HHS Public Access
Author manuscript
Cognition. Author manuscript; available in PMC 2016 May 01.

Published in final edited form as:
Cognition. 2016 May ; 150: 200–212. doi:10.1016/j.cognition.2016.01.006.

Training children's theory-of-mind: A meta-analysis of controlled studies

Stefan Hofmann[a], **Stacey N. Doan**[b], **Manuel Sprung**[c], **Anne Wilson**[c], **Chad Ebesutani**[d], **Leigh Andrews**[a], **Joshua Curtiss**[a], and **Paul L. Harris**[e]

Stacey N. Doan: sdoan@cmc.edu; Manuel Sprung: manuel.sprung@gmail.com; Anne Wilson: anne.wilson80@gmail.com; Chad Ebesutani: ebesutani@duksung.ac.kr; Paul L. Harris: paul_harris@gse.harvard.edu

[a]Department of Psychological and Brain Sciences, Boston University, 648 Beacon Street, 6th Floor, Boston, MA 02215, U.S.A

[b]Department of Psychology, Claremont McKenna College, 850 Columbia Ave., Claremont, CA, U.S.A., 91711

[c]International Psychology & Psychotherapy Center, Vienna (Austria)

[c]Department of Psychology, Ohio State University, 225 Psychology Building, 1835 Neil Avenue, Columbus, Ohio 43210, U.S.A

[d]Department of Psychology, Duksung Women's University, 33, Samyangro 144-gil, (419 Ssangmun Dong), Dobong Gu, Seoul 132-714, South Korea

[e]Graduate School of Education, Harvard University, 503A Larsen Hall, Appian Way, Cambridge, MA 02138, U.S.A

Abstract

Background—Theory-of-mind (ToM) refers to knowledge and awareness of mental states in oneself and others. Various training programs have been developed to improve ToM in children.

Objectives—In the present study, we conducted a quantitative review of ToM training programs that have been tested in controlled studies. DATA SOURCES: A literature search was conducted using PubMed, PsycInfo, the Cochrane Library, and manual searches.

Review Methods—We identified 32 papers with 45 studies or experiments that included1529 children with an average age of 63 months (SD= 28.7).

Results—ToM training procedures were more effective than control procedures and their aggregate effect size was moderately strong (Hedges'g = 0.75, CI = 0.60 - 0.894, p<.001). Moderator analyses revealed that although ToM training programs were generally effective, ToM skill-related outcomes increased with length of training sessions and were significantly higher in active control studies.

Conclusion—ToM training procedures can effectively enhance ToM in children.

[*]Corresponding author : Stefan G. Hofmann, Ph.D., Department of Psychological and Brain Sciences, Boston University, 648 Beacon Street, 6th Fl, Boston, MA 02215, Tel: 617 353 9233, Fax: 617 353 9609, shofmann@bu.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Keywords

Theory of Mind; Training; Intervention; Meta-Analysis; False Beliefs; Appearance- Reality

The ability to interpret and understand behavior in the terms of psychological states is crucial for successful social interactions. Theory-of-mind (ToM) refers to knowledge and awareness of mental states, (perceptions, emotions and thoughts) in oneself and others. Initial research on ToM focused extensively on the ability to understand that someone can believe something that is not true (i.e., hold a false belief), finding that this ability emerges around 4 years of age. More recently, researchers have adopted a broad conceptualization of ToM as a series of understandings which progresses from understanding simple desires and emotions to more complex mental states (for reviews see Flavell, 2004 and Wellman, 2014). Under this conceptualization, evidence suggests that desire reasoning is readily seen in children as early as 2-years of age, before the ability to reason about cognitions (notably false beliefs) (Gopnik & Wellman, 1992, 1994; Wellman, Cross & Watson, 2001). This trajectory appears robust and relatively stable across cultures (Liu, Wellman, Tardif & Sabbagh, 2008). It is also consistent with children's use of mental state language (Bartsch & Wellman, 1995), which initially focuses on desires and emotions and only later gradually advances to include references to thoughts and beliefs. (Note, however, that using non-verbal tasks, false belief understanding can be detected in the second year of life (Baillargeon, Scott & He, 2010; Onishi & Baillargeon, 2005).

