The relationship between theory of mind and emotion regulation in the preschool years. The role of language and gender

Daniela Bulgarelli°*, Maria Nives Sala°**, Beatrice Ragaglia°***, Paola Molina§

°Department of Psychology,
Università degli Studi di Torino,
Torino, Italia;
*e-mail: daniela.bulgarelli@unito.it.
** e-mail: nives.sala@gmail.com.
*** e-mail: beatrice.ragaglia@unito.it.
§Interuniversity Department of Regional and Urban Studies and Planning,
Università degli Studi di Torino,
Torino, Italia;
e-mail: paola.molina@unito.it.

Ricevuto: 08.05.2024. - **Accettato**: 15.07.2024

Pubblicato online: 23.12.2024

Abstract

A few studies investigated the relationship between Theory of Mind (ToM) and Emotion Regulation (ER) in preschoolers, mainly using False Belief tasks for assessing ToM, and disappointment procedures for measuring ER. Mixed findings were reported, as some researchers did not find any association and others did. The current study aims at testing the positive correlation between ToM and ER, using different measures, i.e., the comprehensive test ToM Storybooks, which allows for qualitative, quantitative, and total scores, and the narrative story stem procedure SIRE, which returns four scores: Behavioural Strategy, Cognitive Reappraisal, Social Support, and Attentional Deployment. Moreover, the study aims at testing the role of receptive vocabulary size (VS) and gender on the relationship between ToM and ER. Sixty-two Italian 3- to 6-year-old children participated. The three ToM scores, VS score, and ER Behavioural Strategy and Cognitive Reappraisal scores were associated with age. Gender differences were only found in ER Behavioural Strategy and Social Support scores. VS was significantly associated with the three ToM scores and with the ER Behavioural Strategy and Cognitive Reappraisal.

ISSNe 1972-5620, Doi: 10.3280/rip2024oa18947

Partial correlations controlling for age, gender, and VS showed that ToM Qualitative score was associated with ER Cognitive Reappraisal: more complex ToM abilities were associated with ER cognitive strategies.

Keywords: ToM Storybooks; Emotion Regulation Story Stems; Social Cognition; Receptive Vocabulary Size; Gender differences.

Introduction

The current study aims at analysing the relationships among Theory of Mind (ToM), Emotion Regulation (ER), language, and gender in typically developing preschoolers. In what follows, the literature about the associations among these competencies is presented. As discussed below, only five studies from 2007 and 2015 analysed together the four skills.

Theory of Mind, language, and gender

ToM is the ability to attribute mental states, such as desires, thoughts, intentions, and beliefs to self and others; to understand that others' mental states differ from one's own; and to predict one's own and others' behaviour (Wellman, 2018). ToM develops from early infancy: starting from 2 years of age, children talk about emotional experiences, and understand that desires cause behaviours and emotions; at about 3 years, they use beliefs to predict behaviours and, from 4 years onwards, they understand that mental states can differ from reality and be untrue, and children can use false beliefs to foretell the other people's behaviours (Blijd-Hoogewys et al., 2022; Wellman, 2018).

Since the early Eighties, False Belief tasks have been mainly used to assess ToM abilities in preschool age (Wellman, 2018). Several versions of the False Belief tasks have been developed, but all are characterized by asking the child to predict the behaviour of another person, based on the fact that this person has an erroneous belief in reality. The children are told a story or observe some puppets interacting with each other, and have to solve such problems: where Ann will look for a doll if she thinks it is in a basket, even though it is actually in a box? What will Maxi expect to find in a closed box of Smarties that instead contains pencils? The developmental trend of children's responses to the False Belief procedures does not depend on the type of task, type of questions, nature of the protagonist, etc. (Wellman, 2018). Researchers grew the awareness that ToM is not simply understanding False Beliefs and several instruments have been designed to better analyse the

construct. A systematic review reported that, from 1983 to 2019, 220 different measures have been used to assess ToM in children from 0 to 5 years of age (Beaudoin et al., 2020). These tasks and tools measure abilities that could be grouped into seven clusters of psychological aspects that contribute to the multifaceted construct of ToM: emotions, desires, intentions, percepts, knowledge, beliefs, and mentalistic understanding of non-literal communication. Each cluster included from four to eight sub-abilities. To obtain a more reliable and complex assessment, currently 18 comprehensive measures, composed of multiple ToM tasks, are available. The ToM Storybooks is one of them (Blijd-Hoogewys et al., 2008; Bulgarelli et al., 2015). In this tool, five ToM components are evaluated: psychology of emotions, desires, beliefs, perception leads to knowledge, and distinction between mental and physical phenomena. The components are presented via two or more similar tasks contextualised in different stories, thus allowing for more reliable and stable measurement than FB tasks alone. A detailed description of the test is reported in the Supplementary materials.

ToM and language are interdependent competences (Milligan et al., 2007): longitudinal studies showed that significant predictive correlations between language and ToM were greater in magnitude than those between ToM and language. The relationship between ToM and several linguistic and communication features – such as lexicon, syntax, and complementation, and being involved in the conversation – has been studied. Specifically, vocabulary size (VS) was found to partly explain preschoolers' performances in False Beliefs tasks. While language fosters ToM development by putting one's own mental states and that of others into words, developing ToM abilities permits to be included more often in communication exchanges and also to be more effective while interacting with others (for a literature review, see Belacchi, 2022; Bulgarelli, Henning, & Bertin, 2022; Farrar et al., 2017).

Concerning gender differences in preschooler's ToM performances, a few studies found that girls were advantaged in False Belief tasks compared to boys, while often such differences were not detected (see Blijd-Hoogewys et al. 2022, for a review). Yet Charman et al. (2002) discussed that most of the studies lacked statistical power to demonstrate small gender differences.

