Check for updates

Infant and Child Development

WILEY

REPORT OPEN ACCESS

Relations Between the Development of False Belief and Scientific Reasoning in Turkish Preschoolers

Mesut Saçkes[1] | Sonnur Işıtan[1] | Sinem Güçhan Özgül[1] | Kerem Avci[1] | Kathy Cabe Trundle[2] | David M. Sobel[3]

[1]Balikesir University, Balikesir, Turkey | [2]Utah State University, Logan, Utah, USA | [3]Brown University, Providence, Rhode Island, USA

Correspondence: Mesut Saçkes (msackes@gmail.com; msackes@balikesir.edu.tr)

Received: 9 June 2025 | **Revised:** 7 February 2026 | **Accepted:** 17 February 2026

Handling Editor: Jonathan Kominsky

Keywords: false belief | preschoolers | scientific reasoning | Turkish sample

ABSTRACT

Previous studies have shown relations between the emergence of theory of mind capacities and children's scientific reasoning in Western samples. Previous studies have also shown that Turkish preschoolers have a different trajectory for theory of mind development than their Western counterparts. This study extends these previous findings to examine relations between the emergence of false belief understanding and different facets of scientific reasoning in Turkish preschoolers (aged 3 to 5), particularly focused on the difference between understanding others' false belief and one's own representational change. Study 1 revealed a positive relation between preschoolers' capacity to infer another's false belief and their experimentation skills when unconfounded interventions were presented to them. This finding did not extend to children's ability to report their own representational change. Study 2, however, showed that this relation with false belief did not extend to a case in which children had to design their own interventions to implement the control of variables strategy. The results suggest links between children's understanding of false belief and their nascent scientific reasoning among Turkish preschoolers.

The cornerstone of children's emerging scientific reasoning is the ability to differentiate between theory and evidence (e.g., Kuhn 1989, 2010; Schauble et al. 1995). Scientific reasoning necessitates that children recognise that certain kinds of beliefs have distinct epistemic values. Moreover, these truth values change with evidence. This makes a fundamental part of scientific reasoning the understanding that hypotheses—the propositional beliefs used in scientific reasoning—are not simply true by virtue of being expressed, but are capable of being falsified through observing or generating data (e.g., Kuhn and Pearsall 2000). In this way, scientific reasoning requires the representational capacity inherent in understanding the epistemic nature of belief—to represent belief states and manipulate those belief states upon observation or interaction with the world. In order to engage in scientific reasoning—particularly hypothesis testing—one must be able to represent and reason across multiple hypotheses.

Understanding this representational structure is part of children's developing theory of mind. In particular, this representational capacity—recognising that there is a difference between viewing beliefs as a reflection of reality as opposed to viewing beliefs as a representation of reality that might not be veridical—has often been operationalised through measures of understanding others' false belief. Using a relatively global sample, Wellman et al. (2001) documented that children understand that others can have false beliefs about the world around age 4, indicating the emergence of a representational theory of mind (see also Perner 1991). We posit here that this representational capacity is a component of children's emerging scientific reasoning capacity. A representational theory of mind suggests that children can hold multiple potential belief states simultaneously and reason among them based on observed data. For example, in an unexpected contents task with a crayon box that contains candles, once one knows the container is deceptive, to reason

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *Infant and Child Development* published by John Wiley & Sons Ltd.

Highlights

- False belief understanding predicts early hypothesis testing capabilities in Turkish preschoolers.
- Performance on false belief tasks does not generalize to the design of unconfounded experiments.
- The association between Theory of Mind (ToM) and scientific reasoning appears multifaceted and contingent upon specific task demands.

appropriately, one must inhibit the representation of the situation (that there are candles in the box) to respond on the basis of a representation of another's knowledge (that the other person thinks there are crayons in the box). That is, although a successful reasoner believes that there are candles in the box, they respond that the other person will believe there are crayons. This ability to hold multiple belief states seems analogous to holding multiple hypotheses in mind in measures of scientific reasoning.

Scientific reasoning, however, has not often been investigated in preschoolers, making evidence in favour of a hypothesized relation between understanding others' false beliefs and children's scientific reasoning capacities sparse. Experimentation—the cornerstone of scientific reasoning—involves designing and conducting investigations to test different hypotheses. Typically, experimentation has been measured through children's understanding and use of the *control of variables strategy* (henceforth, CVS). CVS involves manipulating one variable while holding all other variables constant (e.g., Tschirgi 1980; Zimmerman 2007). Children in kindergarten and early elementary school often find it challenging to consistently apply CVS (e.g., Croker and Buchanan 2011; Chen and Klahr 1999), although some modifications can make these measures applicable to 4–7-year-olds (e.g., Sobel 2023; van der Graaf et al. 2015, 2016, 2018). What preschoolers can more clearly do is identify confounded experiments as less informative than unconfounded ones, and recognise the need to isolate variables to make appropriate causal conclusions when those interventions are presented to them (e.g., Lapidow and Walker 2020; Moeller et al. 2022; Shavlik et al. 2022). Young children appear capable of holding multiple hypotheses in mind and determining what data are necessary to discern among those possibilities. The underlying hypothesis guiding this investigation is that this ability is related to children's developing false belief.

The majority of research examining relations between scientific reasoning and theory of mind development focused on older children, rather than preschoolers. Astington et al. (2002) found that 7-year-olds' performance on a second-order theory of mind measure—representing that mental states have a recursive quality (Perner and Wimmer 1985)—related to performance on measures of understanding evidence, independent of other cognitive capacities. Similarly, Osterhaus et al. (2017) found that performance on similar second-order false belief measures—measures showing belief states are recursive—as well as interpretive theory of mind measures (e.g., Carpendale and Chandler 1996)—measures related to recognising that events can have multiple interpretations (what they refer to together as *Advanced Theory of Mind* or *AToM*)—related to 8- to 10-year-olds' beliefs about what science is (their understanding of the nature of science, following Lederman et al. 2002). Koerber and Osterhaus (2019) found that 6-year-olds' performance on AToM measures related to their scientific reasoning, using a measure they developed called the *Science-K*, which tests their understanding of experimentation, how data are interpreted related to a hypothesis, and similar nature of science questions (see also Lecce et al. 2025; Osterhaus and Koerber 2023; Weisberg and Sobel 2022, for similar findings relating AToM measures to children's scientific reasoning).

Success on the advanced theory of mind measures suggests that children appreciate that different individuals can hold different beliefs regarding the same evidence or event, such as distinguishing between the cause of an event and a person's reason for believing that cause occurred (Astington et al. 2002; Kyriakopoulou and Vosniadou 2020; Osterhaus and Koerber 2023). A 'reason' in this context is a belief that provides evidence for another belief—that is, that beliefs can lead to other beliefs and conclusions in an interpretive manner. Such a capacity appears to be more readily available to older children. For example, Eisbach (2004) showed that 9-year-olds, but not 5-year-olds, could recognise that two people could perceive the same stimulus (e.g., a picture of the ocean), and have the same thought about it (e.g., swimming is water is fun), but could then have thoughts that diverge (i.e., if one person next thought about fish, the other person did not necessarily have to have the same thought). This suggests that only older children understand that two people need not reach the same conclusion even though they see the same data (see also Flavell et al. 1995; Lagattuta and Wellman 2001).

Children's emerging understanding of false belief relates to their later success on more advanced theory of mind measures (e.g., Osterhaus and Koerber 2021; see also Lockl and Schneider 2007). Moreover, several studies suggest that the metacognitive abilities required for scientific reasoning originate in the representational capacity evident in performance on false belief tasks (e.g., Köksal-Tuncer and Sodian 2018; Köksal et al. 2021; Ruffman et al. 1993). Few studies, however, examine direct links between performance on false belief measures and measures of children's scientific reasoning. Having the representational capacity indicated by success on explicit false belief measures suggests that children have the ability to hold multiple representations of the same event (e.g., Gopnik and Rosati 2001; Doherty and Perner 2020; Perner 1991), which might be the capacity necessary to reason about the relation between evidence and alternative hypotheses. Indeed, Piekny et al. (2013) found that children who passed a first-order false belief task at age 4 were more likely to choose a conclusive over an inconclusive test for a hypothesis (following the *Big-Mouse/Little-Mouse* procedure developed by Sodian et al. 1991) when they were tested longitudinally at age 5.