Although the majority of studies on ToM are during the preschool period, there is evidence to suggest that ToM abilities continue to develop continues to develop in middle childhood and early adolescence (Devine & Hughes, 2013; Grazzani & Ornaghi, 2012). ToM advances to include the ability to consider more complex psychological states such as double bluffs, and second order false beliefs (making inferences about someone's false attribution of belief).

In line with this perspective, the specific type of task used to measure ToM abilities varies as a function of child age. In the younger years, researchers have employed tasks that tap into the understanding of diverse desires, the appearance reality distinction, perceptual and perspectival understanding, as well as false beliefs. In middle childhood and early adolescence, researchers have often employed tests that assess second-order false beliefs, as well as the ability to make inferences about double bluffs and misunderstandings. More recently however, instead of any single tasks, among preschoolers and reflecting the broader conceptualization of ToM, researchers have often employed a battery of ToM abilities, which allows researchers to score ToM on more of a continuum (Sabbagh, Xu, Carlson, Moses & Lee, 2006; Wellman, Fuxi & Peterson, 2011).

Delays in the development of ToM skills have been documented among children with various clinical conditions (for reviews see: Corcoran, 2000; Sprung, 2010) including deaf and hard-of-hearing children born to hearing parents, suggesting the importance of environmental factors, notably language and verbal interactions (Stanzione & Schick, 2014). There are also considerable individual differences among typically developing children in ToM skills (Pons & Harris, 2005; Repacholi & Slaughter, 2003) that are likely due to both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

individual characteristics as well as the social contextual environment. With regard to the former, language (Astington & Baird, 2005; de Villiers, 2005; Milligan, Astington & Dack, 2007) and executive function (EF; Benson & Sabbagh, 2010; Carlson, 2005; Devine & Hughes, 2014; Hughes & Graham, 2002) are thought to be important factors. Successful completion of ToM tasks often requires comprehension of mental state terms, of complex syntax (specifically understanding of sentential complement clauses), working memory and the ability to inhibit a prepotent response, the latter two being important aspects of executive function.

With regard to the social and contextual environment, observational studies have demonstrated that family size, namely the number of siblings (McAlister & Peterson, 2007; Perner, Ruffman, & Leekam, 1994), socioeconomic status (Cutting & Dunn, 1999), and conversational discourse (Astington & Baird, 2005; de Rosnay & Hughes, 2006) are powerful factors that influence children's ToM abilies. ToM skills are associated with a variety of positive outcomes and behaviors, including social skills (e.g., Banerjee & Yuill, 1999; Banerjee, Watling & Caputi, 2011; Bosacki & Astington, 1999; Cassidy, Parke, Butkovsky & Braungart, 1992; McDowell, O'Neil & Parke, 2000; Slaughter, Dennis & Pritschard, 2002; Watson, Nixon, Wilson & Capage, 1999), and better academic performance (e.g. Doudin, Martin & Albanese, 2001; Jones, Brown & Aber, 2011; Jones, Brown, Hoglund & Aber, 2010; Lecce, Caputi & Hughes, 2011; Pons, Doudin, Harris & de Rosnay, 2002). However, ToM skills have also been linked to changes in internalizing problems, such as increases in levels of intrusive thoughts, worry and rumination (Rieffe, Oosterveld, Miers, Terwogt & Ly, 2008; Rieffe et al., 2007; Sprung, 2008; Sprung & Harris, 2010) and increased sensitivity to criticism (Cutting & Dunn, 2002; Lecce et al., 2011).