Emotion Regulation, language, and gender

ER refers to the processes used to modify and manage the way emotions are experienced and how they are expressed through behaviour. The three core features of ER are the activation of a regulatory goal; the engagement of regulatory processes, that range from explicit, conscious, effortful, and

controlled regulation to implicit, unconscious, effortless, and automatic regulation; and the modulation of the emotion trajectory (Gross, 2024). In 1998, Gross elaborated a process model that highlights five families of emotional regulation processes in adults: situation selection to avoid unpleasant circumstances; situation modification, which are behavioural strategies such as physically removing from view the stimulus generating the emotion; attentional deployment, as searching for distraction to move the attention away from the situation; cognitive change to modify the emotional significance of a situation (which also includes cognitive reappraisal); and modulation of the emotional response itself.

ER strategies that are connected to Gross' model have been also used to describe ER in childhood. Attentional deployment seems quite stable in the preschool years, the use of behavioural strategies to regulate emotions improves from 3 to 6 years of age, and cognitive reappraisal appears at 3 years, if adults instruct the children to do so, and is spontaneously used from 5 years onwards (Davis et al., 2010; Sala et al., 2014).

In childhood, ER is measured through different methodological approaches: adult-report and self-report questionnaires, task-based procedures which often are designed to cause frustration or disappointment in the child, and narrative story stem tools (for a brief review, see Sala et al., 2014). Specifically, story stem completion procedures elicit children's narratives, to identify their representations through the analysis of both content and narrative style. The Emotion Regulation Story Stems (*Storie Interrotte per la Regolazione Emotiva*-SIRE; Sala et al., 2014) is one of these instruments. It allows to analyse and categorise the preschoolers' narratives about the regulation of five negative emotions, namely fear, anger, sadness, shame, and guilt. A detailed description of the test is reported in the Supplementary materials.

Language and ER are related: Eisenberg and colleagues (2005) proposed that linguistic skills promote ER, as children with higher communication abilities have more opportunities to learn about mental states, as just discussed. This model has been confirmed in subsequent studies: a small correlation between ER and receptive VS is observed in preschoolers (Cohen & Mendez, 2009; Reylly & Downer, 2019) and better language skills in toddlerhood predict the use of strategies to regulate emotion at 4 years of age (Roben et al., 2013). Cole et al. (2009) also found that expressive language predicts the generation of appropriate and effective strategies to regulate negative emotions (i.e., anger) in 3- and 4-year-olds. Ren et al. (2009) confirmed that both measures of expressive and receptive language are correlated to positive emotion regulation and emotion dysregulation in 3- to 5-year-olds.

Chaplin's bio-psycho-social contextual model is meant to explain the role

of gender on emotion expression: «Small biologically related gender differences in behaviours are present at birth, such as boys' greater reactivity and energy level and girls' greater language skills, and these elicit and are affected by socialization pressures from adults (and peers) to channel these differences into gender-role-consistent gender differences in emotion expressions as youth develop from infancy into the toddler/preschool period and childhood. In addition, there are contextual forces at work, shaping when and whether children express emotions according to gender roles depending on an interaction between their (biologically and/or socialization-influenced) tendency to express gender-role-consistent emotions and the particular social environment and larger cultural context in which they are found» (Chaplin, 2015, p. 10; see also Chaplin & Aldao, 2013). The model was derived from the meta-analysis of 164 studies, involving 21,709 participants from early infancy to adolescence. Girls display more positive emotions and internalizing emotions (e.g., sadness, anxiety, sympathy) than boys, and boys display more externalizing emotions than girls. Gender differences in positive emotions and externalizing emotions increase in middle childhood and adolescence. Children also use different display rules according to the person they are interacting with, and gender differences are less pronounced when children interact with parents, are more pronounced with unfamiliar adults (for positive emotions), and with peers or when they are alone (for externalizing emotions). In the literature, the effect of gender on ER development has not be detected systematically. In some studies, boys and girls did not show differences in their ER (Gilpin et al. 2015; Jahromi & Stifter, 2008), while in other research differences were found (Sala et al., 2014) in the use of ER strategies, with boys using more often behavioural strategies and girls referring more often to social support. Charman et al. (2002) discussion about the lack of statistical power to demonstrate small gender differences could also be the case for the research about ER in preschoolers.

Theory of Mind and Emotion Regulation

To our knowledge, only five studies investigated the relationship between ToM and ER in typically developing preschoolers: Baurain & Nader-Grosbois (2013), Gilpin et al. (2015), Hudson & Jacques (2014), Jahromi & Stifter (2008), and Lieberman et al. (2007). They involved in average 80 subjects (N range = 45-107) from 3 to 7 years of age; the participants were English-speaking (4 studies) or French-speaking children (Baurain & Nader-Grosbois, 2013). Three studies detected small to medium correlations between ToM and ER, where False Belief or ToM-emotion tasks were used to measure ToM, while ER scores were obtained during disappointment tasks

or dyadic games inducing positive versus negative emotions. According to Baurain & Nader-Grosbois (2013), children's performances on ToM tasks regarding causes and consequences of emotion correlated to two ER aspects coded by trained observers, namely emotional expression (r = .54) and behaviour towards social rules (r = .40). Hudson & Jacques (2014) found a small correlation between preschoolers' scores on a False Belief task and ER measured through disappointing gift task (r = .19). Lieberman et al. (2007) found that the association between False Belief tasks and ER, measured through a disappointment procedure and parents' report about their children emotional control, approached significance (r = .26, p < .054) only when controlling for verbal ability, measured with the PPVT-3 or PPVT-4. Two studies found no significant association between ToM and ER (Gilpin et al. 2015; Jahromi & Stifter, 2008): also in this case, False Belief tasks were used to assess ToM, and a disappointment task or a teacher-report tool was used for measuring ER.

Three of these studies also investigated the role of language on the relationship between ToM and ER, using different versions of the PPVT test, that measures the receptive VS of children: Gilpin et al. (2015) only detected an association between ToM and language (r = .43); Jahromi & Stifter (2008) and Lieberman et al. (2007) used receptive VS as a control variable in correlations between ToM and ER, and the first study found no association between the two constructs, while the second one found a positive correlation, as discussed above.