The two studies presented here extend this finding by examining the relations between children's understanding of others' false belief and their scientific reasoning in a sample of Turkish preschoolers. Study 1 examines relations between false belief understanding and recognising the difference between confounded and unconfounded experiments that are presented to children to determine what data are necessary to observe to make particular kinds of causal conclusions (i.e., a

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cross-sectional reproduction of the Piekny et al. finding). Study 2 extends these findings by examining relations between the emergence of false belief understanding and children's ability to design unconfounded experiments to generate data for causal conclusions, as opposed to reasoning about experiments that are designed ahead of time. Study 1, therefore, tests whether there is a relation between false belief understanding and choosing an appropriate experiment to make an inference. Study 2 tests whether a similar relation exists for designing that experiment.

1 | Theory of Mind Development in Turkish-Speaking Children

Examining the relations between false belief understanding and scientific reasoning in Turkish children is theoretically significant, given Turkish linguistic characteristics that may relate to the development of theory of mind. The Turkish language includes explicit mental state verbs that denote false thoughts (*sanmak* and *zannetmek*) as opposed to a more epistemically neutral verb for think (*düşünmek*). Turkish also contains obligatory evidential markers that require speakers to indicate and track the source of their knowledge (Aksu-Koç et al. 2009; Aksu-Koç and Alici 2000; Granti 2004). Some suggest that exposure to such linguistic features facilitates children's understanding of false beliefs (Aksu-Koç et al. 2009; Lucas et al. 2013; Shatz et al. 2003). For example, in a combined sample of Turkish and Puerto Rican Spanish-speaking children (two languages with lexical items that distinguish "think" from "falsely think"), more than 70% of the 3- and 4-year-olds tested responded correctly on a false belief question when asked with the language-specific verb meaning "falsely think." This performance was superior to that of Brazilian Portuguese-speaking and American English-speaking preschoolers, two languages that lack lexical items that distinguish think from falsely think (Shatz et al. 2003). Even using the neutral verb "think" (*düşünmek*), Turkish preschoolers performed better than English- and Chinese-speaking children on an unexpected contents task (Lucas et al. 2013). A more recent study (Kandemirci et al. 2023) further compared the link among evidential markers, source monitoring, and false belief understanding in Turkish and English-speaking children and found that Turkish children performed better than a matched English-speaking sample. Further, children in Türkiye, like children in China and Iran, show a different pattern of results on the theory of mind scales than children in the United States; they pass the Knowledge Access task prior to the Diverse Beliefs task (e.g., Haskarca & Ilgaz, Haskaraca and Ilgaz 2023; Selcuk et al. 2018). Given these findings, we sought to examine the relations between the emergence of false belief understanding and facets of scientific reasoning in a Turkish-speaking sample.

At this point, we want to acknowledge that the unexpected contents task can ask two questions to children—one about another's false belief and one about their understanding of their own representational change. Some studies only evaluate responses to questions about others' false beliefs (e.g., theory of mind scales, Wellman and Liu 2004). However, other studies, including many cited above, combine responses to questions about others' false beliefs and children's own representational change into a single measure. Both Perner et al. (1987) and Gopnik and Astington (1988)—the two papers that originally described the unexpected contents task—suggest that these questions consider different capacities (see also Gopnik 1993). Perner et al. (1987), for example, argued that when assessing performance on an unexpected contents task, questions about understanding one's own representational change, although related to understanding others' false belief because similar cognitive control capacities are required, might be more of a measure of children's memory for a previous belief state as opposed to children's representational capacity. Sobel (2023) showed that performance on these questions follows different developmental trajectories, even when both questions are asked of the same child; both meta-analytic and mega-analytic investigations suggest that children use different cognitive capacities to respond to these questions.

In other findings, we (Saçkes et al. 2025) showed that preschoolers in Türkiye were more likely to respond correctly to questions about their own representational change in an unexpected contents task than appreciating another's false beliefs on the same task (see also Kaysılı and Acarlar 2011). This suggests that, in a Turkish sample, we should consider how children's responses to questions about others' false beliefs and their own representational change relate to their scientific reasoning independently. Following Perner et al. (1987), we hypothesize that performance on the false belief question might have more of a relation with children's scientific reasoning than performance on the representational change question, as the former relates more to the representational capacities necessary to hold multiple hypotheses in mind. The latter, however, shares many lower-level cognitive control capacities with the former. Thus, any unique variance explained by performance on the false belief question, independent of the representational change question, might indicate how children's emerging representational capacity relates to their scientific reasoning.

To this end, Study 1 reproduces (cross-sectionally) the Piekny et al. (2013) investigation, examining the relation between children's performance on false belief and representational change questions separately with performance on the Big Mouse/Little Mouse task originally described by Sodian et al. (1991). We hypothesize that any relation between performance on the unexpected contents task and scientific reasoning in a Turkish sample will be limited to the false belief question and not a representational change question (note that Piekny et al. (2013) asked only a false belief question, so we cannot determine whether such relations exist in other samples). Moreover, we hypothesize that even the youngest children in the sample of Turkish preschoolers will show a relation between their understanding of experimentation and their performance on false belief tasks.

The scientific reasoning measure used in Study 1, however, involves interventions generated by the experimenter that are described only to the child. Although this is similar to several studies of early scientific reasoning in which children select the appropriate intervention (e.g., Lapidow and Walker 2020; Moeller et al. 2022; Sodian et al. 1991), these procedures are structurally different from studies in which children generate their own interventions, which show more of a developmental trajectory between ages 4–6 (e.g., van der Graaf et al. 2015). Study 2 considers a similar relation between Turkish preschoolers' performance on a false belief (vs. representational change) question and their performance

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

on a modified version of the Ramps task used by van der Graaf et al. (2015). This latter study involves children using CVS to design their own unconfounded interventions in order to learn a causal structure. To our knowledge, no study has examined this relation in a sample of children this young. It is possible that the Turkish sample we collected will show a correlation similar to that hypothesized in Study 1. However, given that previous studies have shown subsequent development in these measures, it is also possible that children may not be able to use CVS at this age, which would prevent any correlation with false belief from being present.

Finally, we note that in both studies we examine relations between false belief and scientific reasoning, rather than any first-order theory of mind measure. This hypothesis posits that relations between responses to false belief questions and scientific reasoning will be present regardless of how children perform on representational change questions, which involve similar cognitive-control capacities (e.g., both involve inhibiting the actual state of the world in favour of a prepotent response). This hypothesis aligns with our finding that preschoolers in a Turkish sample differed in their performance on representational change and false belief questions.

2 | Study 1

2.1 | Participants

The final sample consisted of 110 children (M_{age} = 55.67 months, Range 40–71 months, 56 Girls and 54 Boys). Twenty-five children were 3 years old (M_{age} = 44.56 months), 47 children were 4 years old (M_{age} = 53.06 months), and 38 children were 5 years old (M_{age} = 66.21 months). Children were recruited from eight preschools serving predominantly middle-income families in an urban area in Türkiye. Sample size was determined by a power analysis for a binomial logistic regression analysis, assuming a two-tailed test with $\alpha = 0.05$, $\beta = 0.20$, and a medium effect size, Odds Ratio = 3.5. This power analysis indicated a required sample of 100 participants. However, all children whose families agreed at these preschools were tested. When the same analyses were run on only a subset of the first 100 children tested, the pattern of significance in the results reported below was identical; we report data from the entire sample. Eighteen additional children were tested but excluded from the final sample because they did not pass control questions (see below).

2.2 | Materials and Procedure

Children were tested individually in a quiet room at their preschool. Children received two measures—an unexpected contents task in which both a question about children's false belief and their representational change were measured, as well as a version of the Big Mouse/Little Mouse task that measured children's ability to determine an intervention for experimentation as opposed to satisfying a different goal.

