Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-026-07393-2

ORIGINAL ARTICLE

Teaching Theory of Mind Skills to Individuals with Autism Spectrum Disorder: A Systematic Review and Meta-analysis

Betül Kat Şen[1,2] · **Aydan Aydın**[2]

Received: 18 August 2025 / Accepted: 27 May 2026
© The Author(s) 2026

Abstract

Purpose In this study, we examined the effectiveness of Theory of Mind (ToM) intervention programs on the ToM skills of individuals with Autism Spectrum Disorder (ASD) through a systematic review and meta-analysis.

Methods A comprehensive search was conducted in Web of Science, Scopus, EBSCO ERIC, Academic Search Complete, PubMed, ProQuest, and the Türkiye Council of Higher Education Thesis Center, completed on January 5, 2024. Studies were included if published in English or Turkish, involved participants with ASD, implemented interventions targeting ToM, used at least one ToM assessment tool, employed a randomized controlled trial (RCT) or quasi-experimental design, and provided sufficient data for meta-analysis. Methodological quality was evaluated using What Works Clearinghouse (WWC) standards. The meta-analysis used a fixed-effects model with Hedges' g in Comprehensive Meta-Analysis Software 4.

Results We identified 20 studies ($n=924$) meeting the inclusion criteria, of which 15 were included in the meta-analysis. Most interventions targeted understanding first- and second-order false beliefs and frequently used role-play and picture-based storytelling techniques. The results indicated a moderate positive effect ($g=0.492$, 95% CI [0.322, 0.662]). Subgroup analyses showed no significant differences based on study characteristics.

Conclusion Findings should be interpreted cautiously due to publication bias, small sample sizes, and the lack of IQ, language, or prerequisite skill assessments. Crucially, insufficient reporting of social validity and generalization represents a major barrier to assessing real-world utility. Future research must prioritize rigorous RCTs and systematic functional reporting to strengthen the evidence base.

Keywords Autism spectrum disorder · Theory of mind · Intervention · Meta-analysis · Systematic review

Autism Spectrum Disorder (ASD) has been defined as a neurodevelopmental difference characterized by limitations in social interaction and communication skills, which are noticed in early childhood, along with patterns of repetitive behaviors and activities (American Psychiatric Association, 2013). Theory of mind (ToM), a concept defined as the ability to understand others' thoughts, feelings, desires, intentions, beliefs, and mental states (Premack & Woodruff, 1978; Wellman et al., 2001), is a critical component of social interaction and communication in individuals diagnosed with ASD (Zwaigenbaum et al., 2005). This skill enables individuals to predict others' behavior based on these inferences (Gallagher & Frith, 2003). ToM, a core component of social cognition, has been theorized as a potential underlying factor in the social interaction and communication difficulties observed in individuals with ASD (Korkmaz, 2011).

One of the most critical skills targeted in ToM, regarded as a developmental ability, is understanding beliefs. The ability to recognize that an individual can hold a false belief typically emerges around the age of four (Kobayashi Frank & Temple, 2009; Perner & Roessler, 2012; Tomasello et al., 2005). However, precursor skills, including joint attention, intentionality, perspective-taking, and symbolic play, emerge in infancy (Miller, 2006). Children demonstrate first-order perspective-taking by age two, understanding what others can or cannot see (Moll & Tomasello, 2006). Around

✉ Betül Kat Şen
betul.kat@izu.edu.tr

Aydan Aydın
aydanaydin@marmara.edu.tr

[1] Department of Special Education, Faculty of Education, Istanbul Sabahattin Zaim University, Istanbul, Turkey

[2] Department of Special Education, Marmara University, Istanbul, Turkey

Published online: 19 June 2026

age three, they understand that seeing leads to knowing and recognize that others may have different desires (Pratt & Bryant, 1990). At this stage, they also begin to use terms such as "want" and "think," reflecting an emerging grasp of desires, emotions, and beliefs (Bretherton & Beeghly, 1982; J. Brown & Dunn, 1991; Wellman et al., 2000). By the same age, children start to distinguish between mental and physical entities and to realize that individuals can hold different thoughts (Flavell et al., 1990; Watson et al., 1998; Wellman & Estes, 1986; Wellman & Woolley, 1990). Between four and five years, they comprehend that beliefs can guide behavior, even if those beliefs are false (Wimmer & Perner, 1983). By six to seven years, they typically acquire second-order false belief understanding, enabling them to think about what others are thinking (Baron-Cohen, 2000; Miller, 2013). More advanced ToM skills continue to emerge later, with recursive thinking developing around age seven and reasoning about social events around age nine (Wang et al., 2016).

This developmental trajectory, however, differs for children with Autism Spectrum Disorder (ASD). ToM development is often delayed in these children and follows an atypical course (Hoogenhout & Malcolm-Smith, 2014). Research consistently shows that individuals with ASD experience significant challenges in acquiring ToM compared to typically developing peers and other diagnostic groups (Baron-Cohen et al., 1997; Brewer et al., 2017; Kaland et al., 2002; Mazza et al., 2017; Schwartz Offek & Segal, 2022; Zalla et al., 2009).

Individuals diagnosed with ASD demonstrate distinct impairments in social interaction and communication, which are associated with impaired ToM skills, or the capacity to comprehend the intentions, desires, and thoughts of others (Devine & Apperly, 2021). Individuals with ASD have been shown to experience difficulties in ToM that can impact their social interaction and communication skills (Chiu et al., 2023; Mao et al., 2023), empathic abilities (Mathersul et al., 2013), friendships and romantic relationships (Banerjee et al., 2011; Bauminger et al., 2010; Dodell-Feder et al., 2016; Fink et al., 2015), and comprehension of implicit concepts such as irony, similes, and metaphors (Happé, 1999). Developing ToM skills can facilitate the social participation of individuals with ASD by enabling them to establish healthy social relationships (Ma et al., 2023). In this context, researchers have developed ToM intervention programs focusing on emotions, perspective taking, ToM precursors, first-order false belief, and second-order false belief in order to enhance ToM skills in individuals with ASD, and have tested their effectiveness in ToM skill domains (Begeer et al., 2015; Bylemans et al., 2022; Cheung et al., 2018; Cheung et al., 2020; de Veld et al., 2017; Durrleman et al., 2022; Ma et al., 2023; Ozonoff & Miller, 1995; Rice et al., 2015; Waugh & Peskin, 2015). Therefore, to determine the direction and magnitude of an intervention program's effect, it is necessary to calculate the effect size and identify which variables account for it (Sullivan & Feinn, 2012). A meta-analysis was conducted by Hofmann et al. (2016) to investigate the effectiveness of ToM interventions for children. The meta-analysis results indicate that ToM intervention programs have a large effect on children's ToM skills, as determined by the random-effects model ($g = 0.75$). In another meta-analysis of Fletcher-Watson et al. (2014), the effectiveness of ToM interventions implemented with individuals diagnosed with ASD was evaluated. The data obtained from these studies were subsequently analyzed under the following categories: emotion recognition, joint attention, imitation, social communication, and theory of mind skills. The study revealed that the effect sizes of ToM interventions demonstrated inconsistent findings ($g = 0.3$–0.6) and were methodologically deficient. While the study by Hofmann et al. (2016) is a pioneering work demonstrating the general effectiveness of theory of mind interventions, it is argued that the combined evaluation of typically developing samples and clinical populations may obscure intervention outcomes specific to ASD. In contrast, this study focuses exclusively on individuals diagnosed with ASD and aims to address the needs of this heterogeneous group in a more specific manner. Although limitations in theory of mind are often presented as a uniform concept in the ASD literature, individuals with ASD may experience difficulties with an intuitive and spontaneously developed "implicit" mind-reading system. However, they can develop an "explicit" compensatory mechanism to make up for this deficit by using their logical and linguistic abilities. This creates diversity, leading to distinct social skill profiles in each individual (Happé & Frith, 2014). In particular, individuals with ASD who possess high cognitive capacity can conceal their social-cognitive deficits through compensatory mechanisms and masking strategies (Livingston & Happé, 2017; Hull et al., 2017). Rather than intuitively interpreting mental states as neurotypical individuals do (consciously or unconsciously), these individuals may develop rule-based methods that enable them to perform better on standard theory of mind tests (Livingston et al., 2019). Therefore, when evaluating the effectiveness of theory of mind-based interventions using explicit or implicit theory of mind tests, determining whether the results reflect actual social cognitive development or result from developed compensatory strategies is crucial to the clinical validity of intervention outcomes (Happé & Frith, 2014). This study aims to evaluate the generalizability and maintenance of intervention outcomes, moving beyond static ToM test scores to consider the real-life application of implicit ToM abilities alongside other qualitative features. Furthermore,

it seeks to examine how moderator variables that contribute to heterogeneity in ASD influence overall mean effect sizes. Finally, in this study, the literature on theory of mind interventions was expanded to include the most recent studies up to 2024, including Turkish studies, drawing on the Cochrane review by Fletcher-Watson et al. (2014). Additionally, to address the methodological limitations noted by Fletcher-Watson et al., the existing studies were evaluated using the WWC standards, which are widely used in education.