Of the five studies considered, two did not analyse gender differences in ToM or ER performance (Baurain & Nader-Grosbois, 2013; Lieberman et al., 2007), while three found no differences (Gilpin et al., 2015; Hudson & Jacques, 2014; Jahromi & Stifter, 2008).

The current study

Summarizing, ToM and ER Cognitive Reappraisal strategies develop from 3 to 6 years of age; both these competencies are related to language development and partly to gender, thus it is worth analysing the relationship between them. The present study aims to investigate the relationship between ToM and ER in typically developing preschoolers, also considering the role of age, receptive VS, and gender as control variables.

Our study might contribute to the current literature for many reasons. First, the scarce literature devoted to this topic shows mixed results, with some studies reporting positive correlations between ToM and ER, and others detecting a lack of association: our research might allow for a better understanding of the relationship between these constructs. Second, in this

paper ToM is assessed through a comprehensive test, the ToM Storybooks, consisting of repeated different tasks that allow a more reliable measure, as previously discussed. Third, ER is assessed with a narrative story stem procedure, that allows to directly observe and categorise the children's representations of ER strategies. Fourth, given the existing association between language and ToM or ER, and given that most of studies has been performed with English-speaking participants, it is worth testing the literature findings on different linguistic communities, as we did in this research conducted with Italian children.

Concerning the hypotheses of the current study, first, we expect an overall effect of age on ToM and language (specifically, VS), on ER *Behavioural* and *Cognitive Reappraisal Strategies* but not on ER *Attentional Deployment* and *Social Support* strategies. Second, we expect limited gender differences, particularly in ER *Attentional Deployment* and *Social Support* strategies, like in previous studies. Third, according to the existing literature, we expect language to be related to both ToM and ER. Fourth, given that the few studies in the literature analyzing the relationship between ToM and ER showed mixed results, we will explore the associations between the three ToM scores (*Quantitative*, *Qualitative*, and *Total*) and the four ER strategies.

Method

Participants

The participants were 62 Italian children (32 girls, 51.6%), aged from 3 to 6 years (range: 37-75 months, $M = 59.09$, $SD = 10.30$). Their average non-verbal IQ was 98.97 ($SD = 13.98$).

The inclusion criteria consisted of having a chronological age between 3 and 6 years and being enrolled in kindergarten; the main exclusion criterion was the presence of developmental difficulties or disabilities. Parents and teachers reported that the children had no health or mental issues. All the children were attending kindergarten and none of them had been already attending primary school. The participants of the current study were also included in the study by [Blind reference 1] and were part of the ToM Storybooks Italian standardisation sample [Blind reference 2].

Procedures and Measures

To take part in the study, the parents gave their written informed consent and the children their verbal consent. The research was conducted according

to the ethical principles of the Declaration of Helsinki. Data collection took place in 2011-2012. For the familiarisation, the administers of the tests spent a few hours with the children in the class to present themselves and freely play together. In the following weeks, the ToM Storybooks, the SIRE, and the PPVT-R were individually administered in the same period, in a quiet room at the kindergarten; the average time interval between the administration of the ToM Storybooks and the SIRE was 8 days (*SD* = 24 days).

The ToM Storybooks is a comprehensive test evaluating five ToM components: recognizing and naming emotions, making a distinction between physical and mental phenomena, understanding that perception leads to knowledge, understanding that desires and beliefs affect behaviours; a classic content False Belief task is also included. The test consists of six colour illustrated books telling the stories of Sam, a 5-year-old boy. Some tasks are proposed several times, to allow a more reliable measure. The administration requires about 40 minutes. The test has 95 items, that contribute to three scores. The *Quantitative score* varies from 0 to 77 and sums 77 close-ended questions. The *Qualitative score* varies from 0 to 36 and sums 18 open-ended questions. The children get 2 points if they spontaneously attribute mental states to Sam; 1 point if they simply explain situational aspects, or 0 points if they give incorrect answers. The *Total score* is the sum of the *Quantitative* and the *Qualitative score*. The ToM Storybooks have good internal consistency, test-retest reliability, inter-rater reliability, divergent and convergent validity (Blijd-Hoogewys et al., 2008; Bulgarelli et al., 2015; Bulgarelli, Testa, & Molina, 2022).

The Emotion Regulation Story Stems (*Storie Interrotte per la Regolazione Emotiva*-SIRE) is an experimental story stem procedure; cloth and wooden puppets are used to animate the stories (Sala et al., 2014). The order in which the narratives are told is always the same and follows the increasing complexity of the emotions cited in the different stories (fear, anger, sadness, shame, and guilt). The first story is a *warm-up story*: a narrative that has no implications with the theme of emotions. The administration of the procedure required about 20 minutes. The procedure was videotaped, transcribed, and coded by two independent judges, and the interrater reliability based on 20% of the tapes, evaluated through Cohen's K, ranged from .64 and 1.0 (Sala et al., 2014). The following coding system, developed by the Modal Model proposed by Gross (Gross, 2024), was used to assess the four ER strategies addressed in the children's responses: *Behavioural Strategy* was coded when the child refers to the protagonist's actions that are directed to change the situation in order to manage the emotion (e.g.: "He fixes the broken toy"); *Social Support* was coded when the narrative mentions the intervention of another character helping the protagonist to overcome the negative emotion

(e.g.: “The teacher comes and sits close to her”); *Attentional Deployment* was coded when the child responds refereeing to actions or thoughts distracting the protagonist and also the core of the narration from the emotion (e.g.: “He starts to play with the toys that he likes”); while Cognitive Reappraisal was coded when the child refers to protagonist’s thoughts or evaluations that, through the meaning’s modification of the situation, permit the regulation of the emotion (e.g.: “She understands that the next year she also will go to another school and she will find her friend again”). The coding system was meant to be mutually exclusive, and only one strategy was assigned to each of the child’s five responses. Thus, each strategy score varied from 0 to 5, because the child could have potentially used the same strategy in every answer.

The receptive VS was measured with the Italian version of the PPVT-R (Stella et al., 2000). Refer to the Supplementary Materials for detailed information about the tests.