2.2.1 | Unexpected Contents Task

We employed a version of the unexpected-contents task (Gopnik and Astington 1988; Perner et al. 1987), which had been modified and used in various studies with Turkish children (Kaysılı 2013). The researcher introduced a well-known chocolate box in Türkiye to the preschoolers and asked them to guess what was inside the box. If a child did not respond, the researcher used further prompts (e.g., make your best guess) to motivate the child to provide an answer. The researcher then revealed the contents of the box: three colouring pencils. We prompted the child to identify the box's contents if he or she did not spontaneously do so. The child was allowed to touch the box and its contents. The researcher then placed the colouring pencils inside the box and asked the child about their *representational change*; they were asked to reiterate his or her initial guess about the content of the box using the Turkish verb for falsely think (i.e., when you first saw the box, before you opened it, what did you falsely think was inside the box?).

The researcher then retrieved a plush toy character from a closed bag and stated that he/she (the toy) was sleeping inside the bag while talking. The researcher named the toy with the child. Using the given name of the toy, the researcher emphasised that the toy was sleeping inside the bag, had not seen or heard what happened, and had never looked inside the box. Then, the researcher posed two questions to the child: the *false belief* question (Now, if I ask the toy what is inside this box, what would she/he respond to me?) and a *memory* question (Had the toy seen inside the box?).

If children did not respond correctly on the memory question (15 children) or could not identify the contents of the box (5 children), they were not included in the final analysis (2 children responded incorrectly on both of these questions, meaning that 18 children were tested but not included in the analysis; note that including these children does not change the pattern of significance reported below).

2.2.2 | Experimentation Measure

We used the Big Mouse/Little Mouse task (Sodian et al. 1991) to assess children's domain-general experimentation skills. The researcher introduced two plush toys named Can and Ece to children. The researcher explained that a mouse was living in the basement of Can and Ece's house. Can and Ece could not decide whether the mouse was big or small, and they were curious to find the answer. The researcher then showed a picture of two mouse houses and explained that Can and Ece made two, one with a big door and one with a small one. The researcher emphasized that a big mouse can only enter the mouse house with a big door, and a small mouse can enter both houses. After ensuring that the child understood the information, the researcher asked the children two questions and additional justification questions: the feed question and the find out question. In the feed question, we told the children that Can and Ece wanted to feed the mouse in their basement. We asked the children, "In which house should Can and Ece put the food to make sure that the mouse gets the food?" In the find-out question, we told the children that Can and Ece wanted to find out whether the mouse in their basement was big or small. The children were asked, "In which house do you think Can and Ece should put food to determine if the mouse in their basement is a big or small mouse?" Immediately after each question type, we prompted children to

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

provide rationales for their responses. Children who responded correctly to the feed and find-out questions passed the task. This parallels the scoring used by Piekny et al. (2013) and Sodian et al. (1991).

3 | Results and Discussion

Out of the 110 children, 32 children responded correctly to the false belief question (29%) and 91 children in the sample (83%) responded correctly to the representational change question. Preliminary analysis showed that boys and girls did not differ in their response to either question, both $\chi^2(1)$-values <1.82 both p-values >0.17. A GLMM assuming a binomial logistic response with age (centered) and question type as fixed factors and participant ID as a random factor revealed only a main effect of question type, with superior performance on the representational change question than the false belief question, $\beta=8.64$, SE $=0.88$, $t=9.77$, $p<0.001$, Odds Ratio >100. Overall, children were above chance on their performance on the representational change question, Binomial test, $p<0.001$, and below chance on their performance on the false belief question, Binomial test, $p<0.001$, suggesting that responses on these two questions had different developmental trajectories.

Thirty-eight children in the sample (35%) passed the Experimentation measure by responding correctly on both the Feed and Find Out questions. This indicates that they not only selected the right intervention to find out information, but they appreciated that the different interventions were useful for different goals (i.e., they did not always choose the house with the smaller opening). Boys and girls did not differ in their response to these questions, $\chi^2(1)=2.15$, $p=0.14$. Performance on this measure overall was greater than chance performance (25%), Binomial test, $p=0.02$, and marginally correlated with age, $r_s(108)=0.18$, $p=0.07$, due to 5-year-olds (50%) outperforming the combined performance of the 3-year-olds and 4-year-olds (28% and 26% accurate respectively), $\chi^2(1)=6.13$, $p=0.01$. This age-related trend in experimental reasoning is consistent with research highlighting the gradual development of scientific thinking skills throughout early childhood (Piekny et al. 2013, 2014; Sodian et al. 1991).

We performed a binary logistic regression using a hierarchical design on whether children succeeded on the experimentation measure. In the first model, we entered children's age (in months) and performance on the representational change question. The second model then added performance on the false belief question. This second step resulted in explaining significantly more variance, $\chi^2(1)=4.74$, $p=0.03$, and the final model was significant, $\chi^2(3)=8.97$, $p=0.03$. Performance on the false belief measure predicted a unique amount of variance in this final model, $B=1.00$, SE $=0.46$, Wald $\chi^2(1)=4.71$, $p=0.03$, Odds Ratio $=2.71$. Neither age, Wald $\chi^2(1)=2.19$, $p=0.14$, nor performance on the representational change measure, Wald $\chi^2(1)=0.01$, $p=0.98$, predicted a significant amount of variance.

We supplemented this analysis by coding how children explained their choices on the feed and find-out questions by scoring whether children justified their response in a manner that elicited their understanding of why their choice would result in the desired outcome. To examine the reliability of the scoring, all responses to the questions targeting children's rationale for the feed and find out questions were scored independently by two researchers. For the feed question, responses that indicated both mice could fit through the large door were marked as correct. (e. g, "If it's small, it can pass, if it's big, it can still pass", "If there is a small door, the big mouse cannot enter, they both enter this [big door] one"). The responses to the 'find-out' question were marked as correct if they indicated that only the small mouse could enter the small door. (e.g., "because the door is too small for a big mouse to get in", "The big mouse cannot pass through the small door. If the mouse is small, it eats the food"). The inter-rater agreements between the two coders were high (Feed: $\kappa=0.903$; 95% CI $=0.81$–0.98; Find-out: $\kappa=0.897$; 95% CI $=0.79$–0.98), indicating high inter-rater agreements between the two coders. Overall, only 15 children both made the right choice on both the feed and find-out questions as well as justified their response in terms of understanding the rationale for their responses in terms of scientific reasoning (14% of the sample). This finding aligns with previous studies (Piekny and Maehler 2013; Piekny et al. 2014) where a limited percentage of children provided correct rationales for the feed and find-out questions.

We considered the subset of the sample who justified the experimentation questions appropriately with the same binary logistic regression model. The initial model with age and performance on the representational change measure was not significant. Adding the false belief question showed a trend towards a better model, $\chi^2(1)=2.83$, $p=0.09$. In this final model, neither age nor performance on the representational change measure predicted whether children could explain the logic behind the experimentation and feed questions, both Wald $\chi^2(1)$-values <0.77, both p-values >0.38. Performance on the false belief question was a marginally significant predictor of children's performance, $B=1.00$, SE $=0.59$, Wald $\chi^2(1)=2.84$, $p=0.09$. These data suggest that although children may begin to grasp the difference between producing an effect and testing a hypothesis, their ability to articulate their understanding of the underlying logic of experimentation is still developing. There is, however, some evidence that such explanatory ability is related to success on the false belief question.

The present study showed that responses to a false belief question, but not a representational change question on an unexpected contents task related to children's scientific reasoning—particularly their ability to discern between confounded and unconfounded interventions—in a sample of Turkish preschoolers. This works expands on a broader literature suggesting relations between children's developing theory of mind and metacognitive reasoning abilities and their scientific reasoning more generally (e.g., Astington et al. 2002; Klein 1998; Koerber and Osterhaus 2019; Ruffman et al. 1993; Sodian et al. 2016), and suggests the possibility that early theory of mind mastery—particularly false belief understanding—serves as a foundation for determine what data are necessary to come to an appropriate conclusion. Of particular importance is that only the false belief question and not the representational change question was relevant to children's experimentation. This suggests that it is the representational capacity inherent understanding others' false belief and not the memory capacity

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in remembering a past mental state that is necessary for inferences about experimentation.

In Study 1, the measure of experimentation involved children choosing between confounded and unconfounded actions to accomplish a goal. The interventions, however, were designed for the child, who was just required to select the appropriate option. When we considered children's justifications for their responses, relatively few of the children tested appropriately explained their choice in terms of scientific reasoning. This suggests that although children's understanding of false belief might be foundational for scientific reasoning, even children in our Turkish sample, who might have a different developmental trajectory for understanding others' false belief and their own representational change, might not possess the ability to design their own interventions to use the control of variables strategy in the preschool years. We investigated this possibility in Study 2.