A sizable number of studies have been conducted with the aim of improving children's ToM skills. Demonstrating the ability to enhance ToM via training would have three main implications. First, if ToM can be accelerated through training, this would reinforce the claim that individual differences in ToM skills are not merely the result of maturational changes or broad developments in cognitive processing, but are certainly associated with, and may be influenced by the specific type of social-environmental input that children receive. Second, if particular types of training are more effective than others, such findings may cast light on competing theoretical models of theory-of mind-development (e.g., Brockmeyer, 2010; Guajardo & Watson, 2002; Slaughter & Gopnik, 1996). For example, it might be possible to show that specific types of language training (Milligan et al., 2007), or exposure to certain types of pretend play (Lillard et al., 2012) are particularly effective. Alternatively, the existence of equifinality might be demonstrated, whereby a number of different interventions lead to a similar level of ToM mastery. Finally, successful ToM training offers a way to assess the impact of ToM on other domains, such as language, executive function and social skills (e.g., Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Peskin & Astington, 2004; Sprung, Lohmann, Zauner & Bauer, 2011; Sprung, Ristic, Fritz & Markova, 2012).

ToM training studies have built upon observational research suggesting that exposure to certain socio-linguistic contexts is associated with better ToM abilities. Thus in training studies, children are often exposed to language characteristics that have been associated with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

better ToM abilities, particularly explicit exposure to mental state language (e.g., Guajardo & Watson, 2002; Ornahgi, Brockmeier & Gavazzi, 2011). Several studies, for example, have engaged children in conversations with rich discussion of psychological states and explicit labeling of desire, perceptual and cognitive states (Slaugher & Gopnik, 1996; Appleton & Reddy, 1996). Others have focused on the importance of sentential complement constructions (sentences that take a full clause as their object complement, e.g., “Stephen thinks the ball is in the brown box.”) (Hale & Tager-Flushberg, 2003; Lohman and Tomasello, 2003) and still others have encouraged children to reason about alternative perspectives or mental states (e.g., Zhang, Wu & Zheng, 2008; Sprung et al., 2011; Sprung et al., 2012).

Despite the substantial number of controlled training studies that have been conducted, no quantitative review of their impact on ToM skill development has yet been conducted. The purpose of this paper was to conduct a meta-analysis of training studies designed to facilitate children's ToM skills. We sought to evaluate the efficacy of ToM training on children's ToM skills and to identify moderators that might influence the effectiveness of ToM training programs. Specifically, we investigate age, gender, and the quality of the intervention as potential moderators.

Methods

Search

Studies were identified by searching PsycInfo, PubMed, and the Cochrane Library. Search methods were specified in advance (see below) and inclusion criteria required that studies be controlled studies. Other selection criteria are listed below (see “Study Selection” section). We conducted searches for studies published between 1977 (i.e., the first available year) and November 06, 2015, using the following search term combination: (“*theory of mind*” OR “*emotion understanding*” OR “*cognitive empathy”*) AND *children* AND (*training* OR *teaching* OR *intervention* OR *program* OR *therapy*). Additionally, an extensive manual review of reference lists of relevant studies and review articles extracted from the database searches was conducted. Articles determined to be related to the topic of ToM or emotion understanding interventions/trainings were selected for further examination. To protect against publication bias, unpublished dissertations were included in the initial search. In addition, the authors of studies selected for inclusion in the meta-analysis were contacted for unpublished studies (and the first author's own unpublished studies were added).

Study Selection

Studies were included if the training program: (a) was designed to improve ToM capabilities;(b) included one or more pre- and post-test measure of ToM; (c) included child samples; (d) included a control group; and (e) provided sufficient data to perform effect size analyses (i.e., means and standard deviations, F values, t values, or change scores). If all criteria except for criterion (e) were met, the authors were contacted for additional data.

Figure 1 details our study selection process. Of the initial 757 articles initially identified, 32 were included in the present meta-analysis. The majority of excluded studies either did not

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contain any type of training (n=338) or were review articles (n=176). We decided to exclude emotion understanding training studies (n=24) because the outcome measures and training procedures for emotion understanding and ToM were not comparable to the ToM outcomes that were aggregated and examined in the present meta-analysis.