Data analysis

We conducted a cross-sectional study, comparing different age groups. The raw scores of the three ToM scores, the four ER scores, and the VS score were used. Two different methods for age calculation have been used depending on the analysis. A categorical measure was used in ANOVA analysis. The children were divided into four age groups: 3-, 4-, 5-, and 6-year-olds; in the 3-year-old group, the children’s age ranged from 2 years, 6 months, and 1 day to 3 years and six months; the same strategy was followed for the other age groups. In the correlation analysis, we utilized an approximate decimal age, calculated from the date of birth to the date of measurement of the ToM Storybooks. First, we tested the presence of age and gender differences in the participants’ performances. One-way Anova, with Bonferroni post-hoc test, was run on ratio measures, i.e., the ToM and VS average scores, and non-parametric Kruskal-Wallis (for k-independent samples) or Mann-Whitney (for 2 independent samples) exact tests (Monte Carlo method) were conducted on the ordinal measures, i.e., the ER median scores. Pearson correlation was used to test the associations between the ratio measures, i.e., ToM scores, decimal age, and the other variables, while Spearman correlation was used for the associations between the ordinal measures, i.e., ER scores, decimal age, and other variables. Finally, partial correlations were run to test the associations between ToM and ER scores, controlling for age, and VS. The statistical analyses were conducted using SPSS 28.0 for Windows (SPSS Inc., Chicago, IL, USA).

Results

Age differences in Theory of Mind, Emotion regulation, and Vocabulary Size

The descriptives of ToM, ER, and VS scores by age groups are reported in Table 1. The one-way Anova analysis showed that the ToM performances grew regularly as children got older: *Quantitative* score ($F(3,58) = 19.668$, $p < .001$), *Qualitative* score ($F(3,58) = 20.490$, $p < .001$), and *Total* score ($F(3,58) = 22.824$, $p < .001$). Among the ER scores, the Kruskal-Wallis exact test (Monte Carlo method) confirmed that the use of *Behavioural Strategy* improved with age ($H = 9.235$, $p = .023$), as well as the level of appeal to *Cognitive Reappraisal* strategies ($H = 11.261$, $p = .009$); both *Social Support* and *Attentional Deployment* did not grow as a function of age (all $p > .20$). Also VS significantly increased with age (one-way Anova, $F(3,58) = 7.660$, $p < .001$). Post-hoc comparisons between the age groups' scores are reported in Table 1 (see Tab.1).

The correlations between ToM, ER, VS scores, and age are reported in Table 2. The three ToM scores were strongly and positively associated with age; among the four ER scores, only *Behavioural Strategy* and *Cognitive Reappraisal* were positively and significantly associated with age. VS score also significantly varied as a function of age (see Tab. 2).

Tab. 1 – *Descriptives and age differences between the age groups' average scores (SD) of Theory of Mind, Emotion Recognition, and Vocabulary Size*

	Age groups			
	3 years (n = 6)	*4 years (n = 14)*	*5 years (n = 26)*	*6 years (n = 16)*
ToM Quantitative °	32.50 (6.31) [a,b]	42.57 (9.57) [c]	54.26 (6.35) [a,c]	56.56 (9.35) [b,c]
ToM Qualitative °	.33 (.52) [a,b]	3.00 (2.80) [c,d]	8.00 (3.63) [a,c]	12.63 (5.92) [b,d]
ToM Total °	32.83 (6.56) [a,b]	45.57 (12.16) [c,d]	62.26 (8.95) [a,c]	69.18 (14.13) [b,d]
ER Behavioural Strategy ^	.00 (.00) [a,b,c]	.57 (.65) [a]	1.15 (1.32) [b]	1.38 (1.41) [c]
ER Social Support ^	1.33 (1.21)	1.93 (1.64)	1.88 (1.42)	1.94 (1.12)
ER Attentional Deployment ^	.00 (.00)	.00 (.00)	.27 (.60)	.19 (.40)
ER Cognitive Reappraisal ^	.00 (.00)	.000 (.00) [a]	.27 (.60) [b]	.81 (1.05) [a,b]

VS°	23.50 (16.91) [a,b]	46.14 (18.97) [c]	56.38 (21.02) [a]	69.88 (23.35) [b,c]

Note: Within each score reported in the raw, corresponding superscript letters refer to significantly different groups (°Anova, Bonferroni test, $p < .05$; ^Mann-Whitney exact test, Monte Carlo method, $p < .05$).

Gender Differences in Theory of Mind, Emotion Regulation, and Vocabulary Size

No significant gender differences were found in the three ToM scores (One-way Anova; all $p > .60$) and in VS (One-way Anova; $p > .30$). As to the four ER scores, boys and girls differed in the use of *Behavioural Strategy*, as they appeared more frequently in the males' answers (boys: median = 1.00, girls: median = 0.00; Mann-Whitney exact test, Monte Carlo method, $U = 328.000$, $p = .022$). They also differed in the use of *Social Support*, that appeared more frequently in the females' answers (boys: median = 1.00, girls: median = 2.00; Mann-Whitney exact test, Monte Carlo method, $U = 327.000$, $p = .026$). No gender differences were found in *Attentional Deployment* ($p > .50$) and Cognitive Reappraisal ($p > .50$).

Tab. 2 – *Correlations between age, Vocabulary Size (VS), Theory of Mind (ToM), and Emotion Recognition (ER) scores*

		Age	**VS**	**ER BS**	**ER SS**	**ER AD**	**ER CR**
Correlations	*Age*	1.00	.565**	.370**	.085	.085	.462**
	VS	.565**	1.00	.423**	.119	-.129	.325*
	ToM Quant.	.727**	.684**	.310*	.104	.111	.409**
	ToM Qual.	.727**	.663**	.300*	.065	.120	.531**
	ToM Total	.763**	.711**	.322*	.095	.120	.473**
Partial correlations controlling for age, gender, and VS	*ToM Quant.*	--	--	-.035	-.052	.188	.108
	ToM Qual.	--	--	-.049	-.105	.191	.345**
	ToM Total	--	--	-.046	-.082	.219	.221

Note: BS = Behavioural strategy; SS = Social support; AD = Attentional deployment; CR = Cognitive reappraisal. Peason's correlations were run for Theory of Mind scores and Vocabulary Size score and Spearman correlations for Emotion Regulation scores; * $p < .05$, ** $p < .01$.