4 | Study 2

Study 2 used the same design as Study 1, but instead of presenting children with a CVS measure in which the interventions were designed for and told to the child, children were asked to generate interventions to test hypotheses. Performance on false belief and representational change questions on the unexpected contents task were again considered in relation to performance on a version of the Ramps task (Chen and Klahr 1999) adapted by van der Graaf et al. (2015). Note that they found that 5- and 6-year-olds outperformed 4-year-olds in their capacity to use CVS when designing interventions on their measure. Four-year-olds' performance in van der Graaf et al.'s study was below chance. Study 2 investigated whether these findings were similar in a sample of Turkish preschoolers as well as whether they related to performance on the false belief (but not representational change) question on the unexpected contents task.

In the present version of the ramps task, children were introduced to the measure 1 week prior to the test session. This allowed us to consider an auxiliary hypothesis, which concerns how children were introduced to the measure and the stimuli. Half of the children we tested were given a structured introduction to the ramps measure and were guided through the procedure, whereas the other half of the sample (randomly determined) was allowed simply to play freely with the apparatus. van der Graaf et al. (2015, 2016, 2018) have shown 5 and 6-year-olds' can use CVS when they are given guided feedback to scaffold their performance. In addition to considering relations between this measure and performance on the unexpected contents task, the present study examines whether introducing children to this structure facilitates their reasoning.

4.1 | Participants

The final sample consisted of 76, 4 and 5-year-olds (M_{age} = 60.64 months, SD = 6.97, Range = 48 to 70 months). Thirty-six children were 4 years old; 40 children were 5 years old. Children were recruited from nine preschools serving predominantly middle-income families in an urban area in Türkiye. Sample size was based on a one-tailed power analysis, assuming $\alpha = 0.05$, $\beta = 0.20$, and a medium to large effect size (OR = 3.5) for the analysis comparing the play conditions (described below). This analysis yielded a necessary sample of 78 children. All children available at the included preschoolers were tested, but 4 children who were tested were not included in the final analysis because they failed the inclusion measure (the same measure as Study 1, see below).

4.2 | Materials and Procedure

All children were administered the unexpected contents task, using the same materials with the same procedure as Study 1. Four children did not respond correctly on the control question in this task, so their data were not included in the final analysis.

On the day children were administered the unexpected contents measure, they were also introduced to the stimuli used in the ramp task assessment. Children were randomly assigned to be introduced to these stimuli in a free play condition or a guided play condition. Forty children were assigned to each group initially, but the four children who were not included in the final analysis all came from the guided play condition, meaning there were only 36 children in this group. These conditions are described below:

4.2.1 | Free Play Condition

In the free play condition, the ramp task materials (two toy cars, two wooden ramps, and five plastic blocks) were introduced to the children. The researcher set the play environment by stacking two sets of blocks on top of each other to create ramps. Children were encouraged to explore and compare how far a small and a large car would travel when released down the ramp. Would they travel the same distance, or would one travel farther? The researcher encouraged children to test their ideas using the materials. Children were allowed to play with the ramps independently, using the cars and other materials available in the classroom and without adult intervention. We provided no explicit instruction, probe questions, or guidance to the children in the free-play condition. Children were allowed to play for up to 50 min.

4.2.2 | Guided Play Condition

The same materials as in the Free Play condition were used. As with the free play condition, the researcher set up the play environment by stacking two sets of blocks on top of each other to create ramps. Children were encouraged to explore and compare how far a small and a large car would travel when released down the ramp. Would they travel the same distance, or would one travel farther? The researcher encouraged children to test their ideas using the materials. Unlike the free play condition, children in the guided play condition were provided explicit probing questions, information, and guidance during their play with the materials. Children were encouraged to explore the relationship between changing the ramp's angle and the distance the cars travelled. They were encouraged to observe that different cars and objects can travel different distances depending on

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the type and slope of the ramp, the type of cars, and the type of release (i.e., a push vs. a release) of the cars (or objects from the top of the ramp). The researcher asked the children to compare two ramps: a narrow ramp with a smooth surface and a wider ramp with a rough surface (i.e., a wider ramp covered with thick fabric). The children were encouraged to explore which cars and objects could move down each ramp (e.g., the small car fit both ramps, and the large car only fit the wide ramp). Children were allowed to play for up to 50 min with the cars, ramps, and other materials in the classroom.

4.2.3 | Ramp Task Assessment

A week after this play intervention, we administered the ramp task individually to each child in a quiet room located in their preschool. This task measured whether children could use CVS to design an unconfounded experiment to answer a research question. The materials used in the ramp task included the following: (1) Two toy cars with the same colour and model, one small (less than 5 cm wide) and one large (6–8 cm wide). (2) Two wooden ramps of 1 m in length each. The ramps were 5 cm and 8 cm wide, respectively. We covered the surface of the wider ramp (8 cm) with a thick fabric, which increased friction. (3) Five plastic blocks to vary the heights and inclines of the ramps.

The assessment started with a warm-up talk, during which the researcher reminded the children that they had played with ramps and cars the previous week and explained that the researcher had brought the ramps and cars for more play today. The researcher initially probed the children's predictions about the distances the different cars would travel when released from the top of the ramp, and the researcher asked the children to explain their predictions. Next, the researcher asked the children to test their predictions using the materials. We observed the children during experimentation and recorded field notes of their actions and utterances. The researcher asked the children to describe the outcome of their experiments by responding to the following questions: "Which of these cars did you predict would go the furthest before the experiment?" "Which car went the furthest?" "Why do you think the small/large car went farther?" We recorded the responses of the children in a field notebook. We measured whether children changed a single variable (i.e., used CVS by changing only the car's size, the ramp's slope, the ramp, type of the ramp, or whether the car was pushed or released) and kept the other variables constant while testing their predictions as opposed to whether children changed all variables or kept one variable constant while changing the remaining variables.

5 | Results and Discussion

We first examined the variations in children's performance on the unexpected contents measure. Twenty-seven (out of 76, 36%) children responded correctly to the false-belief task. In contrast, 73 children (96%) responded correctly to the representational change question. Girls and boys did not respond differently to either question, both $\chi^2(1)$-values <0.35, both p-values >0.55. We ran a GLMM assuming a binomial logistic distribution with response on these two questions as the dependent variable, with age (centered) and question type as independent variables. This analysis revealed only a main effect of question type, with superior performance on the representational change question than the false belief question, $\beta=2.41$, SE $=0.88$, $t=2.75$, $p=0.006$, Odds Ratio $=15.64$. Similar to the previous study, the difference in performance between these questions suggests that they have distinct developmental trajectories in this sample.

Next, we examined the children's use of the control of variables strategy, as measured by their ramp task performance. Only six children passed the ramp task (8%). All six were 5-year-olds, and although there was a significant correlation between children's use of CVS and age, $r_s(74)=0.31$, $p=0.006$, this correlation may reflect a floor effect. Four of the children who used CVS were in the free play condition and two were in the guided play condition. This was not a significant difference, Fisher Exact Test, $p=0.31$.[1] The correlation between performance on the false belief measure and children's use of CVS was marginally significant, $r_s(74)=0.19$, $p=0.09$. No such relation with the representational change measure was present, $r_s(74)=0.06$, $p=0.61$. To isolate any unique effect of these variables, we performed a binary logistic regression with CVS usage as the dependent variable, with age, performance on the representational change and false belief measures, and play condition as the independent variables. The overall model was significant, $\chi^2(4)=12.33$, $p=0.02$. Only age explained any unique variance, $B=0.26$, SE $=0.13$, Wald $\chi^2(1)=4.10$, $p=0.04$, Odds Ratio $=1.30$. Performance on the false belief measure was not significant in this model, Wald $\chi^2(1)=1.68$, $p=0.20$.