The remaining 32 articles were comprised of 45 studies, experiments, or comparisons. Of the studies included in the meta-analysis, all except 10 studies used random assignment of participants to experimental and control group. Of the 10 studies that we judged to be based on non-random assignment, 6 studies were “group matched”, often by classroom, and groups were found to not differ with respect to IQ/language. In three of the studies, “quasi” random assignment was used (i.e., children were randomly assigned after matching for IQ, language and ToM). In the sixth study deemed ‘non-random,’ it was not possible to determine how participants were assigned to groups, and so we conservatively assumed that assignment was not random. In studies with multiple training or control conditions, we compared the most complete training condition (i.e., the training condition that the authors hypothesized to be the most effective) to the most active, but neutral control condition (i.e., the condition that controlled for the nonspecific aspects of training, such as amount of time with the experimenter and general discussion). In cases where the author included multiple groups in order to analyze the contribution of different training components to ToM improvement (i.e., exposure to sentential complements vs. discussion of mental states vs. exposure to deceptive experience), we compared each of these training groups to the neutral control group.

Study Characteristics

Table 1 provides a description of the characteristics of the studies included in the present meta-analysis. Studies were conducted in 10 different countries (Austria, Canada, China, Singapore, Germany, Netherlands, Italy, UK, Spain, and USA) and included a total of 1529 children. The age of participants ranged from 31 to 194 months, with an average mean age of 63 months (SD= 32.14; median=46.65). The number of training sessions ranged from 1-32 sessions (mean=7.72, SD=7.36; median=13), with individual session length between 5 and 90 minutes (mean=23.01, SD=19.60; median=10), and overall training periods spanning from 1 to 189 days (mean=32.5, SD=39.73; median=33.5). Total Jadad scores ranged from 2 to 9, with a median of 5 (M=5.47, SD=1.69).

Data Abstraction

We extracted data from tasks designed to measure children's ToM capabilities. These measures include the false belief task (e.g., Wimmer & Perner, 1983), the appearance-reality task (e.g., Flavell, Flavell, & Green, 1983), and ToM batteries (e.g., Wellman & Liu, 2004). Some studies administered multiple tasks and reported the aggregate score for all of the tasks combined. Whenever possible, we extracted data separately for each ToM measure. If the study reported an aggregate ToM score, we contacted the author to request the scores for each individual measure. We categorized each measure according to which ToM domain it measured (i.e., false belief, appearance-reality, aggregate ToM). Additionally, we extracted numerical and categorical data for the purposes of conducting moderator analyses. If a study

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

did not report one or more variables targeted in our moderator analyses, we requested these data from the author.

To explore variability in study results, we examined whether effect size estimates differed depending on the methodological quality of the studies. We evaluated the design of each study to determine its relative quality using the following modified Jadad criteria (Jadad et al., 1996): (a) the study's control group participated in some sort of activity (1 point), (b) the study's training and control groups were matched according to language ability (1 point), (c) the tasks used during training differed from the tasks used at pre- and post-test (Different domain: 4 points; Additional domain: 3 points, Same domain, Different task: 2 points; Same domain, Similar task: 1 point; Same domain, Identical task: 0 points), and (d) the post-test measure differed from the pre-test measure (Different domain: 4 points; Additional domain: 3 points; Same domain, Different task: 2 points; Same domain, Similar task: 1 point; Same domain, Identical task: 0 points).

Quantitative Data Synthesis

We calculated pre-post treatment effect sizes for continuous measures of ToM using Hedges' *g* (a variation of Cohen's d that corrects for biases due to small sample size [Hedges & Olkin, 1985]) and its 95% confidence interval. Due to the small sample sizes in some of the studies, we calculated effect size estimates using a random-effects model (Hedges & Vevea, 1998). We followed the recommendation by Rosenthal (1993) and assumed a conservative estimate of $r = 0.7$ in all cases where the correlation between pre- and post-treatment measures was unavailable, which was every article extracted for this analysis.

We pooled effect sizes across studies to obtain a summary, between-group effect size. To assess the success of different training regimens according to specific components involved in training, we obtained a summary within-group effect size. The magnitude of Hedges' *g* corresponds to Cohen's recommendations for interpreting effect sizes as small (0.2), medium (0.5), and large (0.8).

