. University of Florence, via San Salvi, 12–50125 Florence, Italy. E-mail: giulia.sanna5@stud.unifi.it

Margherita Conti. University of Florence, via San Salvi, 12–50125 Florence, Italy. E-mail: margherita.conti2@stud.unifi.it

Current themes of research:

Reading comprehension and multiple-text comprehension.
Argumentative writing and multiple-text comprehension.
Multiple-text comprehension.

Most relevant publications in the field of Psychology of Education:

Mason, L., Tarchi, C., Ronconi, A., Manzione, L., Latini, N., & Bråten, I. (2022). Do medium and context matter when learning from multiple complementary digital texts and videos? *Instructional Science, 50*, 653–679. https://doi.org/10.1007/s11251-022-09591-8

Tarchi, C., Villalón, R., & Lamanda, E. (2022). Investigating the effect of actively-openminded thinking on source-based writing: A randomized control trial. *Thinking Skills and Creativity, 46,* 101128. https://doi.org/10.1016/j.tsc.2022.101128

Tarchi, C., & Mason, L. (2022). Learning across media in a second language. *European Journal of Psychology of Education.* https://doi.org/10.1007/s10212-022-00652-7

Casado-Ledesma, L., Cuevas, I., Van den Bergh, H., Rijlaarsdam, G., Mateos, M., Granado- Peinado, M., & Martín, E. (2021). Teaching argumentative synthesis writing through deliberative dialogues: Instructional practices in secondary education. *Instructional Science, 49*, 515-559. https://doi.org/10.1007/s11251-021-09548-3

Casado-Ledesma, L., Cuevas, I., & Martín, E. (2021). Learning science through argumentative synthesis writing and deliberative dialogues: A comprehensive and effective methodology in secondary education. *Reading and Writing.* https://doi.org/10.1007/s11145-021-10191-0

Granado-Peinado, M., Cuevas, I., Olmos, R., Martín, E., Casado-Ledesma, L., & Mateos, M. (2022). Collaborative writing of argumentative syntheses by low-performing undergraduate writers: explicit instruction and practice. *Reading and Writing.* https://doi.org/10.1007/s11145-022-10318-x