This difference in findings between Study 1 and the Big Mouse/Little Mouse task may be attributed to the greater complexity of the Ramps task. The ramps task required children to consider multiple variables and manipulate them systematically, whereas the mouse task presented a more straightforward experimental scenario with fewer variables to manage. The limited success rate observed here, even among older preschool children, highlights the complexity of this scientific reasoning skill and the challenges young children face in isolating and controlling variables (Tschirgi 1980; Zimmerman 2007). In a previous study using the same ramp task, children performed better than the sample in the present study. This discrepancy may be attributable to the sample composition in van der Graaf et al.'s (2015) study, which recruited slightly older children from kindergartens with a science and technology-focused curriculum. In contrast, the current sample was recruited from preschools in which science and technology were not integral to the curriculum, and the children were slightly younger. These differences in sample characteristics may account for the observed floor effect and the resulting nonsignificant result. Research suggests that kindergarten and early elementary children often struggle to consistently apply CVS (Chen and Klahr 1999; Siegler and Chen 1998). Although younger children may perceive confounded experiments as less informative, their ability to design unconfounded experiments, particularly with multiple variables, develops gradually throughout childhood (e.g., Croker and Buchanan 2011; Köksal-Tuncer and Sodian 2018; Shavlik et al. 2022).

6 | General Discussion

A key element of scientific reasoning is the ability to design and interpret well-controlled experiments to test hypotheses

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Zimmerman 2007). The ability to reason about relations between hypotheses and evidence emerges in early childhood, particularly in the form of selecting unconfounded interventions as more informative than confounded ones (e.g., Koerber et al. 2005; Moeller et al. 2022; Piekny and Maehler 2013; Ruffman et al. 1993; Sodian et al. 1991). Moreover, the design and selection of such interventions appear to be related to children's developing theory of mind, specifically their understanding of belief (e.g., Astington et al. 2002; Klein 1998; Koerber and Osterhaus 2019). The present study aimed to reproduce the previously reported association between false belief understanding and scientific reasoning with a sample of younger children who speak a language where performance on questions asking about others' false belief and their own representational change might differ. In both of our studies, as well as in a previous investigation (Saçkes et al. 2025), we found that Turkish children were better at tracking their own representational change on an unexpected contents measure than reasoning about another's false belief. The question we asked here was whether these children's understanding of false belief specifically related to their ability to select and design unconfounded interventions over confounded ones.

Study 1 found that Turkish preschoolers' performance on a measure of understanding others' false belief, but not performance on a measure of their own representational change, related to their ability to determine the difference between confounded and unconfounded interventions. Study 2, however, did not find that this relation held for a more complex measure of scientific reasoning, in which children had to design their own unconfounded interventions, as opposed to selecting an unconfounded intervention over a confounded one. These findings suggest that false belief understanding alone is insufficient for children to use CVS. Moreover, in Study 2, children in both play conditions showed similar low levels of CVS usage, indicating that preschoolers struggled to discover and acquire the necessary experimental skills for an unconfounded test. The explicit guidance in the guided play condition was insufficient in helping preschoolers develop CVS beyond the free play condition. The free play condition may have placed too much demand on children's cognitive processing, while the scaffolding in the guided play condition did not adequately reduce cognitive load.

Although Study 1 identified a positive link between false-belief understanding and differentiating between generating an effect and testing a hypothesis, the absence of a similar relation in Study 2 suggests that different scientific reasoning skills may have distinct developmental trajectories and be influenced by ToM development (and specifically false belief development) to different degrees. Success on a false belief measure might be indicative of the representational capacity necessary to hold multiple hypotheses in mind, so that the child could reason about which action could discern among these representations, but further capacities might be necessary to design those interventions themselves. This finding aligns with a broader literature where some studies have found that specific aspects of advanced theory of mind, such as second- or third-order false belief understanding, are more strongly linked to specific scientific reasoning skills, such as experimental design (e.g., Koerber and Osterhaus 2019; Osterhaus et al. 2017). The type of experimental reasoning assessed by the Big Mouse/Little Mouse task involves recognising which action is needed to change the epistemic value of a state (from uncertainty to certainty) when the interventions are presented to the child. In contrast, the type of reasoning assessed by the Ramps task builds on this reasoning capacity to require children to design those unconfounded interventions. An analogy is the difference between asking the child whether another would have a false belief when they observe that the container is deceptive, and having the child create a deceptive container independently. The former clearly develops during preschool, whereas our intuition is that the latter would be more difficult for preschoolers without explicit instruction on the nature of the deception (see also Lapidow and Walker 2022). These data suggest that the representational capacities indicated by success on false belief measures might serve as a foundation for later scientific reasoning.

We want to highlight two additional facets of these findings. First, neither study found a relation between children's scientific reasoning and sex; girls and boys showed no difference in scientific reasoning in either study. These findings are consistent with the existing literature, which suggests no significant difference between boys and girls in their science achievement prior to elementary school, although there are age-related trends favoring older children (e.g., Curran and Kellogg 2016; Kinzie et al. 2014; Morgan et al. 2016; Quinn and Cooc 2015; Saçkes et al. 2011, 2013, 2020; Spelke 2005).

Second, these two studies examined more domain-general scientific reasoning abilities in children, rather than their more domain-specific scientific knowledge. Children develop substantial domain-specific scientific content knowledge during the preschool and early elementary school years. For instance, in earth and space science, research has explored children's understanding of the day and night cycle (Vosniadou 1994), mechanisms of rainfall (Saçkes et al. 2010), the shape of the Earth (Vosniadou and Brewer 1992), and the phases of the Moon (Hobson et al. 2010). In the physical sciences, children develop an understanding of the properties of matter (Krnel et al. 2005), sound (Mazens and Lautrey 2003), and density, particularly related to sinking and floating (Kohn 1993). In the life sciences, children develop distinct concepts of what it means to be alive (Laurendeau and Pinard 1962; Opfer and Siegler 2004), differential properties of animals and plants (Carey 1985), and biological processes like growth, healing, and heredity (Backscheider et al. 1993; Johnson and Solomon 1997). The acquisition of these concepts typically responds to instructional interventions (e.g., Yoon and Onchwari 2006). Together, these studies indicate that children develop explanations of the natural world at different rates, and that at any given point in development, children's concepts may differ from scientifically accepted models (e.g., Gelman and Wellman 1991; Schult and Wellman 1997).

Indeed, some studies on the acquisition of scientific reasoning and its relation to education emphasise the acquisition of domain-specific knowledge (e.g., Dunbar and Fugelsang 2005; Dündar-Coecke and Tolmie 2020; Klahr et al. 2011). However, the acquisition of this knowledge is not necessarily related to the representational capacities inherent in succeeding on a measure of false belief. What both types of measures share, however, are verbal demands and other general cognitive capacities, and we do not see a reason for there to be a unique relation between

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

children's understanding of others' false belief and their content knowledge. In another study (described in the supplemental materials section), we measured 3- and 4-year-olds' understanding of basic science concepts via the Science Concepts and Skills Assessment (Samarapungavan et al. 2009; Senocak et al. 2013). Children's understanding of others' false belief, their own representational change, and their age (whether they were 3 or 4 years old) all uniquely predicted performance on this scientific knowledge measure, suggesting that children's broader cognitive capacities might potentially relate to their acquisition of domain-specific knowledge.

6.1 | Limitations and Further Research

The linguistic factors present in Turkish, such as the presence of the obligatory evidential system and different lexical items indicating the epistemic state of a belief, have developmental trajectories—in some cases beyond the preschool years. As Turkish children's linguistic abilities mature, they might become more likely to use these features of their language to support their reasoning. Future work should examine older children to see whether relations between theory of mind development and scientific reasoning are task-specific. For example, Study 2 showed limited success in children's use of CVS, with only a small percentage demonstrating this skill. The intervention did not significantly improve CVS use, and no direct relation was found between false-belief understanding and CVS performance, highlighting the complexity of CVS development in young children and the need for more research into effective methods for promoting scientific reasoning skills in early childhood. Investigating an older sample could determine whether more advanced theory of mind measures are related to the emergence of CVS in a Turkish sample, and whether such advanced measures relate to either performance on false belief or representational change questions independently.

In addition to considering future work on older children, what is absent in the present study is a cross-cultural comparison to a matched sample in either the United States or Europe, particularly testing a sample of children in a language that does not have either the evidentiary markers or the distinction between the lexical items 'think' and 'falsely think'. It would also be interesting to consider samples from countries like China or Iran, which show a similar "collectivist acquisition pattern" (Selcuk et al. 2018, p. e2098) for theory of mind development to Turkish children (e.g., Shahaeian et al. 2011; Wellman et al. 2006), to see whether children in these cultures show similar relations as documented here.