Risk of Bias

We assessed the risk of publication bias by calculating the *fail-safe N* to estimate the number of unpublished studies with an effect size of zero needed to nullify the significant effect (Rosenthal, 1991; Rosenthal & Rubin, 1988). The effect size can be considered robust if the required number of studies to reduce the overall effect size to a nonsignificant level exceeds $5K + 10$, where K is the number of studies included in the analysis (Rosenthal, 1991). We completed all analyses using the software program Comprehensive Meta-Analysis, Version 2 (Borenstein, Hedges, Higgins, & Rothstein, 2005). In addition, because of the tendency of the fail-safe N to be upwardly biased, we constructed a funnel plot to examine the presence of publication bias. The absence of a publication bias results in a funnel plot that is symmetrical, with studies equally dispersed around either side of the mean effect size. Duval and Tweedie's ‘Trim and Fill method’ examines where missing studies are likely to fall to make the plot symmetrical, and adds them to the analysis in order to recalculate the effect size estimate.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Moderator Analyses

We conducted meta-regression analyses to examine whether the effect sizes varied as a function of the following variables: length of each training session, number of training sessions, total length of the training period, post-test delay (i.e., the amount of time between the last day of training and the post-test), total sample mean age, continent of origin, gender distribution, and modified Jadad scores (Jadad et al., 1996).

Results

Quantitative Data Synthesis

Between-Group Effect Size—The average between-group effect size estimate (Hedges' *g*) based on the 45 comparisons was 0.75, (*CI* = 0.60 - 0.89, $p<.001$, Q = 98.89, $I^2 = 55.51$) for improving ToM skills. Table 2 illustrates the effect size data for all comparisons included in the overall analysis.

Within-Group Effect Size—The average within-group effect size estimate (Hedges' *g*) based on the 44 (one article did not provide sufficient information) comparisons was 0.93(95% *CI* [0.79, .97], $p<.001$, Q = 361.22, $I^2 = 84.77$) for improving ToM skills.

Publication Bias—Based on the mean overall pre-post effect size ($z = 14.79$), our results indicated that 2519 studies with an effect size of zero would be needed to nullify the significant effect. Because the *fail-safe N* (1,208) exceeds 5*K* + 10, our results can be considered “robust” regarding the effectiveness of ToM programs in improving ToM skills. The funnel plot (Figure 2) depicts observed (white circles) and imputed (black circles) studies. The Trim and Fill method suggests that only 1 study would need to fall to the right of the mean effect size in order to make the plot symmetric, suggesting a conservative estimate. The Egger's regression intercept was not significant (intercept = 1.567, 2-tailed p = 0.09), suggesting that the parameter estimates were not influenced by the number of studies. Under the random effects model, the imputed mean effect size is Hedges' $g = 0.72$ ($CI_{95\%}$: 0.62, .81).

Moderator Analyses

Table 3 outlines the results of the meta-regression analyses conducted to assess the moderating effects of the number of training sessions and the length of each session (measured in minutes). The results of our analyses indicate that Hedges' *g* was moderated by the length of each session ($\beta = -0.01$, $SE = 0.00$, $p = 0.02$) and the length of the training period ($\beta = 0.01$, $SE = 0.00$, $p = 0.03$), as well as marginally by the number of sessions ($\beta = -0.02$, $SE = 0.00$, $p = 0.07$), and the gender distribution of the sample ($\beta = -0.91$, $SE = 0.60$, $p = 0.06$. Longer sessions and the number of sessions was associated with higher impact of the training, while longer training periods decreased the impact. Interventions also seem marginally more effective for females. Effect sizes were not moderated by primary continent status (i.e., Europe vs. North America) ($Q_B = 2.93$, $df = 1$, p = n.s.). We examined other variables as potential moderators, including *training characteristics* (i.e. post-test delay) and sample mean age. However, none of these variables significantly moderated the effect size. Also, Hedges' *g* was not moderated by *study year, publication status,* or *Jadad* scores.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hedges' *g* was also not moderated by the publication status of the article (Published Hedges' *g*= .942, Hedges' Unpublished *g* = .882, Q_B = .129, p = .719).