Relationship between Theory of Mind, Emotion Regulation, and Vocabulary Size

Table 2 also reports the correlations between the children's scores. The three ToM scores were significantly correlated to the ER *Behavioural Strategy* and the ER *Cognitive Reappraisal* scores; the second type of association was stronger. VS was significantly correlated to the three ToM scores and to ER *Behavioural Strategy* and ER *Cognitive Reappraisal* (see Table 2).

To better describe the relationship between ToM and ER scores, a partial correlation was conducted, controlling for age in months, IQ, and VS: only the correlation between the ToM *Qualitative* score and the ER *Cognitive Reappraisal* score was still moderate and significant (see Table 2).

Discussion

The current study investigated the relationship between ToM and ER, controlling for age, gender, and VS in 3- to 6-year-old Italian children. Our first hypothesis, based on the consolidated literature findings, was that the children's performances on different tasks would improve with age. The ToM Storybooks scores (*Qualitative*, *Quantitative*, and *Total*) and the VS score grew regularly with age, as expected from previous research using the same instruments (Blijd-Hoogewys et al., 2008; Bulgarelli et al., 2015; Lieberman et al., 2007). Also, the SIRE *Behavioural Strategy* and *Cognitive Reappraisal* scores were associated with age, showing that these two strategies are susceptible to age-related developmental processes in preschool years. Besides, the SIRE *Attentional Deployment* and *Social Support* scores were not correlated with age: probably, these two strategies depended more on socialization process (see below) or were more susceptible to inter-individual differences than to age. This pattern of results was also expected, as the current study involved the same children of [Blind reference 1]. Moreover, the literature devoted to ER development in preschool age systematically reported that this competence improved in the first years of life (Baurain & Nader-Grosbois, 2013; Lieberman et al., 2007).

The second hypothesis was also confirmed. No significant gender differences were found in the three ToM scores and in VS, coherently with previous research (Blijd-Hoogewys et al., 2022; Bulgarelli et al., 2015; Charman et al., 2002). As to the four ER scores, as expected, we observed the same pattern of results of Sala et al. (2014): the use of *Behavioural Strategy* was

more frequent in the males' answers, and the reference to *Social Support* appeared more frequently in the females' responses, while no differences were found in *Attentional Deployment* ($p > .50$) and *Cognitive Reappraisal*. Interestingly, ER *Social Support* strategy was susceptible to gender and did not grow with age, showing that, probably, socialisation processes played a crucial role in this case, as expected by Chaplin's bio-psycho-social contextual model of emotion expression (Chaplin & Aldao, 2013). The model was derived from the meta-analysis of 164 studies, involving 21,709 participants from early infancy to adolescence. Girls display more positive emotions and internalizing emotions (e.g., sadness, anxiety, sympathy) than boys, and boys display more externalizing emotions than girls. Children also use different display rules according to the person they are interacting with.

Considering the third hypothesis regarding language, we found that VS was significantly associated with the three ToM scores, as expected from the literature (Bulgarelli, Henning & Bertin, 2022; Milligan et al., 2007). VS was also associated with ER scores – namely *Behavioural Strategy* and *Cognitive Reappraisal* –, as already detected in previous studies (Cohen & Mendez, 2009; Gilpin et al., 2015; Ren et al., 2009; Reylly & Downer, 2019; Roben et al., 2013).

Moving towards the fourth main objective of the current study, our aim was to explore the interrelations between the ToM and the ER scores, as some previous findings showed that these constructs are correlated (Baurain & Nader-Grosbois, 2013; Hudson & Jacques, 2014; Lieberman et al., 2007), while other did not find such an association (Gilpin et al., 2015; Jahromi & Stifter, 2008). The three ToM scores (*Qualitative*, *Quantitative*, and *Total* scores) and the ER *Behavioural Strategy* and ER *Cognitive Reappraisal* scores were associated; actually, these scores also showed a clear developmental trend, as they were correlated with age. Conversely, the three ToM scores did not correlate with ER *Social Support* and *Attentional Deployment*, which were not associated with age and depended on gender.

Taking into consideration other variables that could influence the relationship between ToM and ER, we conducted partial correlations controlling for age, gender, and language: the correlation between the ToM *Qualitative* score and the ER *Cognitive Reappraisal* score turned out to be the only significant one. Also Lieberman et al. (2007) found a positive correlation between ToM and ER to approach significance ($r = .26$, $p < .054$) only when controlling for verbal ability, measured with the same tool that we used, the PPTV. Such a result is highly interesting, given the nature of the measures that we used in this study. In fact, the *Qualitative* ToM score depended on the ability of the child to reason about mental states: highest *Qualitative* ToM scores were reached when the child spontaneously talked about emotions,

desires, and beliefs, attributing them to the protagonists of the stories while answering to the open-ended questions of the ToM Storybooks. Coherently, the ER *Cognitive Reappraisal* score was attributed when the child referred to the use of cognitive abilities (thoughts, evaluations, etc.) that act on the meaning of the situation. Thus, ToM abilities measured through the ToM Storybooks were associated with more complex use of ER strategies in the SIRE story stem procedure.

The current study adds to the literature thanks to a series of methodological features. First, we used two new tools. The ToM Storybooks is a comprehensive test that consists of several tasks tapping five components of basic ToM. This characteristic allows for a more stable and reliable measure, that differentiates our study from the other five that already analysed the relationship between ToM and ER, using only False Belief or ToM-emotion tasks. Second, ER was assessed here with the SIRE story stem, and this makes the current study the first one to analyse ToM and ER association using a narrative procedure, which permits to observe how children develop the script and which kind of strategies to regulate emotion they use. In previous research, disappointment tasks, teacher-report tools, or dyadic games inducing positive versus negative emotions were used. Third, to our knowledge, this was the first study about the relationship among ToM, ER, and language to be conducted with Italian speaking children. As discussed in the introduction, ToM and language are inter-independent competencies (Milligan et al., 2007) and ToM processing during False Belief tasks was proved to partly depend on cultural and linguistic features, both when comparisons were made between Western and Eastern countries, and among Western European countries belonging to two cultural Latin and Germanic clusters (Blijd-Hoogewys et al., 2022; Wellman, 2018). Given that the literature devoted to the relationship between ToM, ER, and language is mainly run on English-speaking children, the current study contributes to filling a gap in the field.