The two studies presented here relied heavily on single measures to assess false belief and representational change understanding, as well as each facet of scientific reasoning. Only one theory of mind measure, the unexpected contents task, was used, which might not capture the full range of children's social cognitive understanding. Although a strength of this method is that the representational change and false belief questions can be asked of the same participant and share similar demand characteristics, they yield only a single binary response. It would be important to replicate these findings using a battery of false belief measures, including not only unexpected contents tasks but also unexpected transfer or appearance/reality procedures in which a false belief question can be posed. Likewise, the Big Mouse/Little Mouse measure (Study 1) and the Ramps task (Study 2) represent only one aspect of scientific thinking. Using multiple measures would provide a more comprehensive picture of children's scientific reasoning skills. Furthermore, general cognitive ability, language skills, and executive functioning skills correlate with both false belief and scientific thinking skills, but were not considered in the current studies. Although the false belief and representational change questions share similar language and cognitive abilities, they offer useful comparisons. Future studies should examine how performance on inhibitory control and working memory tasks relate to Turkish children's false belief and scientific reasoning.

The present studies also primarily focused on a narrow age range of preschool children (ages 3 to 5 years). This limitation limits the understanding of how the relation between false belief understanding and scientific thinking evolves over time among Turkish children. Longitudinal studies encompassing a wider age range would provide a more complete picture of the developmental trajectory of these skills. Addressing these limitations in future research would strengthen the understanding of the complex interplay between false belief understanding and scientific thinking in young children. Research with more diverse samples, the use of multiple, more sensitive measurement tools, the development and use of more targeted interventions, and the incorporation of cross-cultural perspectives will contribute to a more nuanced and comprehensive understanding of this crucial area of child development.

7 | Conclusion

The present studies suggest a relation between the emergence of representational capacity, as indicated by understanding others' false beliefs and appreciating what data are necessary (i.e., the difference between a confounded and unconfounded intervention) to make a causal conclusion in a sample of Turkish preschoolers. The relation is nuanced and task-specific: we observed significant relations for false belief understanding but not for understanding one's own representational change, and only for children's understanding of experimentation, not for a more complex measure of designing interventions to control variables.

Finally, the differences in performance on false belief and representational change questions across the two studies parallel other findings (Saçkes et al. 2025; Kaysılı and Acarlar 2011). This suggests the possibility that other relations with the development of representational theory of mind capacities, usually indicated by performance on the false belief task, might want to be considered in terms of the different capacities presented by a false belief and representational change question, as opposed to lumping performance on these two measures together. Although this conclusion may apply only to a Turkish sample, the differences in developmental trajectories between false belief and representational change described by Sobel (2023) suggest that there might be more global differences between these two measures. That said, these data suggest that children in Türkiye exhibit a different developmental trajectory in theory of mind than children in other

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cultures (see also Selcuk et al. 2018). An open question is how that potentially different developmental trajectory relates to other facets of cognitive development thought to be related to theory of mind.

Author Contributions

Mesut Saçkes: conceptualization, writing – original draft, methodology, investigation, writing – review and editing, formal analysis. **Sonnur Işıtan:** investigation, data curation, writing – review and editing. **Sinem Güçhan Özgül:** investigation, data curation, writing – review and editing. **Kerem Avci:** investigation, data curation. **Kathy Cabe Trundle:** conceptualization, writing – original draft, writing – review and editing. **David M. Sobel:** conceptualization, writing – original draft, writing – review and editing, formal analysis.

Acknowledgements

We would like to thank children, parents, teachers, and directors of preschools that took part in our study.

Ethics Statement

The study was approved by the Balikesir University Social Sciences and Humanities Ethics Commission (Approval No 2021/02).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/yfz7b/?view_only=bbc4c9c270654e42a7e878c0d305e8b3.

Endnotes

[1] Note that for Fisher Exact Tests reported here, one-tailed tests were used, as there were clear hypotheses regarding the advantage of guided play over free play and subsequently, the advantage of passing measures over not passing them.

References

Aksu-Koç, A., and D. M. Alici. 2000. "Understanding Sources of Beliefs and Marking of Uncertainty: The Child's Theory of Evidentiality." In *The Proceedings of the Thirtieth Annual Child Language Research Forum*, edited by E. V. Clark, 123–130. Center for the Study of Language and Information.

Aksu-Koç, A., H. Ögel-Balaban, and I. E. Alp. 2009. "Evidentials and Source Knowledge in Turkish." In *Evidentiality: A Window Into Language and Cognitive Development, New Directions for Child and Adolescent Development, 125*, edited by S. A. Fitneva and T. Matsui, 13–28. Jossey-Bass. https://doi.org/10.1002/cd.247.

Astington, J. W., J. Pelletier, and B. Homer. 2002. "Theory of Mind and Epistemological Development: The Relation Between Children's Second-Order False-Belief Understanding and Their Ability to Reason About Evidence." *New Ideas in Psychology* 20, no. 2–3: 131–144.

Backscheider, A. G., M. Shatz, and S. A. Gelman. 1993. "Preschoolers' Ability to Distinguish Living Kinds as a Function of Regrowth." *Child Development* 64, no. 4: 1242–1257. https://doi.org/10.2307/1131337.

Carey, S. 1985. *Conceptual Change in Childhood*. MIT Press.

Carpendale, J. I., and M. J. Chandler. 1996. "On the Distinction Between False Belief Understanding and Subscribing to an Interpretive Theory of Mind." *Child Development* 67, no. 4: 1686–1706. https://doi.org/10.2307/1131725.

Chen, Z., and D. Klahr. 1999. "All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy." *Child Development* 70, no. 5: 1098–1120. https://doi.org/10.1111/1467-8624.00081.

Croker, S., and H. Buchanan. 2011. "Scientific Reasoning in a Real-World Context: The Effect of Prior Belief and Outcome on Children's Hypothesis-Testing Strategies." *British Journal of Developmental Psychology* 29, no. 3: 409–424. https://doi.org/10.1348/026151010X496906.

Curran, F. C., and A. T. Kellogg. 2016. "Understanding Science Achievement Gaps by Race/Ethnicity and Gender in Kindergarten and First Grade." *Educational Researcher* 45, no. 5: 273–282.

Doherty, M. J., and J. Perner. 2020. "Mental Files: Developmental Integration of Dual Naming and Theory of Mind." *Developmental Review* 56: 100909. https://doi.org/10.1016/j.dr.2020.100909.

Dunbar, K. N., and J. A. Fugelsang. 2005. "Scientific Thinking and Reasoning." In *The Cambridge Handbook of Thinking and Reasoning*, edited by K. J. Holyoak and R. G. Morrison, 702–725. Cambridge University Press.

Dündar-Coecke, S., and A. Tolmie. 2020. "Nonverbal Ability and Scientific Vocabulary Predict Children's Causal Reasoning in Science Better Than Generic Language." *Mind, Brain, and Education* 14, no. 2: 130–138. https://doi.org/10.1111/mbe.12226.

Eisbach, A. O. 2004. "Children's Developing Awareness of Diversity in People's Trains of Thought." *Child Development* 75, no. 6: 1694–1707. https://doi.org/10.1111/j.1467-8624.2004.00810.x.

Flavell, J. H., F. L. Green, E. R. Flavell, P. L. Harris, and J. W. Astington. 1995. "Young Children's Knowledge About Thinking." *Monographs of the Society for Research in Child Development* 60, no. 1: i–113.

Gelman, S. A., and H. M. Wellman. 1991. "Insides and Essence: Early Understandings of the Non-Obvious." *Cognition* 38, no. 3: 213–244. https://doi.org/10.1016/0010-0277(91)90007-Q.

Gopnik, A. 1993. "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality." *Behavioral and Brain Sciences* 16, no. 1: 1–14. https://doi.org/10.1017/S0140525X00028636.

Gopnik, A., and J. W. Astington. 1988. "Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction." *Child Development* 59, no. 1: 26–37. https://doi.org/10.2307/1130386.

Gopnik, A., and A. Rosati. 2001. "Duck or Rabbit? Reversing Ambiguous Figures and Understanding Ambiguous Representations." *Developmental Science* 4, no. 2: 175–183. https://doi.org/10.1111/1467-7687.00163.