In its report on evidence-based practices, the National Autism Center (NAC) stated that ToM interventions were not yet fully supported by evidence for individuals with ASD and were categorized as emerging interventions—defined as approaches that have shown positive outcomes in one or a few studies (NAC, 2015). For these interventions to be regarded as evidence-based, consistent results from high-quality studies are necessary. Although previous meta-analyses have estimated overall effect sizes, it is necessary to systematically examine study characteristics that may explain variability in these effects and investigate moderator effects. The developmental structure of the theory of mind ability (Wellman, 2017) means that we need to examine age as a moderator variable. Furthermore, previous meta-analyses have found that the length and number of sessions in theory-of-mind interventions moderate their effects (Hofmann, 2016). Therefore, it is important to assess for a similar moderating effect in this study to ensure consistency. Although group and individual interventions on children with autism are known to have similar effects, the fact that individual interventions yield more reliable results on reciprocal social interaction (Tachibana et al., 2018) suggests that the mode of delivery may be a potential moderator in interventions such as theory of mind. The differences in research designs regarding counterfactual inferences and control of external variables between randomised controlled and quasi-experimental studies (Shadish et al., 2002) necessitate the examination of moderator variables.

By identifying the characteristics of studies that can explain the average effect size of ToM interventions on individuals with ASD and the variability in outcomes, the explanation of the moderator effect will guide researchers and practitioners in educational and clinical settings on whether ToM interventions should be incorporated into teaching practices for individuals with ASD and, if so, under what conditions they may be most effective. It is expected that this study will guide the decision on whether ToM interventions should be incorporated into teaching practices for individuals with ASD and, if so, under which conditions they may be most effective, thereby providing information for evidence-based decision-making.

In this context, the present study aims to (a) evaluate the methodological quality of intervention programs designed to improve ToM skills based on WWC standards, (b) estimate the individual and overall effect sizes of the included studies, (c) describe the qualitative and methodological characteristics of the included studies, and (d) examine whether selected study characteristics (namely participant age, intervention delivery format, intervention duration, and research design) serve as significant moderators of the intervention effects. Accordingly, the objective of this study is to address the following three research questions:

1. What is the effect of intervention programs aimed at improving ToM skills on the ToM abilities of individuals with ASD?
2. What are the qualitative characteristics (e.g., demographic information, methodology, intervention features) of the experimental studies that examine the effects of these intervention programs on the ToM skills of individuals with ASD?
3. To what extent do selected qualitative and methodological characteristics of the included studies—namely participant age, intervention delivery format, intervention duration, and research design—function as potential moderators of the effect sizes of ToM interventions for individuals with ASD?
4. To what extent do the experimental studies examining the effects of these intervention programs on the ToM skills of individuals with ASD meet the WWC standards?

Method

In this study, we conducted a systematic review and meta-analysis to examine interventions to improve ToM skills in individuals with ASD. A systematic review synthesizes multiple studies addressing the same research question (Cook et al., 1997), while meta-analysis applies statistical methods to aggregate findings and identify evidence-based practices (Borenstein et al., 2013; Mulrow, 1994; Tranfield et al., 2003). The present study followed the PRISMA checklist (Page et al., 2021).

Search Strategy

We aimed to identify experimental articles and theses, published in either English or Turkish, that implemented intervention programs to improve ToM skills in individuals diagnosed with ASD—including Asperger's syndrome, high-functioning autism, and pervasive developmental disorder not otherwise specified—without applying any date restrictions. We selected seven databases—Web of Science, Scopus, EBSCO ERIC, Academic Search Complete, PubMed, ProQuest, and the Türkiye Council of Higher

Education (CoHE) Thesis Center. We identified three primary keyword groups: target population, skill domain, and intervention, and applied Boolean logic (e.g., AND, OR). The keywords were: ("theory of mind" OR "mind reading" OR "social cognition" OR "mental state" OR "cognitive empathy" OR "mentalization" OR "social intelligence" OR "perspective taking" OR "false belief") AND ("train" OR "training" OR "teach" OR "teaching" OR "intervention" OR "instruction") AND ("autism" OR "autism spectrum disorder" OR "high functioning autism" OR "Asperger syndrome" OR "pervasive developmental disorder" OR "ASD" OR "PDD"). We completed the search on January 5, 2024, identifying 2,919 studies. We also reviewed the reference lists of published systematic reviews, which yielded an additional 20 studies. All studies obtained were imported into EndNote 21 for classification according to the inclusion and exclusion criteria.

Inclusion and Exclusion Criteria

We did not include all studies identified during the screening process for the meta-analysis. We applied the following inclusion criteria: (a) publication as an article in an international peer-reviewed journal or as a thesis in the ProQuest or CoHE Thesis Center databases, (b) published in either English or Turkish, (c) participants diagnosed with ASD, Asperger syndrome, high-functioning autism, or pervasive developmental disorder-not otherwise specified, (d) implementation of an intervention program aimed at teaching ToM skills (including subskills such as perspective-taking), (e) inclusion of at least one measurement tool assessing ToM skills, (f) use of a RCT or quasi-experimental research design, (g) provision of an appropriate dataset (e.g., sample size, mean, standard deviation), and (h) for studies included in the descriptive analysis, fulfillment of all criteria except the seventh. We excluded studies that (a) were published in languages other than Turkish or English, (b) used research methods other than RCT or quasi-experimental designs, (c) involved only participants from other diagnostic groups or typically developed individuals without an ASD diagnosis, (d) implemented an intervention program targeting skills other than ToM, (e) lacked any measurement tools for assessing ToM skills, or (f) lacked a suitable dataset for meta-analysis (e.g., sample size, mean, standard deviation).

Study Selection

Our database and reference list searches yielded a total of 2,939 studies. We excluded 1,182 of these as duplicates. From the remaining 1,757 studies, we excluded 640 for using research methods other than intervention programs (e.g., survey, correlational), 392 for employing qualitative research methods or review studies, 174 for being published in formats other than articles or theses, 32 for being meta-analyses, 21 for being written in language other than Turkish or English, and 15 for being inaccessible. These steps resulted in 483 studies that included an intervention program. We excluded 317 studies because the intervention targeted skills outside the ToM domain, leaving 166 studies aimed at improving ToM skills. Of these, we excluded 68 for using different research designs (e.g., weak experimental designs, single-subject studies), 54 for targeting populations other than individuals with ASD (e.g., different diagnostic groups or typically developing children), and 24 for not employing appropriate measurement tools. After these exclusions, 20 studies remained eligible for descriptive analysis. We excluded five of these from the meta-analysis because of the absence of a suitable dataset for meta-analytic calculations. Ultimately, we included 15 studies in the meta-analysis. Fig.1 presents the PRISMA flow diagram summarizing the evaluation process of the studies according to the inclusion and exclusion criteria.

Data Extraction and Coding

The first author coded all studies included in the research via a coding form. This form consisted of three sections: bibliographic information, study-related content, and numerical data required for the meta-analysis. The first section recorded the code, study title, author(s), year, publication type, journal name or university, and country codes. In the second section, we documented details about the research design, ToM intervention themes, methods and techniques used, number of sessions, session duration, implementation format (individual or group), practitioner, sample characteristics (gender, mean age, diagnosis, comparison group), setting, measurement tools, overall effect, generalization, follow-up, and social validity codes. In the third section, we entered the sample sizes for the experimental and control groups, the pre-test and post-test mean scores, the pre-test and post-test standard deviations, and, when available, the F and p values.

Data Reliability

Interrater reliability refers to the consistency between two or more raters evaluating the same measure (Cohen & Swerdlik, 2009), and the literature indicates that this value should be at least 80% (Miles & Huberman, 1994). In this context, we used multiple methods to assess the reliability of the research during screening, data extraction, and moderator rating. We adopted a consensus-based approach to selecting studies, evaluating their eligibility according to inclusion criteria, and rating the moderator variables. We