Limitations of the current study

Concerning the limitations of our study, a rather small sample of children was involved (N = 62) and, specifically, the size of the younger group of participants could impact the robustness of the results. The study has an exploratory nature and the findings cannot be generalised to different populations.

The study was correlational, thus no casual relationship among the variables could be tested. For instance, language and ToM are inter-related competences, yet the impact of early language on future ToM abilities was proved to be stronger than the opposite (Milligam et al., 2007). Therefore,

future longitudinal research or intervention studies about ToM and ER could help better understand if ToM and ER are inter-related competence and if one of the two has greater impact on developing the other one. Also, future studies could further analyse the role of moderating variables on the relationship between ToM and ER, such as language and gender.

In summary, the current study showed that ToM and ER are associated competencies in preschool Italian children and that the child's ability to talk about emotions, desires, and beliefs spontaneously, is connected to higher cognitive ER strategies, even when language ability, gender, and age are controlled for.

Acknowledgments

We are grateful to the parents and children who participated in this study and to the teachers and the directors of the schools where the children were recruited. We also thank the students who contributed to the data collection.

Disclosure statement

No potential conflict of interest was reported by the authors.

Author contributions

Authors NS and PM designed the study. Author DB and NS collected the data. BR, DB, NS, and PM analysed the data. Author DB wrote the first versions of the paper and author BR, NS, and PM edited the final paper. All authors approved the final version of the manuscript for submission.

Data availability

Anonymised data are available from the authors upon reasonable request. Enquiries can be sent to the corresponding author.

References

Baurain, C., & Nader-Grosbois, N. (2013). Theory of mind, socio-emotional problem-solving, socio-emotional regulation in children with intellectual disability and in typically developing children. *Journal of Autism and Developmental Disorders*, *43*, 1080-1097. DOI: 10.1007/s10803-012-1651-4.

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology*, *10*(2905), 1-23. DOI: 10.3389/fpsyg.2019.02905.

Belacchi, C. (2022). Teoria della Mente e linguaggio: come condividere emozioni e pensieri con le parole [Theory of Mind and language: how to share emotions and thoughts with words]. In I. Castelli, D. Massaro, & A. Marchetti (Eds.), *La competenza sociale nei bambini* (pp. 101-134). Bergamo: Edizioni Junior-Bambini srl.

Blijd-Hoogewys, E. M. A., Bulgarelli, D., Molina, P., & van Geert, P. L. C. (2022). Nonlinearities in theory-of-mind development: New evidence from Dutch and Italian boys and girls. *European Journal of Developmental Psychology*, *19*(6), 905-924. DOI: 10.1080/17405629.2021.2020636.

Blijd-Hoogewys, E. M. A., van Geert, P. L. C., Serra, M., & Minderaa, R. B. (2008). Measuring theory of mind in children. Psychometric properties of the ToM Storybooks. *Journal of Autism and Developmental Disorders*, *38*(10), 1907-1930. DOI: 10.1007/s10803-008-0585-3.

Bulgarelli, D., Henning, A., & Bertin, E. (2022). Social cognition and its main correlates in childhood. In P. K. Smith & C. H. Craing (Eds.), *The Wiley-Blackwell Handbook of Childhood Social Development* (pp. 613-630). DOI: 10.1002/9781119679028.ch33.

Bulgarelli, D., Testa, S., & Molina, P. (2015). Factorial structure of the 'ToM Storybooks': A test evaluating multiple components of theory of mind. *British Journal of Developmental Psychology*, *33*(2), 187-202. DOI: 10.1111/bjdp.12062.

Bulgarelli, D., Testa, S., & Molina, P. (2022). Theory of mind development in Italian children with specific language impairment and autism spectrum disorder: delay, deficit, or neither?. *Journal of Autism and Developmental Disorders*, *52*(12), 5356-5366. DOI: 10.1007/s10803-021-05389-8.

Chaplin, T. M. (2015). Gender and emotion expression: A developmental contextual perspective. *Emotion Review*, *7*(1), 14-21. DOI: 10.1177/1754073914544408.

Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin*, *139*(4), 735-765. DOI : 10.1037/a0030737.

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development?. *Social development*, 11(1), 1-10. DOI: 10.1111/1467-9507.00183.

Cohen, J. S., & Mendez, J. L. (2009). Emotion regulation, language ability, and the stability of preschool children's peer play behavior. *Early Education and Development*, *20*(6), 1016-1037. DOI: 10.1080/10409280903430745.

Cole, P. M., Dennis, T. A., Smith-Simon, K. E., & Cohen, L. H. (2009). Preschoolers' emotion regulation strategy understanding: Relations with emotion socialization and child self-regulation. *Social Development*, *18*(2), 324-352. DOI: 10.1111/j.1467-9507.2008.00503.x.

Davis, E. L., Levine, L. J., Lench, H. C., & Quas, J. A. (2010). Metacognitive emotion regulation: children's awareness that changing thoughts and goals can alleviate negative emotions. *Emotion*, *10*(4), 498. DOI: 10.1037/a0018428.

Eisenberg, N., Sadovsky, A., & Spinrad, T. L. (2005). Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. *New Directions for Child and Adolescent Development*, *109*, 109-118. DOI: 10.1002/cd.143.

Farrar, M. J., Benigno, J. P., Tompkins, V., & Gage, N. A. (2017). Are there different pathways to explicit false belief understanding? General language and complementation in typical and atypical children. *Cognitive Development*, *43*, 49-66.