Granti, T. 2004. "The Relationship Between Metacognitive Vocabulary and Theory of Mind Development (Publication Number 39001102276287)." [Master of Arts, Boğaziçi University] İstanbul.

Haskaraca, F. N., and H. Ilgaz. 2023. "Culturally Constituted Universals: Evidential Basis of Belief Matters." *Developmental Science* 00: e13398. https://doi.org/10.1111/desc.13398.

Hobson, S. M., K. C. Trundle, and M. Sackes. 2010. "Using a Planetarium Software Program to Promote Conceptual Change With Young Children." *Journal of Science Education and Technology* 19, no. 2: 165–176.

Johnson, S. C., and G. E. Solomon. 1997. "Why Dogs Have Puppies and Cats Have Kittens: The Role of Birth in Young Children's Understanding of Biological Origins." *Child Development* 68, no. 3: 404–419.

Kandemirci, B., A. Theakston, D. Boeg Thomsen, and S. Brandt. 2023. "Does Evidentiality Support Source Monitoring and False Belief Understanding? A Cross-Linguistic Study With Turkish- and

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

English-Speaking Children." *Child Development* 94: 889–904. https://doi.org/10.1111/cdev.13905.

Kaysılı, B. K. 2013. "Zihin Kuramı: Otizm Spektrum Bozukluğu Olan Ve Normal Gelişen Çocukların Performanslarının Karşılaştırılması." *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi* 14, no. 1: 83–103.

Kaysılı, B. K., and F. Acarlar. 2011. "Zihin Kuramının 3-5 Yaşları Arasındaki Çocuklarda Gelişiminin Yanlış Inanç Performansına Göre Incelenmesi." *Kuram Ve Uygulamada Eğitim Bilimleri* 11, no. 4: 1809–1826.

Kinzie, M. B., J. V. Whittaker, A. P. Williford, et al. 2014. "MyTeachingPartner-Math/Science Pre-Kindergarten Curricula and Teacher Supports: Associations With Children's Mathematics and Science Learning." *Early Childhood Research Quarterly* 29, no. 4: 586–599.

Klahr, D., C. Zimmerman, and J. Jirout. 2011. "Educational Interventions to Advance Children's Scientific Thinking." *Science* 333, no. 6045: 971–975. https://doi.org/10.1126/science.1204528.

Klein, P. D. 1998. "The Role of Children's Theory of Mind in Science Experimentation." *Journal of Experimental Education* 66, no. 2: 101–124. https://doi.org/10.1080/00220979809601398.

Koerber, S., and C. Osterhaus. 2019. "Individual Differences in Early Scientific Thinking: Assessment, Cognitive Influences, and Their Relevance for Science Learning." *Journal of Cognition and Development* 20, no. 4: 510–533.

Koerber, S., B. Sodian, C. Thoermer, and U. Nett. 2005. "Scientific Reasoning in Young Children: Preschoolers' Ability to Evaluate Covariation Evidence." *Swiss Journal of Psychology* 64, no. 3: 141–152.

Kohn, A. S. 1993. "Preschoolers' Reasoning About Density: Will It Float?" *Child Development* 64: 1637–1650.

Köksal, Ö., B. Sodian, and C. H. Legare. 2021. "Young Children's Metacognitive Awareness of Confounded Evidence." *Journal of Experimental Child Psychology* 205: 105080.

Köksal-Tuncer, Ö., and B. Sodian. 2018. "The Development of Scientific Reasoning: Hypothesis Testing and Argumentation From Evidence in Young Children." *Cognitive Development* 48: 135–145. https://doi.org/10.1016/j.cogdev.2018.06.011.

Krnel, D., R. Watson, and S. A. Glažar. 2005. "The Development of the Concept of 'Matter': A Cross-Age Study of How Children Describe Materials." *International Journal of Science Education* 27, no. 3: 367–383.

Kuhn, D. 1989. "Children and Adults as Intuitive Scientists." *Psychological Review* 96, no. 4: 674.

Kuhn, D. 2010. "What Is Scientific Thinking and How Does It Develop?" In *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, 497–523. Blackwell Publishing. https://doi.org/10.1002/9781444325485.ch19.

Kuhn, D., and S. Pearsall. 2000. "Developmental Origins of Scientific Thinking." *Journal of Cognition and Development* 1, no. 1: 113–129.

Kyriakopoulou, N., and S. Vosniadou. 2020. "Theory of Mind, Personal Epistemology, and Science Learning: Exploring Common Conceptual Components." *Frontiers in Psychology* 11: 531223.

Lagattuta, K. H., and H. M. Wellman. 2001. "Thinking About the Past: Early Knowledge About Links Between Prior Experience, Thinking, and Emotion." *Child Development* 72, no. 1: 82–102. https://doi.org/10.1111/1467-8624.00267.

Lapidow, E., and C. M. Walker. 2020. "Informative Experiments in Intuitive Science: Children Select and Learn From Their Own Causal Interventions." *Cognition* 201: 104315.

Lapidow, E., and C. M. Walker. 2022. "Rethinking the "Gap": Self-Directed Learning in Cognitive Development and Scientific Reasoning." *Wiley Interdisciplinary Reviews: Cognitive Science* 13, no. 2: e1580.

Laurendeau, M., and A. Pinard. 1962. *Causal Thinking in Children*. International Universities Press.

Lecce, S., S. Mascheretti, S. M. Stagnitto, and C. Osterhaus. 2025. "Individual Differences in Theory of Mind and Children's Advanced Scientific Reasoning in Elementary School." *International Journal of Behavioral Development* 49, no. 4: 382–388. https://doi.org/10.1177/01650254241305545.

Lederman, N. G., F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz. 2002. "Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science." *Journal of Research in Science Teaching* 39, no. 6: 497–521. https://doi.org/10.1002/tea.10034.

Lockl, K., and W. Schneider. 2007. "Knowledge About the Mind: Links Between Theory of Mind and Later Metamemory." *Child Development* 78, no. 1: 148–167.

Lucas, A. J., C. Lewis, F. C. Pala, K. Wong, and D. Berridge. 2013. "Social-Cognitive Processes in Preschoolers' Selective Trust: Three Cultures Compared." *Developmental Psychology* 49, no. 3: 579.

Mazens, K., and J. Lautrey. 2003. "Conceptual Change in Physics: Children's Naïve Representation of Sound." *Cognitive Development* 18: 159–176.

Moeller, A., B. Sodian, and D. M. Sobel. 2022. "Developmental Trajectories in Diagnostic Reasoning: Understanding Data Are Confounded Develops Independently of Choosing Informative Interventions to Resolve Confounded Data." *Frontiers in Psychology* 13: 800226.

Morgan, P. L., G. Farkas, M. M. Hillemeier, and S. Maczuga. 2016. "Science Achievement Gaps Begin Very Early, Persist, and Are Largely Explained by Modifiable Factors." *Educational Researcher* 45, no. 1: 18–35. https://doi.org/10.3102/0013189X16633182.

Opfer, J. E., and R. S. Siegler. 2004. "Revisiting Preschoolers' Living Things Concept: A Microgenetic Analysis of Conceptual Change in Basic Biology." *Cognitive Psychology* 49: 301–332.

Osterhaus, C., and S. Koerber. 2021. "Social Cognition During and After Kindergarten: The Relations Between First-Order and Advanced Theories of Mind." *European Journal of Developmental Psychology* 18, no. 4: 573–592.

Osterhaus, C., and S. Koerber. 2023. "The Complex Associations Between Scientific Reasoning and Advanced Theory of Mind." *Child Development* 94, no. 1: e18–e42.

Osterhaus, C., S. Koerber, and B. Sodian. 2017. "Scientific Thinking in Elementary School: Children's Social Cognition and Their Epistemological Understanding Promote Experimentation Skills." *Developmental Psychology* 53, no. 3: 450.

Perner, J. 1991. *Understanding the Representational Mind*. MIT Press.

Perner, J., S. R. Leekam, and H. Wimmer. 1987. "Three-Year-Olds' Difficulty With False Belief: The Case for a Conceptual Deficit." *British Journal of Developmental Psychology* 5, no. 2: 125–137.