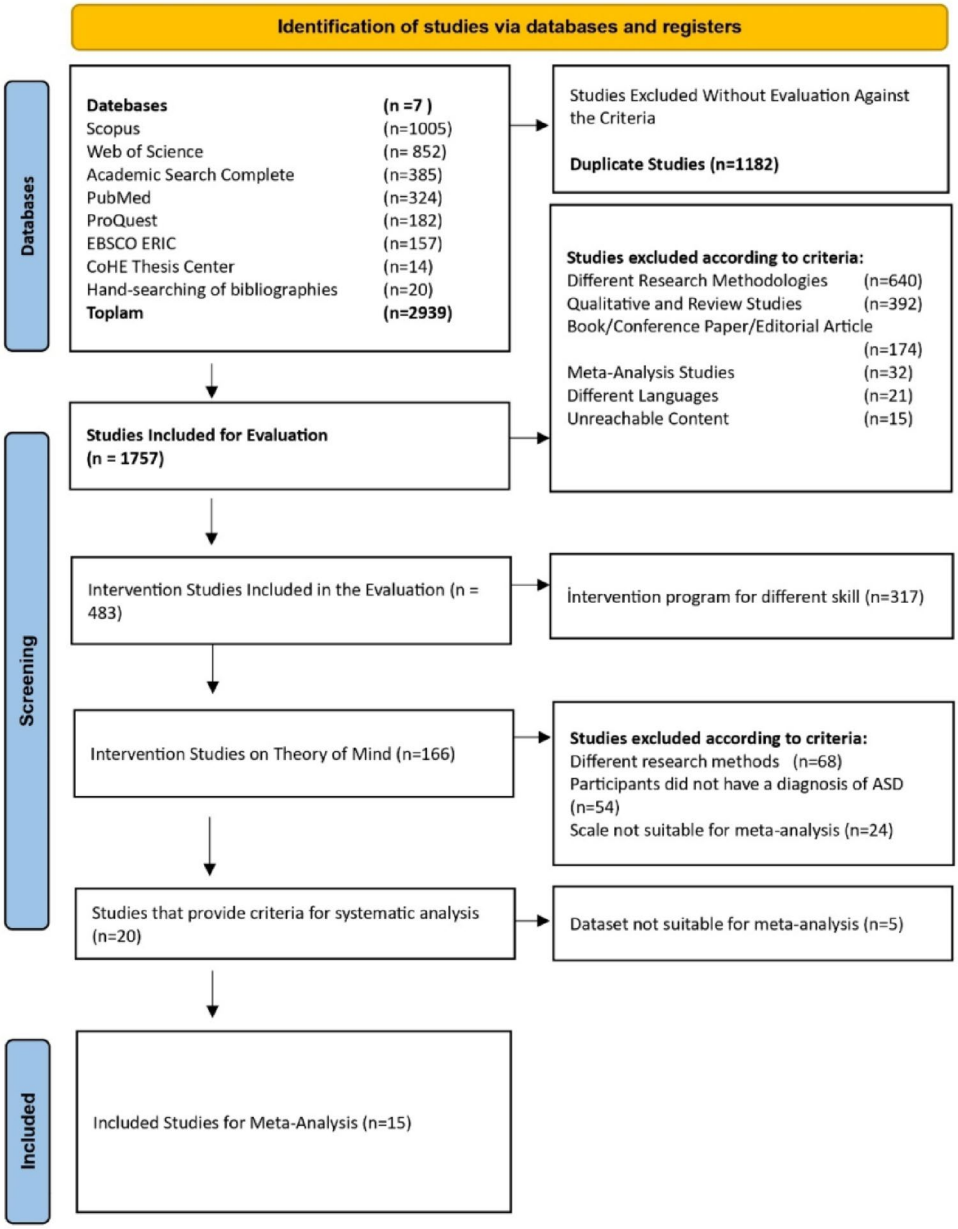


Fig. 1 PRISMA flow diagram of the study evaluation process based on inclusion and exclusion criteria. This flow diagram was created using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) framework to illustrate the study selection process

determined the eligibility of the studies through discussions and consensus within the scope of inclusion and exclusion criteria. We selected moderator variables based on both the developmental literature and a balanced distribution of group studies. We primarily used interrater reliability data from the studies during data extraction. In this process, we created a coding list including the study name, author, year, country, research design, theory of mind training themes, delivery format, number and duration of sessions, practitioner, sample characteristics (mean age, gender, diagnosis), measurement tools, and statistical data (N, M, SD, F, and *p*-values). The first author coded this list, and the second author reviewed it; subsequently, two independent researchers specializing in special education and psychological counseling re-evaluated the data and checked it against the coding list. In the final stage, we calculated reliability using the formula by Miles and Huberman (1994) [Agreement/(Agreement + Disagreement) × 100] and found an average agreement of 97.9% (range: 91–100%).

What Works Clearinghouse (WWC) Procedures and Standards

This study evaluated the included studies according to the standards outlined in the "Reviewing Findings from Randomized Controlled Trials and Quasi-Experimental Designs" section of the WWC Procedures and Standards Handbook (WWC, 2022). We developed a flow diagram to

guide the evaluation process (see Online Resource 1). Each study was assessed twice, with a one-week interval between assessments, and rated as "Meets WWC standards," "Meets WWC standards with reservations," or "Does not meet WWC standards."

Statistical Analysis

We calculated the effect size for the studies in the meta-analysis based on the mean difference between the experimental and control groups. Because the studies used different measurement instruments, we computed effect sizes via the standardized mean difference. We calculated Hedges ' g for studies with fewer than 20 participants to correct for small-sample bias (Cooper et al., 2009).

We assessed homogeneity and heterogeneity using Cochran's Q and I^2 statistics. As shown in Online Resource 2, the Q value for the 15 included studies was 15.373 with 14 degrees of freedom, and the I^2 value was 8.930, indicating homogeneity among the studies (Higgins et al., 2003). However, this result should be interpreted with caution; given the relatively small number of studies and the conceptual differences among various ToM intervention methods, the low I^2 value may not fully reflect the clinical and methodological heterogeneity of the primary studies. The evidence of statistical homogeneity, when combined with the inclusion of both large- and small-sample studies, supports the use of a fixed-effects model (Dettori et al., 2022). However, we remain cautious regarding the generalizability of a single pooled effect size. To preserve the independence of observations, a fundamental assumption of the fixed-effects model, we calculated a single effect size per study in the meta-analysis. As suggested by Lipsey and Wilson (2001, p.78) we selected the most prevalent and comprehensive Theory of Mind scales representing the primary outcomes of the included studies; these included the Theory of Mind Test (Muris et al., 1999), Strange Stories (Happé, 1994; O'Hare et al., 2009), and False Belief Test (Baron-Cohen et al., 1985; Perner & Wimmer 1985). For studies that provided a single composite score representing overall mentalizing ability alongside multiple ToM tasks (e.g., Hadwin et al., 1996; Soorya et al., 2015), we used this composite score to estimate the study's effect size.

We also evaluated publication bias via the Trim and Fill method (Duval & Tweedie, 2000) and the Fail-Safe N method (Rosenthal, 1979). All statistical analyses were conducted with Comprehensive Meta-Analysis Software, version 4.

Result

Descriptive Characteristics

WWC Standards, Publication Year, Country, Research Design, and Publishing Journal/University

This study evaluated the 20 included studies according to the standards outlined in the "Reviewing Findings from Randomized Controlled Trials and Quasi-Experimental Designs" section of the WWC Procedures and Standards Handbook. The evaluation revealed that nine studies met WWC standards without reservations, eight did not meet WWC standards, and three met WWC standards with reservations. The studies were further categorized into five-year intervals: two conducted between 1994 and 1998, two between 1999 and 2003, two between 2004 and 2008, three between 2009 and 2013, six between 2014 and 2018, and five between 2019 and 2023. Geographically, the studies were conducted in the United States, the Netherlands, the United Kingdom, Australia, Belgium, China, Iran, Israel, Switzerland, Japan, Canada, and Türkiye, with the United States and the Netherlands hosting the largest number, followed by the United Kingdom. Regarding publication type, 18 included studies were published as articles in peer-reviewed journals, and two were published as theses—one as a doctoral dissertation and the other as a master's thesis. The Journal of Autism and Developmental Disorders published the most extensive studies ($n=10$). In contrast, the remaining studies appeared in Autism, Autism Research, Clinical Child Psychology and Psychiatry, Development and Psychopathology, Frontiers in Behavioral Neuroscience, Frontiers in Psychology, Journal of the American Academy of Child & Adolescent Psychiatry, and Research in Autism Spectrum Disorders (see Online Resource 3).

Research Design, Sample, and Measurement Instruments, Intervention Themes, Instructional Methods and Techniques, Number and Duration of Sessions, Implementers and Settings

We analyzed the 20 included studies and identified 10 that employed a randomized controlled trial design and 10 that used a quasi-experimental design. Examination of the samples revealed 924 participants, with 51.83% ($n=479$) in the experimental group and 48.16% ($n=445$) in the control group. Five studies did not report sex information; in the remaining 16 studies, 107 female and 659 male participants were included. The mean participant age was 132.21 months (approximately 11.01 years). Participants had diagnoses of ASD, Asperger syndrome, high-functioning autism spectrum disorder, or pervasive developmental disorder–not

otherwise specified. Additional control groups included typically developing children and one group receiving a play-based intervention. Concerning the instruments used, seven studies employed the Strange Story Measure (Happé, 1994) for advanced ToM skills. The Sally-Anne Test was used for first-order false belief, the M&M test for second-order false belief, and the Reading the Mind in the Eyes Test for emotion recognition (see Online Resource 4).

We analyzed the Theory of Mind (ToM) intervention programs under six themes: emotion, perspective-taking, ToM precursors, first-order false belief, second-order false belief, and advanced ToM skills. Frequency analysis revealed that 17 studies targeted first-order false belief, 16 targeted second-order false belief, 13 targeted ToM precursors, eight targeted perspective-taking, eight targeted emotion, and seven targeted advanced ToM skills. Instructional methods and techniques were categorized into ten themes: role-play, reading or telling picture stories, group activities, play, homework, thought bubble technique, technology-assisted instruction, feedback, reinforcement, and prompting. Eleven studies incorporated role-play, ten used picture stories, eight used group activities, seven used play, seven used homework, six employed the thought bubble technique, five used technology-assisted instruction, five provided feedback, four used reinforcement, and fifteen implemented prompting. With respect to program length, the average was 11.9 sessions, most frequently 11–15 sessions ($n=9$), followed by 6–10 ($n=7$), 1–5 ($n=2$), 16–20 ($n=1$), and more than 20 ($n=1$). The average session duration was 51.38 min, with eight studies reporting 31–60 min, seven reporting 0–30 min, and three reporting 61–90 min. Regarding implementers, six interventions were conducted by the author/researcher, five by psychologists, two by teachers, two by therapists, two by research assistants, one by an occupational therapist with students, and one by parents. Settings included five schools, three child psychiatry centers, one special education center, one computer laboratory, one home setting, and one online platform (see Online Resource 4).