Gilpin, A. T., Brown, M. M., & Pierucci, J. M. (2015). Relations between fantasy orientation and emotion regulation in preschool. *Early Education and Development*, *26*(7), 920-932. DOI: 10.1080/10409289.2015.1000716.

Gross, J. J. (2024). Conceptual foundations of Emotion regulation. In J. J. Gross & B. Q. Ford (Eds.), *Handbook of emotion regulation. Third edition* (pp. 3-12). New York, NY: Guilford Press.

Hudson, A., & Jacques, S. (2014). Put on a happy face! Inhibitory control and socioemotional knowledge predict emotion regulation in 5-to 7-year-olds. *Journal of Experimental Child Psychology*, *123*, 36-52. DOI: 10.1016/j.jecp.2014.01.012.

Jahromi, L. B., & Stifter, C. A. (2008). Individual differences in preschoolers' self-regulation and theory of mind. *Merrill-Palmer Quarterly (1982)*, *54*(1), 125-150. DOI: 10.1353/mpq.2008.0007.

Liebermann, D., Giesbrecht, G. F., & Müller, U. (2007). Cognitive and emotional aspects of self-regulation in preschoolers. *Cognitive Development*, *22*(4), 511-529. DOI: 10.1016/j.cogdev.2007.08.005.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622-646. DOI: 10.1111/j.1467-8624.2007.01018.x.

Reilly, S. E., & Downer, J. T. (2019). Roles of executive functioning and language in developing low-income preschoolers' behavior and emotion regulation. *Early Childhood Research Quarterly*, *49*, 229-240. DOI: 10.1016/j.ecresq.2019.07.006.

Ren, Y., Wyver, S., Xu Rattanasone, N., & Demuth, K. (2016). Social competence and language skills in Mandarin-English bilingual preschoolers: The moderation effect of emotion regulation. *Early Education and Development*, *27*(3), 303-317. DOI: 10.1080/10409289.2015.1066639.

Roben, C. K., Cole, P. M., & Armstrong, L. M. (2013). Longitudinal relations among language skills, anger expression, and regulatory strategies in early childhood. *Child Development*, *84*(3), 891-905. DOI: 10.1111/cdev.12027.

Sala, M. N., Pons, F., & Molina, P. (2014). Emotion regulation strategies in preschool children. *British Journal of Developmental Psychology*, *32*(4), 440-453. DOI: 10.1111/bjdp.12055.
Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *PPVT-R, Peabody Picture Vocabulary Test-Revised*. Torino: Omega Edizioni.
Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728-755. DOI: 10.1080/17405629.2018.1435413.

Supplementary materials

Procedures and Measures

Description of the tests

The ToM Storybooks is a comprehensive test evaluating five ToM components, based on Wellman's theory (1990): recognizing and naming emotions, making a distinction between physical and mental phenomena, understanding that perception leads to knowledge, understanding that desires and beliefs affect behaviors; a classic content False Belief task is also included. The test consists of six color illustrated books telling the stories of the protagonist Sam, a 5-year-old boy, his family, and friends. Each task is presented within the context of a story and some tasks are proposed several times, to allow a more reliable measure (Figure 1).

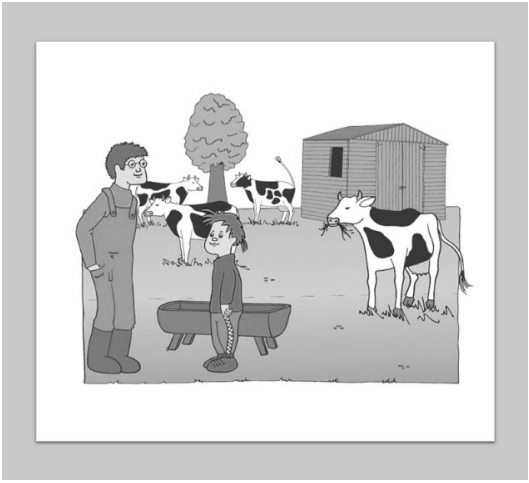

Figure 1. Example of one task of the ToM Storybooks

The administration of the test required about 40 minutes, inclusive of a short break after reading three books. The test has 95 items, that contribute to three scores. The Quantitative score varies from 0 to 77 and sums 77 close-ended questions, some of which require a non-verbal response (e.g., pointing images). The Qualitative score varies from 0 to 36 and sums 18 open-ended questions. As an example, to illustrate the scoring of these answers, let us take into consideration the False Belief task that is included in the test. The experimenter reads this story while the child is looking at the pictures: "Grandpa and grandma are paying Sam a visit. Sam gets rollerblades from grandpa and grandma. […] Sam puts the rollerblades [point out] in the toy trunk. Then, he goes upstairs to play with his crane. When Sam has left, his sister goes to the toy trunk. She likes to tease her brother. Lotje hides the rollerblades [point out] in the box! And then, she goes quickly outside. Then, Sam is coming back. He wants to roller skate. Where will Sam look for his rollerblades? Why is Sam looking [there]?". The children get 2 points if they spontaneously attribute mental states to Sam (e.g.: "Sam looks for the skates in the box because he thinks they are there"), 1 point if they simply explain situational aspects (e.g.: "Sam looks for the skates in the box because he put them there"), or 0 points if they give incorrect answers (e.g.: "Sam looks for the skates in the toy trunk because they are there"). The Total score is the sum of the Quantitative and the Qualitative score. The ToM Storybooks have been standardized in the Netherlands and Italy and it has good internal consistency, test-retest reliability, and inter-rater reliability (Blijd-Hoogewyset al., 2008; Bulgarelli et al., 2015). Both the Dutch and the Italian version show a five-component factorial structure, that are similar: for a discussion about the two factorial structures, see Bulgarelli et al. (2015). The test also has discriminant validity, as it differentiates between autistic and typically developing children, and divergent and convergent validity (Blijd-Hoogewys et al., 2008; Bulgarelli, Testa & Molina, 2022).