Discussion

The results show that, overall, ToM training procedures were effective in improving children's ToM skills (compared with a control group), with a large (Hedges' *g*) average effect size of .747. The overall effect size was also moderated significantly by length of session and length of training period and marginally by number of sessions and gender. Longer sessions and shorter training periods were associated with better outcomes, as marginally were more female samples and greater numbers of sessions. None of the sample characteristics moderated the overall effect size, suggesting that it is possible to improve ToM skills across children of different ages genders. We also found that neither study year nor Jadad scores moderated the effect size.

Scrutiny of the training studies that were analyzed shows that in most cases children were presented with situations (in the form of stories, picture books, videos, etc.) that are similar to ToM tasks. Common training procedures include corrective feedback, use of the imagination, modeling, and role-play. Some of the studies (particularly those with clinical samples) included ToM training as part of a broader program aiming to teach social skills. However, all the training procedures prompted children to reason about alternative perspectives or mental states in one way or another. Some studies did so by engaging children in explicit discussion of mental states. Others did so by putting children in situations that implicitly required them to adopt an alternative perspective or mental state (in most cases that of a story character). For instance, in a study by Peskin and Astington (2004) parents, teachers and graduate assistants read several picture books to children over period of 4-weeks. Participants were randomly assigned to one of two groups. One group received books with text rich in explicit mental state language. The other group received the same books with no mental state language but with most stories and accompanying illustrations implicitly requiring children to reason about mental states (i.e., to think about alternative perspectives).

The effectiveness of using language to draw children's attention to mental states, and more particularly alternative perspectives is fully consistent with studies of individual differences in the acquisition of ToM among typically developing children. Specifically, past studies have shown that children's language ability (Astington & Baird, 2005; Harris et al, 2005; Hughes et al., 2005; Lillard & Kavanaugh, 2014), especially in the context of perspective-taking that is needed for everyday conversation (Deleau, 2012), and children's involvement in pretend play, particularly pretend role play in which children enact another person's perspective (Astington & Jenkins, 1995; Harris, 2000, 2005; Lillard & Kavanaugh, 2014; Taylor, Carlson, Maring, Gerow, & Charley, 2004; Youngblade & Dunn, 1995) are associated with ToM abilities. In addition, a large body of research has shown that the extent to which caregivers attend to and discuss mental states has an impact on ToM development (Meins et al., 2002; Ruffman, Perner, & Parkin, 1999) as does social interaction with children of different ages (Cassidy, Fineberg, Brown, & Perkins, 2005; Cutting & Dunn, 1999; Ruffman, Perner, Naito, Parkin, & Clements, 1998; Wang & Su, 2009).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Typical ToM interventions tend to be one-on-one (i.e., individual format) and last from 15 to 20 minutes. By implication, intervention sessions can be effective despite their brevity. Nevertheless, it may be fruitful for future work to conduct more systematic examinations to better understand the constituent components of ToM interventions. One promising methodology is based on the ‘distillation and matching’ model (Chorpita, Daleiden, & Weisz 2005). Through these ‘distillation’ procedures, discrete common components across effective treatment protocols can be ‘distilled’ and identified. This ‘distillation and matching model’ has recently been applied to several youth problem areas, and has found, for example, that the discrete common practices across all effective youth treatments for autism included communication skills, modeling, social skills training, and goal setting (Chorpita & Daleiden, 2007). The discrete treatment components identified through this ‘distillation and matching’ framework have then been used as the foundation for a modular approach to therapist training and treatment delivery that has not only been shown to be effective with respect to reducing youth behavioral and emotional problems (Weisz et al., 2012), but also associated with more positive attitudes and greater satisfaction among therapists who use this approach (Borntrager et al., 2009). In the current meta-analyses, we found that studies often have multiple discrete practices – this make it difficult to identify which aspect may have the largest impact. Thus, future studies on training should consider a modular approach which would allow an application of this model to ‘distill’ and identify the constituent parts of ToM interventions that is most likely to have an impact. Moreover, it would also allow researchers to test the idea that certain types of interventions may be more beneficial as a function of participant characteristics.