Findings on Generalization, Maintenance, and Social Validity

We found that five studies collected generalization data, whereas fifteen did not—all five measured stimulus generalization, with three examining generalization across materials and two across people. In terms of maintenance, ten studies collected data and ten did not. Of those that did, one measured maintenance over six months, and the others between one and three months, with the average maintenance interval calculated as 48.22 days. Data sources included children, parents, and teachers. Four studies collected data on social validity, whereas sixteen did not (see Online Resource 5).

Meta-analysis

Overall Effect Size ToM Interventions

We conducted a meta-analysis with 15 of the 20 included studies whose datasets met the inclusion requirements. The analysis revealed that the fixed-effects model showed a significant and positive overall effect size of $g=0.49$ (95% CI [0.32, 0.66]; $p<0.001$). The low heterogeneity among studi We conducted a meta-analysis with 15 of the 20 included studies whose datasets met the inclusion requirements. The analysis revealed that the fixed-effects model showed a significant and positive overall effect size of $g=0.49$ (95% CI [0.32, 0.66]; $p<0.001$). The low heterogeneity among studies ($Q(14)=15.37$, $p=0.35$, $I^2 = 8.93\%$) supported selecting this model as the primary analysis method. Additional analyses using a random-effects model to test the robustness of the findings yielded results consistent with the main findings ($g=0.50$, 95% CI [0.32, 0.68]; $p<0.001$) (see Online Resource 2). Due to concerns that the random-effects model would place excessive weight on the results of small-scale studies and increase publication bias by giving greater weight to smaller studies (Deeks et al., 2024), and because of statistical heterogeneity, the fixed-effects model was more appropriate for analyzing the effects of interventions in this review. Nonetheless, given the small number of studies and the conceptual differences among the interventions, this finding should be interpreted with caution. While statistically valid, we adopt a cautious stance regarding the generalizability of the effect size from the fixed-effects model, given differences in intervention methods. On the basis of these results, we concluded that, according to the fixed-effects model, ToM intervention programs aimed at improving ToM skills have a moderate effect on the ToM skills of individuals diagnosed with ASD ($0.2<$Cohen's $d<0.5$ indicates a moderate effect).es ($Q(14)=15.37$, $p=0.35$, $I^2 = 8.93\%$) supported selecting this model as the primary analysis method. Additional analyses using a random-effects model to test the robustness of the findings yielded results consistent with the main findings ($g=0.50$, 95% CI [0.32, 0.68]; $p<0.001$) (see Online Resource 2). Given concerns that the random-effects model may overemphasize the effects of small studies and exacerbate publication bias by assigning greater weight to smaller studies (Desk et al., 2024), the fixed-effects model was deemed more appropriate for evaluating intervention effects in the present study. On the basis of these results, we concluded that, according to the fixed-effects model, ToM intervention

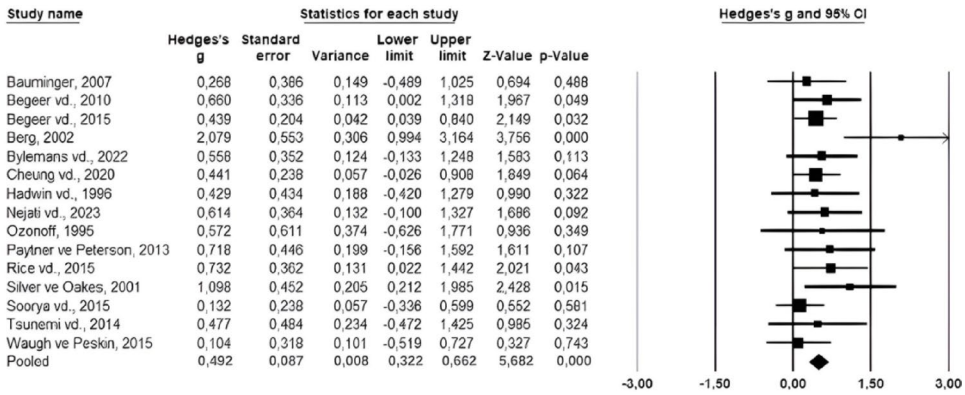


Fig. 2 Forest plot of theory of mind intervention studies. The forest plot displaying Hedges's g effect sizes and 95% confidence intervals for each included study was generated using the Comprehensive Meta-analysis (CMA) software

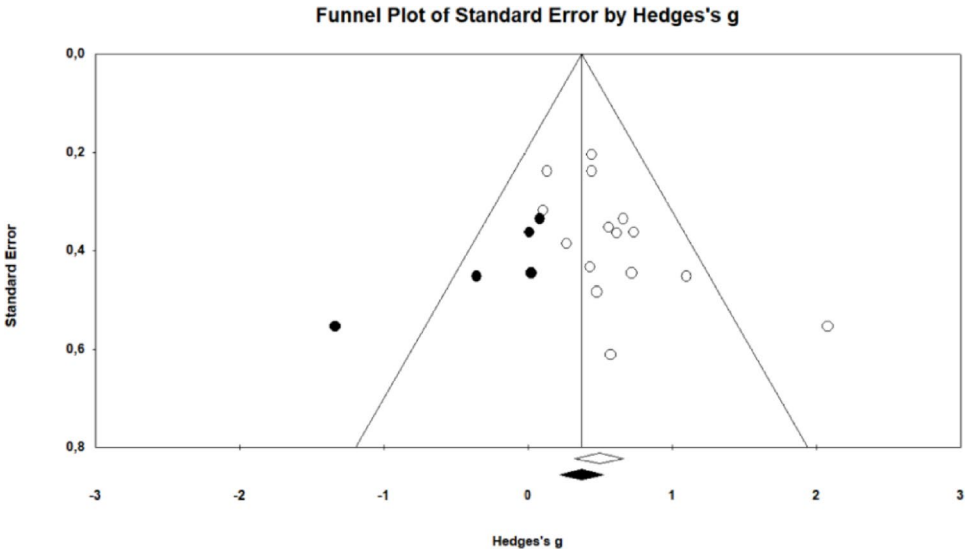


Fig. 3 Trim-and-fill funnel plot for included studies. The funnel plot illustrating the relationship between effect sizes (Hedges's g) and standard errors was generated using the Comprehensive Meta-Analysis (CMA) software

programs aimed at improving ToM skills have a moderate effect on the ToM skills of individuals diagnosed with ASD (0.2 < Cohen's d < 0.5 indicates a moderate effect)

Forest Plot of ToM Intervention

In the forest plot analysis (see Fig. 2), we found that all 15 included studies favored the experimental group. The considerable overlap in confidence intervals indicated a homogeneous structure. One study (Berg, 2002) showed a notably larger effect size (g = 2.07 [95% CI = 0.99, 3.16, p < 0.001]) than the overall confidence interval, suggesting a potential outlier. To examine whether this unusually large effect size influenced the overall findings, a sensitivity analysis was conducted by removing this study from the analysis. After excluding the study, the overall effect size remained similar (g = 0.45, 95% CI = 0.32–0.66; I^2 = 0.00) under the fixed-effects model. Additionally, the standardized residual analysis (SR = 2.90) suggested that the study did not substantially influence the overall results.

Publication Bias

We assessed publication bias via the fail-safe N method and the trim and fill method. For the 15 included studies, the threshold value was 85, and the Fail-Safe N was 130 (see Online Resource 6). We then used trim and fill analysis (Duval & Tweedie, 2000). As shown in Fig. 3, this analysis revealed that five studies contributed to publication bias. According to the fixed-effects model, the observed average effect size was g = 0.49 (95% CI [0.32, 0.66]; Q = 15.372), whereas the adjusted effect size obtained after the trim-and-fill analysis was calculated as g = 0.37 (95% CI [0.21, 0.54]; Q = 31.806); this suggests that the primary findings may have been subject to publication bias. The significant decrease in the adjusted effect size indicates that the original analysis was biased upward due to publication bias in the literature. This suggests that the "file-drawer problem"—where studies with statistically insignificant or small effect sizes are not reported—led to an overestimation of the effect size in our study. Furthermore, this upward bias may also stem from small studies with limited sample sizes (Borenstein et al., 2013; Cooper et al., 2009) and methodological limitations

in studies that do not meet WWC standards (WWC, 2022). Therefore, since the actual effect of ToM interventions in this study is likely lower than initially observed, the overall effect size should be interpreted with caution (see Online Resource 7).

Moderator Effects

We examined the effectiveness of ToM intervention programs in terms of research design, age group, delivery format, number of sessions, and duration (see Online Resource 8). Our analysis by research design revealed no significant difference between randomized controlled trials and quasi-experimental studies (Q_B = 0.13, $p = 0.71$). For age groups, the mean effect sizes were 0.72 for ages 5–7, 0.450 for ages 8–9, 0.91 for ages 10–13, and 0.55 for ages 14 and older, with no significant differences (Q_B = 7.815, $p = 0.56$). Regarding delivery format, individual interventions yielded a mean effect size of 0.678, while group-based intervention yielded 0.457, a non-significant difference (Q_B = 1.05, $p = 0.30$). Based on the number of sessions, we categorized interventions into 1–8 ($g = 0.66$), 9–16 ($g = 0.43$), and 17 or more sessions ($g = 0.551$), with no significant differences (Q_B = 1.29, $p = 0.52$). For session duration, interventions shorter than 30 min and those lasting 30 min or longer (g = 0.47) showed no significant difference (Q_B = 0.25, $p = 0.61$). Given the limited number of studies in each subgroup ($n = 15$), the results of these moderator analyses should be interpreted as exploratory rather than definitive. Similarly, statistically non-significant results may be due to analyses that are not sufficiently powerful to detect small differences between subgroups (Hedges & Pigott, 2004).