The Emotion Regulation Story Stems (*Storie Interrotte per la Regolazione Emotiva*-SIRE) is an experimental story stem procedure; cloth and wooden puppets are used to animate the stories (Sala et al., 2014). The procedure was derived from the story stem tools developed to research attachment representations (Bretherton et al., 1990; Green et al., 2000). One warm-up and five emotional stories take place in the school context; such an environment was chosen for two reasons: first, it permits to investigate the theme of ER without directly involving the representations of the attachment to parents; second, kindergarten was indicated as the first context in which children's ability to regulate emotions take place in addition to the family(Silvers et al., 2012). The order in which the narratives are told is always the same and follows the increasing complexity of the emotions cited in the different

stories (fear, anger, sadness, shame, and guilt). The protagonist of every tale is always the same character, that can be a boy or a girl corresponding to the gender of the child who answers to the procedure. Coherently with other story stems developed in the field of attachment research, such as ASCT (Bretherton et al., 1990), the first story is a warm-up story: a narrative that has no implications with the theme of emotions. The researcher initially tells the warm-up story for the purpose of establishing contact with the child and in order to give her/him the opportunity to familiarize with the task, the context, and the materials. The warm-up story describes an ordinary day at school in which children are going to play in the yard (Figure 2).

Figure 2. Warm-up story of the SIRE

Subsequently, the researcher tells five stories related to the five emotions cited. In the story associated with fear, the protagonist gets lost during school trip. In the narrative related to anger, the protagonist's companions make her/him drop her/his snack. The best friend of the protagonist is moving to another city and leaves the classroom in the story associated with sadness. The following narrative involves the child falling in front of people and feeling ashamed (Figure 3).While in the last story the protagonist breaks a friend's favorite toy and feels guilty. All the stories are narrated both verbally

and by moving the puppets simultaneously. At the end of each story, the researcher labels the emotion involved and asks the child to continue the narrative. The administration of the procedure required about 20 minutes, depending also on the child's playfulness, responses, and narrative skills. The procedure was videotaped, transcribed, and coded by two independent judges, and the interrater reliability based on 20% of the tapes, evaluated through Cohen's K, ranged from .64 and 1.0 (Sala et al., 2014).

Figure 3. Shame story of the SIRE

The following coding system, developed by the Modal Model proposed by Gross (Gross, 1998; 2014), was used to assess the four ER strategies addressed in the children's responses: Behavioral Strategy was coded when the child refers to the protagonist's actions that are directed to change the situation in order to manage the emotion (e.g.: "He fixes the broken toy"); Social Support was coded when the narrative mentions the intervention of another character helping the protagonist to overcome negative emotion (e.g.: "The teacher comes and sits closet to her"); Attentional Deployment was coded when the child responds refereeing to actions or thoughts distracting the protagonist and also the core of the narration from the emotion (e.g.: "He starts

to play with the toys that he likes"); while Cognitive Reappraisal was coded when the child refers to protagonist's thoughts or evaluations that, through the meaning modification of the situation, permit the regulation of the emotion (e.g.: "She understands that the next year she also will go to another school and she will find her friend again"). The coding system was meant to be mutually exclusive, and only one strategy was assigned to each of the child's five responses. Thus, each strategy score varied from 0 to 5, because the child could have potentially used the same strategy in every answer.

The receptive VS was measured with the Italian version of the PPVT-R (Dunn & Dunn, 1981; Stella et al., 2000), a well-known test for evaluating receptive language. It asked the child to recognize the image associated with a target word read by the interviewer; the image was inserted on a page together with three distractors. The tests consisted of 175 tables, of which target words were increasingly more difficult. Thanks to the evaluation of a floor line and a ceiling line, the administration of the test requires about 15 minutes.

References

Blijd-Hoogewys, E. M. A., Van Geert, P. L. C., Serra, M., & Minderaa, R. B. (2008). Measuring theory of mind in children. Psychometric properties of the ToM Storybooks. *Journal of Autism and Developmental Disorders, 38*(10), 1907-1930. DOI: 10.1007/s10803-008-0585-3.

Bretherton, I., Ridgeway, D., & Cassidy, J. (1990). Assessing internal working models of the attachment relationship: An attachment story completion task for 3-year-olds. In M. T. Greenberg, D. Cicchetti & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 273-308). Chicago, IL US: University of Chicago Press.

Bulgarelli, D., Henning, A., & Bertin, E. (2022). Social cognition and its maincorrelates in childhood. In P. K. Smith & C. H. Craing (Eds.), *The Wiley-Blackwell Handbook of Childhood Social Development* (pp. 613-630). DOI: 10.1002/9781119679028.ch33.

Bulgarelli, D., Testa, S., & Molina, P. (2015). Factorial structure of the 'ToM Storybooks': A test evaluating multiple components of theory of mind. *British Journal of Developmental Psychology, 33*(2), 187-202. DOI: 10.1111/bjdp.12062.

Dunn, L M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test Revised*. American Gudance Service.

Green, J., Stanley, C., Smith, V., & Goldwyn, R. (2000). A new method of evaluating attachment representations in the young school-age children: The Manchester Child Attachment Story Task (MCAST). *Attachment & Human Development, 2*(1), 48-70. DOI: 10.1080/146167300361318.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271-299. DOI: 10.1037/1089-2680.2.3.271.
Gross, J. J. (2014). Emotion regulation: Conceptual and empirical foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3-20). New York, NY: Guilford Press.
Sala, M. N., Pons, F., & Molina, P. (2014). Emotion regulation strategies in preschool children. *British Journal of Developmental Psychology, 32*(4), 440-453. DOI: 10.1111/bjdp.12055.
Silvers, J. A., McRae, K., Gabrieli, J. D. E., Gross, J. J., Remy, K. A., & Ochsner, K. N. (2012). Age-related differences in emotional reactivity, regulation, and rejection sensitivity in adolescence. *Emotion, 12*(6), 1235-1247. DOI: 10.1037/a0028297.
Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). PPVT-R, Peabody Picture Vocabulary Test-Revised. Torino: Omega Edizioni.
Wellman, H. M. (1990). *The child's theory of mind*. Cambrige, MA: MIT Press.