Perner, J., and H. Wimmer. 1985. ""Johnthinks That Mary Thinks That...": Attribution of Second-Order Beliefs by 5- To 10-Year-Old Children." *Journal of Experimental Child Psychology* 39, no. 3: 437–471. https://doi.org/10.1016/0022-0965(85)90051-7.

Piekny, J., D. Grube, and C. Maehler. 2013. "The Relation Between Preschool Children's False-Belief Understanding and Domain-General Experimentation Skills." *Metacognition and Learning* 8: 103–119.

Piekny, J., D. Grube, and C. Maehler. 2014. "The Development of Experimentation and Evidence Evaluation Skills at Preschool Age." *International Journal of Science Education* 36: 334–354. https://doi.org/10.1080/09500693.2013.776192.

Piekny, J., and C. Maehler. 2013. "Scientific Reasoning in Early and Middle Childhood: The Development of Domain-General Evidence

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Evaluation, Experimentation, and Hypothesis Generation Skills." *British Journal of Developmental Psychology* 31, no. 2: 153–179.

Quinn, D. M., and N. Cooc. 2015. "Science Achievement Gaps by Gender and Race/Ethnicity in Elementary and Middle School: Trends and Predictors." *Educational Researcher* 44, no. 6: 336–346. https://doi.org/10.3102/0013189X15598539.

Ruffman, T., J. Perner, D. R. Olson, and M. Doherty. 1993. "Reflecting on Scientific Thinking: Children's Understanding of the Hypothesis-Evidence Relation." *Child Development* 64, no. 6: 1617–1636.

Saçkes, M., L. M. Flevares, and K. C. Trundle. 2010. "Four-To Six-Year-Old Children's Conceptions of the Mechanism of Rainfall." *Early Childhood Research Quarterly* 25, no. 4: 536–546.

Saçkes, M., S. Işıtan, K. C. Trundle, and D. M. Sobel. 2025. *Turkish Preschoolers Show an Advantage in Their Understanding of Their Own Representational Change Over Others' False Belief [Manuscript Submitted for Publication] . Necatibey School of Education*. Balikesir University.

Saçkes, M., K. C. Trundle, and R. L. Bell. 2013. "Science Learning Experiences in Kindergarten and Children's Growth in Science Performance in Elementary Grades." *Education and Science* 38, no. 167: 112–125.

Saçkes, M., K. C. Trundle, R. L. Bell, and A. A. O'Connell. 2011. "The Influence of Early Science Experience in Kindergarten on Children's Immediate and Later Science Achievement: Evidence From the Early Childhood Longitudinal Study." *Journal of Research in Science Teaching* 48, no. 2: 217–235.

Saçkes, M., K. C. Trundle, and M. Shaheen. 2020. "The Effect of Balanced Learning Curriculum on Young Children's Learning of Science." *Early Childhood Education Journal* 48: 305–312.

Samarapungavan, A., P. Mantzicopoulos, H. Patrick, and B. French. 2009. "The Development and Validation of the Science Learning Assessment (SLA): A Measure of Kindergarten Science Learning." *Journal of Advanced Academics* 20, no. 3: 502–535.

Schauble, L., R. Glaser, R. Duschl, S. Schulze, and J. John. 1995. "Students' Understanding of the Objectives and Procedures of Experimentation in the Science Classroom." *Journal of the Learning Sciences* 4: 131–166.

Schult, C. A., and H. M. Wellman. 1997. "Explaining Human Movements and Actions: Children's Understanding of the Limits of Psychological Explanation." *Cognition* 62, no. 3: 291–324. https://doi.org/10.1016/S0010-0277(96)00786-X.

Selcuk, B., K. A. Brink, M. Ekerim, and H. M. Wellman. 2018. "Sequence of Theory-Of-Mind Acquisition in Turkish Children From Diverse Social Backgrounds." *Infant and Child Development* 27, no. 4: e2098.

Senocak, E., A. Samarapungavan, P. Aksoy, and C. Tosun. 2013. "A Study on Development of an Instrument to Determine Turkish Kindergarten Students' Understandings of Scientific Concepts and Scientific Inquiry Processes." *Educational Sciences: Theory & Practice* 13, no. 4: 2217–2228.

Shahaeian, A., C. C. Peterson, V. Slaughter, and H. M. Wellman. 2011. "Culture and the Sequence of Steps in Theory of Mind Development." *Developmental Psychology* 47, no. 5: 1239–1247. https://doi.org/10.1037/a0023899.

Shatz, M., G. Diesendruck, I. Martinez-Beck, and D. Akar. 2003. "The Influence of Language and Socioeconomic Status on Children's Understanding of False Belief." *Developmental Psychology* 39, no. 4: 717–729.

Shavlik, M., Ö. Köksal, B. F. French, C. A. Haden, C. H. Legare, and A. E. Booth. 2022. "Contributions of Causal Reasoning to Early Scientific Literacy." *Journal of Experimental Child Psychology* 224: 105509.

Siegler, R. S., and Z. Chen. 1998. "Developmental Differences in Rule Learning: A Microgenetic Analysis." *Cognitive Psychology* 36, no. 3: 273–310.

Sobel, D. M. 2023. "Growing Out of Your Own Mind: Reexamining the Development of the Self-Other Difference in the Unexpected Contents Task." *Cognition* 235: 105403. https://doi.org/10.1016/j.cognition.2023.105403.

Sodian, B., S. Kristen-Antonow, and S. Koerber. "Theory of Mind Predicts Scientific Reasoning. A Longitudinal Study From Preschool to Elementary School Age." Paper Presented at the International Congress of Psychology, Yokohama, Japan. 2016.

Sodian, B., D. Zaitchik, and S. Carey. 1991. "Young Children's Differentiation of Hypothetical Beliefs From Evidence." *Child Development* 62, no. 4: 753–766.

Spelke, E. S. 2005. "Sex Differences in Intrinsic Aptitude for Mathematics and Science?: A Critical Review." *American Psychologist* 60, no. 9: 950–958.

Tschirgi, J. E. 1980. "Sensible Reasoning: A Hypothesis About Hypotheses." *Child Development* 51, no. 1: 1–10.

van der Graaf, J., E. Segers, and L. Verhoeven. 2015. "Scientific Reasoning Abilities in Kindergarten: Dynamic Assessment of the Control of Variables Strategy." *Instructional Science* 43: 381–400.

van der Graaf, J., E. Segers, and L. Verhoeven. 2016. "Scientific Reasoning in Kindergarten: Cognitive Factors in Experimentation and Evidence Evaluation." *Learning and Individual Differences* 49: 190–200.

van der Graaf, J., E. Segers, and L. Verhoeven. 2018. "Individual Differences in the Development of Scientific Thinking in Kindergarten." *Learning and Instruction* 56: 1–9.

Vosniadou, S. 1994. "Capturing and Modeling the Process of Conceptual Change." *Learning and Instruction* 4, no. 1: 45–69.

Vosniadou, S., and W. F. Brewer. 1992. "Mental Models of the Earth: A Study of Conceptual Change in Childhood." *Cognitive Psychology* 24: 535–585.

Weisberg, D. S., and D. M. Sobel. 2022. *Constructing Science: Connecting Causal Reasoning to Scientific Thinking in Young Children*. MIT Press.

Wellman, H. M., D. Cross, and J. Watson. 2001. "Meta-Analysis of Theory-Of-Mind Development: The Truth About False Belief." *Child Development* 72: 655–684. https://doi.org/10.1111/1467-8624.0030.

Wellman, H. M., F. Fang, D. Liu, L. Zhu, and G. Liu. 2006. "Scaling of Theory-of-Mind Understandings in Chinese Children." *Psychological Science* 17, no. 12: 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x.

Wellman, H. M., and D. Liu. 2004. "Scaling of Theory-of-Mind Tasks." *Child Development* 75, no. 2: 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x.

Yoon, J., and J. A. Onchwari. 2006. "Teaching Young Children Science: Three Key Points." *Early Childhood Education Journal* 33: 419–423.

Zimmerman, C. 2007. "The Development of Scientific Thinking Skills in Elementary and Middle School." *Developmental Review* 27, no. 2: 172–223.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.

15227219, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.70093 by Balikesir University, Wiley Online Library on [17/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License