In addition to the variables included in the moderator analysis, we conducted an additional weighted regression analysis comparing studies that met WWC standards with those that did not, to investigate in more detail whether methodological quality assessed using WWC standards was related to effect size. The initial analysis showed a significant difference between the two groups ($\beta = -0.198$, $p = 0.022$). In the next step, assuming that this difference stemmed from an outlier study with a large effect size (Berg, 2002), the outlier study was removed, and the analysis was rerun. After this study was excluded, the difference between studies meeting WWC standards and those not meeting them no longer reached statistical significance ($\beta = -0.066$, $p = 0.457$).

Discussion

This study aimed to examine the effects of intervention programs designed to improve ToM skills on the ToM abilities of individuals diagnosed with ASD. To achieve this aim, we conducted a systematic review and meta-analysis of intervention programs targeting the enhancement of ToM skills in individuals with ASD. The findings of this study, calculated via a fixed-effects model, indicated that ToM intervention programs have a moderate effect on the ToM skills of individuals with ASD ($g = 0.49$). Although the meta-analysis indicates a moderate effect size, this finding should be evaluated in the context of the methodological rigor of the included studies. It was found that nearly half of the studies included in this analysis did not meet WWC standards or met them only with some reservations. This high proportion of methodologically limited studies indicates that the current "evidence-based" status of ToM interventions, as previously categorized by the NAC (2015), remains fragile. It demonstrates that higher-quality, rigorous experimental validation is needed for these interventions to be evaluated as evidence-based.

This result is lower than the high effect size ($g = 0.75$) reported in the meta-analysis of Hofmann et al. (2016). Differences in participant age groups, the scope of intervention programs, and the variety of assessment tools used may explain the discrepancy between the findings. Although the analysis results revealed that theory of mind interventions had a moderate effect on theory of mind skills, the fact that the interventions and dependent variables were very similar in most of the studies examined suggests that these high effect sizes may be context-dependent. The limited availability of generalisation and persistence data in the studies reviewed necessitates a cautious approach regarding the generalisability of these interventions to everyday life.

ToM skills are associated with core aspects of daily life such as social interactions (Hughes & Leekam, 2004; Preckel et al., 2018), friendships (Fink et al., 2015), empathy (Preckel et al., 2018), and communication (Sidera et al., 2018). Therefore, ToM intervention programs are suggested to have the potential to improve the social interaction and communication skills of individuals diagnosed with ASD. From an educational perspective, such programs could offer practical and functional benefits by enhancing the participation of individuals with ASD in social life. For example, in a study by Soorya et al. (2015), a ToM intervention program delivered to youth with ASD aged 8–11 years was evaluated by parents through questionnaires, and the results indicated moderate to high satisfaction. The severe underreporting of findings regarding generalizability and social validity in ToM intervention studies constitutes a significant methodological and clinical limitation. Without this data, the

practical significance of the observed moderate effect size remains largely theoretical. The essence of using social cognitive training with individuals with ASD is in the implementation of these skills in everyday life in order to cope with complex social situations. However, current research does not reveal a clear relationship between test results and actual changes in the lives of individuals and their relatives. Since ToM intervention programs are not yet fully recognized as evidence-based practices (NAC, 2015), systematically reporting generalization and social validity data would significantly strengthen practices in this field. Such data can play a fundamental role in developing interventions that support individuals with ASD in applying socially meaningful skills within real-life contexts.

Moreover, ToM skills are associated with other abilities such as joint attention (Sodian & Kristen-Antonow, 2015), emotion recognition (Mier et al., 2010), imitation (Meltzoff & Decety, 2003), symbolic play (Lam & Yeung, 2012), self-awareness (Vogeley et al., 2001), executive function (Ahmed & Stephen Miller, 2011), and working memory (Lecce et al., 2017). Structuring ToM intervention programs to support these related skills could potentially enhance the programs' overall effect size and promote the long-term maintenance of acquired skills.

In our study, the subgroup analyses revealed no significant differences in the effectiveness of ToM intervention programs based on variables such as research design, age group, delivery format, number of sessions, or session duration. This finding supports the homogeneity among the studies included in the meta-analysis and the suitability of the fixed-effects model. However, Hofmann et al. (2016) reported that intervention duration and session length significantly impact intervention effectiveness. The differences between the present study and the findings of Hofmann et al. may be attributable to methodological factors such as sample characteristics, intervention content, and the variety of methods and techniques employed. Therefore, future research should examine how sample characteristics, intervention content, and instructional methods and techniques influence intervention effectiveness. Such investigations inform the development of more practical applications that enhance the impact of ToM intervention programs on target populations.

In this meta-analysis, approximately half of the included studies employed RCTs, while the other half used quasi-experimental designs. RCTs use random assignment, allowing for causal inferences, whereas quasi-experimental designs do not include random assignment to experimental and control groups (Shadish et al., 2002). This limitation can result in nonequivalent groups at baseline and has contributed to some studies not meeting the WWC standards. Random assignment is not always feasible in applied studies involving individuals with special needs (Gersten et al., 2005). In interventions aimed at improving complex social-cognitive skills such as ToM in individuals with ASD, action research may be preferred, as it allows for the iterative refinement of interventions within a context-sensitive process (Baglin, 2007). This approach enables improvements through feedback from stakeholders involved in the intervention process, offering a new perspective on ToM interventions. Nevertheless, regardless of the design type, it is critical that researchers explicitly report how they ensure internal and external validity, as this transparency is essential for the accurate interpretation of findings (Miller et al., 2020).

Most reviewed studies were conducted with children aged 5–10 years, with only one study including adult participants. This finding aligns with the meta-analysis results of Lewis and van Schalkwyk (2020), who reported that intervention programs for individuals with ASD are typically designed for those under 18 years of age. Similarly, Wong et al. (2015) noted that interventions generally target preschool- and elementary-aged children, with relatively few studies focusing on adolescents. However, adolescents with ASD are known to experience considerable challenges in forming friendships (Petrina et al., 2016), initiating and maintaining social interactions (Schroeder et al., 2014), and developing communication skills (Humphrey & Lewis, 2008), all of which are closely related to ToM (Ma et al., 2023). These findings highlight the need for ToM-based social skills interventions targeting adolescent and adult populations. Expanding ToM-based interventions that foster social-cognitive development to include these age groups is essential for evaluating their effectiveness. Such interventions have the potential to function as practical tools in supporting the independent social lives of adolescents and adults with ASD.

An evaluation of the included studies regarding compliance with the WWC standards revealed that nine studies met the standards without reservations, eight did not, and three met the standards with reservations. The primary reasons for noncompliance were methodological limitations and baseline inequalities between the experimental and control groups. These disparities may stem from the inherent nature of the study design and the natural heterogeneity of individuals with ASD (Geschwind & Levitt, 2007; Grove et al., 2019). In this context, it is important to adopt a cautious approach when interpreting the effect sizes derived from studies that do not meet WWC standards, particularly concerning the generalizability of the findings. Designing studies that account for the individual differences in people with ASD may strengthen the methodological rigor of future intervention research.

Most of the studies included in this research were based on group-based instruction. The literature provides substantial evidence supporting the effectiveness of group interventions for individuals with ASD (Drüsedau et al., 2023; Krasny et al., 2003; Palmen & Didden, 2012). The present study also revealed that interventions that significantly contributed to the overall effect size (Begeer et al., 2011, 2015; Berg, 2002) were implemented in a group format. However, subgroup analyses indicated no significant difference in the interventions' impact when delivered individually or in groups. This finding aligns with the results of Tachibana et al. (2018). While group interventions may offer valuable social opportunities for skills such as ToM that require reciprocal interaction, it should also be acknowledged that individuals with ASD may not fully benefit from group activities owing to their limited holistic perspective (Happé et al., 2001). Therefore, future research could focus on hybrid intervention models that integrate elements of both group and individual instruction, tailored to the specific needs of participants.

The analysis revealed notable limitations in reporting social validity, maintenance, and generalization findings across the included studies. The fact that 80% of the studies did not report social validity data is a significant limitation. Without this information, the practical significance of the moderate effect size found remains debatable. Likewise, there is no data on the extent to which these skills are maintained and used in natural settings. In particular, the lack of examination of skill generalization to natural environments and the absence of maintenance assessments make it challenging to evaluate the long-term effects of intervention programs (Begeer et al., 2011). However, applying social-cognitive skills in everyday life is critical for enabling individuals to participate actively in social contexts. Therefore, systematically reporting social validity data in future research is essential to determine how effective intervention programs are in real-world settings.

It was also observed that generalization data were largely absent from the included studies. However, generalization—transferring acquired skills across different people, settings, and situations—is critical to skill development (Skinner, 1953). Research has shown that individuals with ASD often face difficulties in generalizing learned skills (Brown & Bebko, 2012). Given that ToM is a skill applied continuously across various social contexts, it is necessary to systematically plan opportunities for generalization during the development of ToM-focused interventions. For example, only one study in the present meta-analysis exposed participants to natural environments such as shopping malls and restaurants after the intervention, ensuring that the target skills were practiced in these settings (Ozonoff & Miller, 1995). Such practices represent important strategies for promoting the generalization of ToM skills and should be considered in future intervention designs.