There were some limitations to the present meta-analysis that are worth noting. First, several training studies aimed at improving ToM were excluded because they did not include a control group—several of these studies examined clinical samples. In these cases, a control group may have been excluded because of ethical concerns. Without a (waitlist) control group, it is difficult to separate the effects of training procedures from changes that occur in the course of normal development. Studies evaluating the effectiveness of interventions in clinical settings often include a waitlist control group, in which the waitlist control participants do not receive any active intervention *during* the study (but do so after the study is completed). This type of waitlist design can appease ethical concerns and also allows researchers to parse outchanges due to the training and changes due to the passage of time and natural developmental changes. However, in waitlist control study designs, it is still difficult to tease apart whether changes seen in the active treatment group were due to the specific training components or simply due to other general treatment factors (such as therapist contact and attention). Therefore, it will be important for future training studies to include, at a minimum, attention-control groups, which typically involve an activity that is similar in complexity and interaction to the active treatment, but does not include the critical, “active” elements of training.

Another limitation of the current meta-analysis is that most studies failed to assess general cognitive functioning (IQ), executive functioning and language ability. As a result, we were unable to examine the role of these variables as moderators, despite theoretical proposals suggesting that these may be important factors. The inclusion of these assessments – both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

before and after training – would help to determine if the benefits of ToM training extend to these other cognitive abilities, as well as the extent to which intervention effects might interact with variation in these abilities. For example, a recent study demonstrated that individual differences in executive functioning consistently predicted children's ability to benefit from a ToM training (Benson, Sabbagh, Carlson & Zelazo, 2013).

Relatedly, there was not an adequate number of studies to examine other potentially important moderators of the effectiveness of ToM training, including the type of training intervention, the training setting (e.g., classroom vs. laboratory), the training format (e.g., group vs. individual sessions), and sample type (e.g., typically developing children vs. children with Autism or another clinical diagnosis impairing ToM). Over the next decade of research in this area, attention should be given to these factors to determine whether they differentially affect ToM outcomes. Further, none of the studies provided information about transfer of knowledge (and skills) acquired in training to relevant knowledge (and skills) in children's everyday life. Finally, none of the existing training studies provided information about the durability of training effects. Most of the studies assessed training effects within a few days (max. 13 days) from the last training sessions. In sum, alongside the robust evidence of intervention effects, the present review highlights ways in which future studies might usefully probe the scope of those effects.

It could be argued that the null effects (e.g., non-significant moderator effects) might represent a Type II error. However, the fail-safe N analyses that we conducted suggest that the effect sizes were robust and, therefore, unlikely to be the result of a Type II error (Rosenthal, 1991; Rosenthal & Rubin, 1988). Moreover, it could be argued that clinical populations, such as children with ASD or hearing impairment, should have been excluded and the analysis should have focused exclusively on typically developing children. However, we saw no a priori justification for excluding individuals with clinical diagnoses. Excluding studies from a meta-analysis simply because they introduce some degree of heterogeneity is methodologically problematic, and so we opted to retain studies that included clinical populations (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006).

Given that ToM skills are associated with a variety of positive outcomes (e.g., better social skills, better academic performance), the results of the present meta-analysis highlight both the importance and feasibility of enhancing children's ToM skills with training programs. The moderate effect size associated with these ToM training programs also speaks to the potential utility of ToM training programs for *preventative* purposes, emulating the small number of preventative intervention programs developed in recent years (Bierman, Coie, Dodge, Greenberg, Lochman, McMahon & Pinderhughes, 2010; Jones, et al., 2011; 2010). Indeed, these preventative intervention programs include several elements that are common to the ToM training procedures examined in the present meta-analysis, further increasing the likelihood that such interventions may be effective. As preventative interventions grow, future meta-analytic studies, such as the one reported in the present paper, will be important to systematically evaluate their overall effectiveness and robustness. Finally, to the extent that the ability to report and reflect on one's thoughts and feelings is fundamental to the identification and treatment of many psychological problems, theory-of-mind training holds considerable promise as a tool that might improve diagnostic accuracy and clinical intervention.