Limitations and Recommendations

This study has several limitations. First, many of the included intervention programs targeted ToM and other related skills, which limits the ability to attribute the observed effect size exclusively to ToM. Future research should separate intervention components to enable meta-regression analyses that identify sources of variance more precisely.

Second, the current meta-analysis included only experimental and quasi-experimental studies, even though the literature search identified single-case designs. Incorporating single-case studies in future meta-analyses may expand the evidence base and provide insights from diverse methodological perspectives.

Third, an examination of the sample characteristics of the included studies—notably the average age of 11 and the use of tasks requiring higher-order cognitive abilities, such as second-order false belief and recursive thinking—suggests that the individuals in the sample possess high linguistic and cognitive skills. A significant limitation of this study is that IQ and language ability scores were not systematically reported in the synthesized studies, preventing analysis of the moderating effects of these variables on effect sizes. The exclusion of this data from the analysis process makes it difficult to generalize the moderate effect size to an autistic population with intellectual disabilities or significant language delays. Future research could address cognitive diversity across the spectrum more inclusively and examine the moderating effect on effect size by including participants with different IQ/language profiles in the meta-analysis.

Fourthly, within the scope of this research, it was observed that some studies included in the meta-analysis did not fully meet WWC standards due to initial inequalities between the experimental and control groups. In such situations, where group equivalence cannot be fully ensured through random assignment, validity issues may arise in studies with small sample sizes. Furthermore, in these studies, post-test results were reported without adjusting for pre-test differences between the experimental and control groups. As this may lead to an overestimation of effect sizes, the effect sizes reported in this meta-analysis should be interpreted with caution. Future primary studies involving interventions targeting theory of mind should ensure baseline equivalence between the experimental and control groups and report the intervention's effects using analyses that correct for pre-test differences based on post-test results.

Given the pronounced social and communication challenges in individuals with ASD, ToM interventions remain a promising and functional approach (2015; Hofmann et

al., 2016). Future research could integrate action research approaches (Huang, 2010) that combine practical improvement with theoretical knowledge, allowing for flexibly structured interventions tailored to the heterogeneous needs of this population.

Fifth, the moderator analyses conducted in this study based on variables such as age, intervention type, and session duration may lack sufficient statistical power due to the limited number of studies included in the meta-analysis ($n=15$). Therefore, the findings from these moderator analyses should be interpreted not as definitive conclusions, but as exploratory preliminary evidence for future studies. Future research could include a larger number of carefully designed studies in the analysis to increase statistical power and ensure a more reliable examination of potential moderator effects; it could also explore additional moderators, such as cognitive abilities and language skills, to better reflect individual differences within the autism spectrum.

Sixth, the failure to report or even plan for data on generalizability and social validity in the studies included in this review poses a major obstacle to assessing the long-term clinical benefits of ToM interventions. This dearth of research makes it difficult to know whether the skills learned through ToM interventions are retained across individuals and contexts, or translate into socially meaningful changes for participants' lives. The practical significance of the observed moderate effect size is therefore highly debatable, given that social validity data were reported in only 20% of the included studies ($n=4$) and generalizability to natural settings was examined in even fewer. Future work should focus on including multiple social validity indices and on generalizability to naturalistic settings to address the gap between experimental and real-world effectiveness.

Finally, limitations in the ability to understand others' mental states may increase the risk of experiencing bullying during early adolescence (Shakoor et al., 2012). A meta-analysis by Park et al. (2020) found that 67% of children with Autism Spectrum Disorder (ASD) experienced bullying. One factor that increases this risk in early adolescence is an inability to understand others' mental states adequately. It is important to note that the Theory of Mind (ToM) interventions examined in this study differed in their content. Some programmes primarily target skills such as emotion recognition or understanding false beliefs, while others emphasise interpreting others' intentions and thoughts in social interactions. Therefore, it is thought that designing ToM-based interventions that include perspective-taking and emotion recognition skills, in addition to focusing on interpreting others' intentions, may help reduce the risk of bullying in children with ASD. Furthermore, ToM skills can be indirectly incorporated into various preventive approaches, such as identifying bullying, understanding the bully's thoughts and intentions, and resisting bullying by considering the bully's perspective.

Conclusion

As demonstrated in this review, the positive effects of ToM intervention programs for individuals with ASD are supported by meta-analytic evidence. While the overall findings are positive, it should not be overlooked that nearly half of the included studies did not meet WWC standards or met them only with some reservations when interpreting the average effect size. Additionally, the significant lack of data on social validity and generalizability creates a major gap regarding the validity and practical application of these interventions in daily life. To strengthen our understanding of the effectiveness of ToM interventions for individuals with ASD, additional evidence from high-quality randomized controlled trials is needed.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10803-026-07393-2.

Acknowledgments The authors would like to thank Research Assistant Ece Yağcı AKGÜNDÜZ and Research Assistant Fadime ERKAN for their contributions in recoding the data to establish interrater reliability. The authors are also grateful to Assoc. Prof. Dr. Alev GİRLİ for sharing her valuable expertise on the Theory of Mind. During the preparation of this manuscript, the authors used [Gemini, Deeply, and Grammarly] in order to support translation and improve the language and readability of the text. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the final version of the manuscript. The first author developed this article from her master's thesis.

Author Contributions Conceptualization: Betül Kat Şen, Aydan Aydın. Methodology: Betül Kat Şen, Aydan Aydın, Data Curation: Betül Kat Şen. Formal Analysis: Betül Kat Şen. Writing – Original Draft: Betül Kat Şen, Aydan Aydın. Writing – Review & Editing: Betül Kat Şen, Aydan Aydın.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This research received no external funding.

Data Availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahmed, F. S., & Stephen Miller, L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*(5), 667–678. https://doi.org/10.1007/s10803-010-1087-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

Baglin, M. (2007). All you need to know about action research. *Journal of Advanced Nursing*, *59*(6), 654. https://doi.org/10.1111/J.1365-2648.2007.04363.X

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, *82*(6), 1887–1905. https://doi.org/10.1111/j.1467-8624.2011.01669.x

Baron-Cohen, S. (2000). Is asperger syndrome/high-functioning autism necessarily a disability? *Development and Psychopathology*, *12*(3), 489–500.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal Of Child Psychology And Psychiatry And Allied Disciplines*, *38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. Cognition, 21(1), 37–46. https://pubmed.ncbi.nlm.nih.gov/2934210/

**Bauminger, N. (2007). Brief report: group social-multimodal intervention for HFASD. *Journal of Autism and Developmental Disorders*, *37*(8), 1605–1615. https://doi.org/10.1007/s10803-006-0246-3

Bauminger, N., Solomon, M., & Rogers, S. J. (2010). Predicting friendship quality in autism spectrum disorders and typical development. *Journal of Autism and Developmental Disorders*, *40*(6), 751–761. https://doi.org/10.1007/s10803-009-0928-8

**Begeer, S., Gevers, C., Clifford, P., Verhoeve, M., Kat, K., Hoddenbach, E., & Boer, F. (2011). Theory of Mind training in children with autism: a randomized controlled trial. *Journal of Autism and Developmental Disorders*, *41*(8), 997–1006. https://doi.org/10.1007/s10803-010-1121-9

**Begeer, S., Howlin, P., Hoddenbach, E., Clauser, C., Lindauer, R., Clifford, P., Gevers, C., Boer, F., & Koot, H. M. (2015). Effects and moderators of a short theory of mind intervention for children with Autism Spectrum Disorder: A randomized controlled trial. *Autism Research: Official Journal of The International Society for Autism Research, 8*(6), 738–748. https://doi.org/10.1002/aur.1489

**Berg, H. (2002). P. Remedıatıng cognıtıve perspectıve-takıng ın chıldren wıth autısm (Publication No. 3052862) [Doctoral Dissertation, Columbia University]. ProQuest Dissertations & Theses Global.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). *Introduction to meta-analysis* (S. Dinçer, Trans.). Ankara: Anı Publishing. (Original work published 2009)

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, *18*(6), 906–921. https://doi.org/10.1037/0012-1649.18.6.906

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring theory of mind in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(7), 1927–1941. https://doi.org/10.1007/s10803-017-3080-x

Brown, J. R., & Dunn, J. (1991). You can cry, mum": The social and developmental implications of talk about internal states. *British Journal of Developmental Psychology, 9*(2), 237–256. https://doi.org/10.1111/j.2044-835X.1991.tb00874.x

Brown, S. M., & Bebko, J. M. (2012). Generalization, overselectivity, and discrimination in the autism phenotype: A review. *Research in Autism Spectrum Disorders, 6*(2), 733–740. https://doi.org/10.1016/j.rasd.2011.10.012

**Bylemans, T., Heleven, E., Baetens, K., Deroost, N., Baeken, C., & Van Overwalle, F. (2022). A narrative sequencing and mentalizing training for adults with autism: A pilot study. *Frontiers in Behavioral Neuroscience*, *16*, 941272. https://doi.org/10.3389/fnbeh.2022.941272

**Cheung, P. P. P., Brown, T., Yu, M. L., & Siu, A. M. H. (2020). The effectiveness of a school-based social cognitive intervention on the social participation of Chinese children with Autism. *Journal of Autism and Developmental Disorders, 51*(6), 1894–1908. https://doi.org/10.1007/s10803-020-04683-1

Cheung, P. P. P., Siu, A. M., Brown, T., & Yu, M. (2018). A social-cognitive intervention program for adolescents with autism: A pilot study. *Journal of Occupational Therapy Schools & Early Intervention*, *11*(1), 37–48. https://doi.org/10.1080/19411243.2017.1408442

Chiu, H. M., Chen, C. T., Tsai, C. H., Li, H. J., Wu, C. C., Huang, C. Y., & Chen, K. L. (2023). Theory of mind predicts social interaction in children with autism spectrum disorder: A two-year follow-up study. *Journal of Autism and Developmental Disorders*, *53*(9), 3659–3669. https://doi.org/10.1007/s10803-022-05662-4

Cohen, R. J., & Swerdlik, M. (2009). *Psychological Testing and Assessment* (7th ed.). McGraw-Hill Humanities/Social Sciences/Languages.

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine, 126*(5), 376–380. https://doi.org/10.7326/0003-4819-126-5-199703010-00006

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 1–615). Russell Sage Foundation.

*Dalmış, E. (2013). *Yüksek işlevli otizm ve asperger sendromu tanılı çocuklarda ileri derecede zihin okuma becerileri öğretimi programının etkililiğinin incelenmesi* [Examining the effectiveness of an advanced mindreading skills training program for children diagnosed with high-functioning autism and Asperger's syndrome] (Thesis No. 353284) [Master's thesis, Dokuz Eylül University]. Council of Higher Education Thesis Center.

*de Veld, D. M. J., Howlin, P., Hoddenbach, E., Mulder, F., Wolf, I., Koot, H. M., Lindauer, R., & Begeer, S. (2017). Moderating effects of parental characteristics on the effectiveness of a theory of mind training for children with autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, *47*(7), 1987–1997. https://doi.org/10.1007/s10803-017-3117-1

*de Veld, D. M. J., Scheeren, A. M., Howlin, P., Hoddenbach, E., Mulder, F., Wolf, I., & Begeer, S. (2020). Sibling configuration as a moderator of the effectiveness of a theory of mind training in children with autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, *51*(5), 1719–1728. https://doi.org/10.1007/s10803-020-04649-3

Deeks, J. J., Higgins, J. P. T., Altman, D. G., McKenzie, J. E., & Veroniki, A. A. (2024). Analysing data and undertaking meta-analyses.

In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (version 6.5). Cochrane. www.training.cochrane.org/handbook

Dettori, J. R., Norvell, D. C., & Chapman, J. R. (2022). Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global Spine Journal, 12*(7), 1624–1626. https://doi.org/10.1177/21925682221110527

Devine, R. T., & Apperly, I. A. (2021). Willing and able? Theory of mind, social motivation, and social competence in middle childhood and early adolescence. *Developmental Science*. https://doi.org/10.1111/desc.13137

Dodell-Feder, D., Felix, S., Yung, M. G., & Hooker, C. I. (2016). Theory-of-mind-related neural activity for one's romantic partner predicts partner well-being. *Social Cognitive and Affective Neuroscience*, *11*(4), 593–603. https://doi.org/10.1093/scan/nsv144

Drüsedau, L. L., Götz, A., Kleine Büning, L., Conzelmann, A., Renner, T. J., & Barth, G. M. (2023). Tübinger Training for Autism Spectrum Disorders (TüTASS): A structured group intervention on self-perception and social skills of children with autism spectrum disorder (ASD). *European Archives of Psychiatry and Clinical Neuroscience, 273*(7), 1599–1613. https://doi.org/10.1007/s00406-022-01537-y

*Durrleman, S., Bentea, A., Prisecaru, A., Thommen, E., & Delage, H. (2022). Training syntax to enhance theory of mind in children with ASD. *Journal of Autism and Developmental Disorders, 53*(6), 2444–2457. https://doi.org/10.1007/s10803-022-05507-0

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., & de Rosnay, M. (2015). Friendlessness and theory of mind: A prospective longitudinal study. *British Journal of Developmental Psychology, 33*(1), 1–17. https://doi.org/10.1111/bjdp.12060

*Fisher, N., & Happé, F. (2005). A training study of theory of mind and executive function in children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 35*(6), 757–771. https://doi.org/10.1007/s10803-005-0022-9

Flavell, J. H., Flavell, E. R., Green, F. L., & Moses, L. J. (1990). Young children's understanding of fact beliefs versus value beliefs. *Child Development, 61*(4), 915–928. https://doi.org/10.2307/1130865

Fletcher-Watson, S., McConnell, F., Manola, E., & McConachie, H. (2014). Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). *The Cochrane Database of Systematic Reviews, 2014*(3), Article CD008785. https://doi.org/10.1002/14651858.CD008785.pub2

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences, 7*(2), 77–83. https://doi.org/10.1016/S1364-6613(02)00025-6

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality ındicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149–164. https://doi.org/10.1177/001440290507100202

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: Developmental disconnection syndromes. *Current Opinion in Neurobiology, 17*(1), 103–111. https://doi.org/10.1016/j.conb.2007.01.009

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics, 51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

**Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? *Development and Psychopathology*, *8*(02), 345. https://doi.org/10.1017/s0954579400007136

Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences, 3*(6), 216–222. https://doi.org/10.1016/s1364-6613(99)01318-2

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*(2), 129–154. https://doi.org/10.1007/BF02172093

Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry, 55*(6), 553–577. https://doi.org/10.1111/jcpp.12162

Happé, F., Frith, U., & Briskman, J. (2001). Exploring the cognitive phenotype of autism: Weak central coherence in parents and siblings of children with autism: I. experimental tests. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(3), 299–307. https://doi.org/10.1017/S0021963001006916

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*(4), 426–445. https://doi.org/10.1037/1082-989X.9.4.426

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., Curtiss, J., & Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition, 150*, 200–212. https://doi.org/10.1016/j.cognition.2016.01.006

Hoogenhout, M., & Malcolm-Smith, S. (2014). Theory of mind in autism spectrum disorder: Does DSM classification predict development? *Research in Autism Spectrum Disorders*, *8*(6), 597–607.

Huang, H. B. (2010). What is good action research? Why the resurgent interest? *Action Research*, *8*(1), 93–109. https://doi.org/10.1177/1476750310362435

Hughes, C., & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development, 13*(4), 590–619. https://doi.org/10.1111/j.1467-9507.2004.00285.x

Humphrey, N., & Lewis, S. (2008). Make me normal': the views and experiences of pupils on the autistic spectrum in mainstream secondary schools. *Autism : the international journal of research and practice, 12*(1), 23–46. https://doi.org/10.1177/1362361307085267

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M. C., & Mandy, W. (2017). "Putting on My Best Normal": Social Camouflaging in Adults with Autism Spectrum Conditions. *Journal of autism and developmental disorders, 47*(8), 2519–2534. https://doi.org/10.1007/s10803-017-3166-5

Kaland, N., Møller-Nielsen, A., Callesen, K., Mortensen, E. L., Gottlieb, D., & Smith, L. (2002). A new 'advanced' test of theory of mind: Evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology And Psychiatry and Allied Disciplines*, *43*(4), 517–528. https://doi.org/10.1111/1469-7610.00042

Kobayashi Frank, C., & Temple, E. (2009). Cultural effects on the neural basis of theory of mind. *Progress in Brain Research, 178*, 213–223. https://doi.org/10.1016/S0079-6123(09)17815-9

Korkmaz, B. (2011). Theory of mind and neurodevelopmental disorders of childhood. *Pediatric Research, 69*(5 Pt 2), 101R – 1018. https://doi.org/10.1203/PDR.0b013e318212c177

Krasny, L., Williams, B. J., Provencal, S., & Ozonoff, S. (2003). Social skills interventions for the autism spectrum: Essential ingredients and a model curriculum. *Child and Adolescent Psychiatric Clinics of North America, 12*(1), 107–122. https://doi.org/10.1016/s1056-4993(02)00051-2

Lam, Y. G., & Yeung, S. S. S. (2012). Cognitive deficits and symbolic play in preschoolers with autism. *Research in Autism Spectrum Disorders, 6*(1), 560–564. https://doi.org/10.1016/j.rasd.2011.07.017

Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology, 163*, 69–86. https://doi.org/10.1016/j.jecp.2017.06.011

Lewis, A. S., & van Schalkwyk, G. I. (2020). Systematic review: Distribution of age and intervention modalities in therapeutic clinical trials for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 50*(6), 2208–2216. https://doi.org/10.1007/s10803-019-03942-0

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience and biobehavioral reviews, 80,* 729–742. https://doi.org/10.1016/j.neubiorev.2017.06.005

Livingston, L. A., Shah, P., & Happé, F. (2019). Compensatory strategies below the behavioural surface in autism: a qualitative study. The lancet. Psychiatry, 6(9), 766–777. https://pubmed.ncbi.nlm.nih.gov/31350208/

Ma, W., Mao, J., Xie, Y., Li, S., & Wang, M. (2023). Examining the effects of theory of mind and social skills training on social competence in adolescents with autism. *Behavioral Sciences (Basel, Switzerland), 13*(10), Article 860. https://doi.org/10.3390/bs13100860

Mao, S. Y., Che, H. M., Yu, Y. T., & Chen, K. L. (2023). The associations of theory of mind with both general and theory-of-mind-related social interaction in children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders, 102*, Article Article 102107. https://doi.org/10.1016/j.rasd.2023.102107

Mathersul, D., McDonald, S., & Rushby, J. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology*, *35*, 655–668. https://doi.org/10.1080/13803395.2013.809700

Mazza, M., Mariano, M., Peretti, S., Masedu, F., Pino, M. C., & Valenti, M. (2017). The role of theory of mind on social information processing in children with autism spectrum disorders: A mediation analysis. *Journal of Autism and Developmental Disorders, 47*(5), 1369–1379. https://doi.org/10.1007/s10803-017-3069-5

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London, 358*(1431), 491–500. https://doi.org/10.1098/rstb.2002.1261

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, *47*(6), 1028–1039. https://doi.org/10.1111/j.1469-8986.2010.01031.x

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications, Inc.

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech-Language Pathology*, *15*(2), 142–154. https://doi.org/10.1044/1058-0360(2006/014)

Miller, C. J., Smith, S. N., & Pugatch, M. (2020). Experimental and quasi-experimental designs in implementation research. *Psychiatry Research*, *283*, 1–7. https://doi.org/10.1016/j.psychres.2019.06.027

Miller, S. A. (2013). Children's understanding of second-order false belief: Comparisons of content and method of assessment. *Infant and Child Development, 22*(6), 649–658. https://doi.org/10.1002/icd.1810

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*(3), 603–613. https://doi.org/10.1348/026151005X55370

Mulrow, C. D. (1994). Rationale for systematic reviews. *BMJ, 309*(6954), 597–599. https://doi.org/10.1136/bmj.309.6954.597

Muris, P., Steerneman, P., Meesters, C., Merckelbach, H., Horselenberg, R., van den Hogen, T., & van Dongen, L. (1999). The TOM test: a new instrument for assessing theory of mind in normal children and childr en with pervasive developmental disorders. Journal of autism and developmental disorders, 29(1), 67–80. https://pubmed.ncbi.nlm.nih.gov/10097996/

National Autism Center (NAC). (2015). *National standards project findings and conclusions: Phase 2*. Author. https://www.autismdiagnostics.com/assets/Resources/NSP2.pdf

**Nejati, V., Khankeshlooyee, N., & Pourshahriar, H. (2023). Remediation of theory of mind in children with autism spectrum disorders: Effectiveness and transferability of training effects to behavioral symptoms. *Clinical Child Psychology and Psychiatry*, *29*(1), 259–273. https://doi.org/10.1177/13591045231208580

**Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: a new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders*, *25*(4), 415–433. https://doi.org/10.1007/BF02179376

O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of autism and developmental disorders, 39*(6), 916–928 https://doi.org/10.1007/s10803-009-0699-2

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Palmen, A., & Didden, R. (2012). Task engagement in young adults with high-functioning autism spectrum disorders: Generalization effects of behavioral skills training. *Research in Autism Spectrum Disorders*, *6*, 1377–1388. https://doi.org/10.1016/J.RASD.2012.05.010

Park, I., Gong, J., Lyons, G. L., Hirota, T., Takahashi, M., Kim, B., Lee, S. Y., Kim, Y. S., Lee, J., & Leventhal, B. L. (2020). Prevalence of and factors associated with school bullying in students with autism spectrum disorder: A cross-cultural meta-analysis. *Yonsei Medical Journal, 61*(11), 909–922. https://doi.org/10.3349/ymj.2020.61.11.909

**Paynter, J., & Peterson, C. C. (2013). Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *7*(2), 344–348. https://doi.org/10.1016/j.rasd.2012.10.001

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*(10), 519–525. https://doi.org/10.1016/j.tics.2012.08.004

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5-to 10-year-old

children. *Journal of experimental child psychology, 39*(3), 437-471. https://www.sciencedirect.com/science/article/abs/pii/0022096585900517

Petrina, N., Carter, M., Stephenson, J., & Sweller, N. (2016). Perceived friendship quality of children with autism spectrum disorder as compared to their peers in mixed and non-mixed dyads. *Journal of Autism and Developmental Disorders, 46*(4), 1334–1343. https://doi.org/10.1007/s10803-015-2673-5

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*(4), 973–982. https://doi.org/10.2307/1130869

Preckel, K., Kanske, P., & Singer, T. (2018). On the interaction of social affect and cognition: Empathy, compassion and theory of mind. *Current Opinion in Behavioral Sciences, 19*, 1–6. https://doi.org/10.1016/j.cobeha.2017.07.010

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

**Rice, L. M., Wall, C. A., Fogel, A., & Shic, F. (2015). Computer-assisted face processing ınstruction ımproves emotion recognition, mentalizing, and social skills in students with ASD. *Journal of Autism and Developmental Disorders*, *45*(7), 2176–2186. https://doi.org/10.1007/s10803-015-2380-2

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Schroeder, J. H., Cappadocia, M. C., Bebko, J. M., Pepler, D. J., & Weiss, J. A. (2014). Shedding light on a pervasive problem: A review of research on bullying experiences among children with Autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(7), 1520–1534. https://doi.org/10.1007/s10803-013-2011-8

Schwartz Offek, E., & Segal, O. (2022). Comparing theory of mind development in children with autism spectrum disorder, developmental language disorder, and typical development. *Neuropsychiatric Disease and Treatment, 18*, 2349–2359. https://doi.org/10.2147/NDT.S331988

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal ınference* (2 ed.). Cengage Learning.

Shakoor, S., Jaffee, S. R., Bowes, L., Ouellet-Morin, I., Andreou, P., Happé, F., Moffitt, T. E., & Arseneault, L. (2012). A prospective longitudinal study of children's theory of mind and adolescent involvement in bullying. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 53*(3), 254–261. https://doi.org/10.1111/j.1469-7610.2011.02488.x

Sidera, F., Perpiñà, G., Serrano, J., & Rostan, C. (2018). Why is theory of mind important for referential communication? *Current Psychology, 37*(1), 82–97. https://doi.org/10.1007/s12144-016-9492-5

**Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism: The International Journal of Research and Practice*, *5*(3), 299–316. https://doi.org/10.1177/1362361301005003007

Skinner, B. F. (1953). *Science and human behavior*. Macmillan.

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology*, *51*(9), 1190–1200. https://doi.org/10.1037/dev0000039

**Soorya, L. V., Siper, P. M., Beck, T., Soffes, S., Halpern, D., Gorenstein, M., Kolevzon, A., Buxbaum, J., & Wang, A. T. (2015). Randomized comparative trial of a social cognitive skills group for children with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(3), 208–216e1. https://doi.org/10.1016/j.jaac.2014.12.005

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Tachibana, Y., Miyazaki, C., Mikami, M., Ota, E., Mori, R., Hwang, Y., Terasaka, A., Kobayashi, E., & Kamio, Y. (2018). Meta-analyses of individual versus group interventions for pre-school children with Autism Spectrum Disorder (ASD). *PloS One, 13*(5), Article e0196272. https://doi.org/10.1371/journal.pone.0196272

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences, 28*(5), 675–735. https://doi.org/10.1017/S0140525X05000129

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-ınformed management knowledge by means of systematic review. *British Journal of Management, 14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

**Tsunemi, K., Tamura, A., Ogawa, S., Isomura, T., Ito, H., Ida, M., & Masataka, N. (2014). Intensive exposure to narrative in story books as a possibly effective treatment of social perspective-taking in schoolchildren with autism. *Frontiers in Psychology*, *5*, 2. https://doi.org/10.3389/fpsyg.2014.00002

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., & Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage, 14*(1 Pt 1), 170–181. https://doi.org/10.1006/nimg.2001.0789

Wang, Z., Devine, R. T., Wong, K. K., & Hughes, C. (2016). Theory of mind and executive function during middle childhood across cultures. *Journal of Experimental Child Psychology, 149*, 6–22. https://doi.org/10.1016/j.jecp.2015.09.028

Watson, J., Gelman, S. A., & Wellman, H. M. (1998). Young children's understanding of the non-physical nature of thoughts and the physical nature of the brain. *British Journal of Development Psychology*, *16*, 321–335.

**Waugh, C., & Peskin, J. (2015). Improving the social skills of children with hfasd: an ıntervention study. *Journal of Autism and Developmental Disorders*, *45*(9), 2961–2980. https://doi.org/10.1007/s10803-015-2459-9

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives, 11*(3), 207–214. https://doi.org/10.1111/cdep.12236

Wellman, H. M., & Estes, D. (1986). Early understanding of mental entities: A reexamination of childhood realism. *Child Development, 57*(4), 910–923. https://doi.org/10.1111/j.1467-8624.1986.tb00254.x

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., Phillips, A. T., & Rodriguez, T. (2000). Young children's understanding of perception, desire, and emotion. *Child Development*, *71*(4), 895–912. https://doi.org/10.1111/1467-8624.00198

What Works Clearinghouse (2022). What Works Clearinghouse procedures and standards handbook, version 5.0. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/Handbooks

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's

understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*(7), 1951–1966. https://doi.org/10.1007/s10803-014-2351-z

Zalla, T., Sav, A. M., Stopin, A., Ahade, S., & Leboyer, M. (2009). Faux pas detection and intentional action in Asperger Syndrome. A replication on a French sample. *Journal of Autism and Developmental Disorders, 39*(2), 373–382. https://doi.org/10.1007/s10803-008-0634-y

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*(2–3), 143–152. https://doi.org/10.1016/j.ijdevneu.2004.05.001

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.