Acknowledgments

This research was supported in part by a grant from the Austrian Science Foundation (i.e., Erwin- Schrodinger Fellowship Abroad to Manuel Sprung, Project number: J2853-G16) and NIH grant R01AT007257 and R34MH099311 to Stefan G. Hofmann.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Appleton M, Reddy V. Teaching three-year-olds to pass false belief tests: A conversational approach. Social Development. 1996; 5:275–29. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/j.1467-9507.1996.tb00086.x.

Astington, JW.; Baird, JA. Why language matters for theory of mind. New York, NY: Oxford University Press; US; 2005. Why language matters for theory of mind; p. xii, 355

Astington JW, Jenkins JM. Theory of mind development and social understanding. Cognition and Emotion. 1995; 9:151–165. doi: http://dx.doi.org/10.1080/02699939508409006.

Baillargeon R, Scott RM, He Z. False-belief understanding in infants. Trends in Cognitive Science. 2010; 14(3):110–118. doi: http://dx.dio.org/10.1016/j.tics.2009.12.006.

Banerjee R, Watling D, Caputi M. Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. Child Development. 2011; 82:1887–1905. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/j.1467-8624.2011.01669.x. [PubMed: 22023260]

Banerjee R, Yuill N. Children's explanations for self-presentational behaviour. European Journal of Social Psychology. 1999; 29(1):105–111. doi: http://dx.doi.org/10.1002/(SICI)1099-0992(199902)29:1<105: :AID-EJSP910>3.0.CO;2-K.

Bartsch, K.; Wellman, HM. Children talk about the mind. New York: Oxford University Press; 1995.

Begeer S, Gevers C, Clifford P, Verhoeve M, Kat K, Hoddenbach E, Boer F. Theory of mind training in children with autism: A randomized controlled trial. Journal of Autism and Developmental Disorders. 2011; 41(8):997–1006. doi: http://dx.doi.org/10.1007/s10803-010-1121-9. [PubMed: 20976617]

Begeer S, Howlin P, Hoddenbach E, Clauser C, Lindauer R, Clifford P, Gevers C, Boer F, Koot HM. Effects and moderators of a short theory of mind intervention for children with autism spectrum disorders: A randomized controlled trial. Autism Research. 201510.1002/aur.1489

Benson, JE.; Sabbagh, MA. Theory of mind and executive functioning: A developmental neuropsychological approach. In: Zelazo, PD.; Chandler, M.; Crone, E., editors. Developmental social cognitive neuroscience. New York: Psychology Press; 2010. p. 63-80.

Benson JE, Sabbagh MA, Carlson SM, Zelazo PD. Individual differences in executive functioning predict preschoolers's improvement from theory-of-mind training. Developmental Psychology. 2013; 49:1615–1627. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1037/a0031056. [PubMed: 23244411]

Bierman KL, Coie JD, Dodge KA, Greenberg MT, Lochman JE, McMahon RJ, Pinderhughes E. The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. Journal of Consulting and Clinical Psychology. 2010; 78(2):156–168. doi: http://dx.doi.org/10.1037/a0018607. [PubMed: 20350027]

Borenstein, M.; Hedges, L.; Higgins, J.; Rothstein, H. Comprehensive meta-analysis, version 2. Englewood, NJ: Biostat Inc; 2005.

Borntrager CF, Chorpita BF, Higa-McMillan CK, Weisz JR. the Research Network on Youth Mental Health. Provider attitudes towards evidence-based practices: Are the concerns with the evidence or with the manuals? Mental Health Services Research. 2009; 60:677–681. doi: http://dx.doi.org/10.1176/appi.ps.60.5.677.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Bosacki S, Astington JW. Theory of mind in preadolescence: Relations between social understanding and social competence. Social Development. 1999; 8(2):237–255. doi: http://dx.doi.org/10.1111/1467-9507.00093.

*. Brockmeyer C. Re-reading the relationship between narrative and theory of mind: The effects of narrative training on the developing understanding of mind. Dissertation Abstract International: Section B: The Sciences and Engineering. 2010; 70(9):5863.

Carbonero Martín MÁ, Sâiz Manzanares MC, Román Sanchez JM. Effect of a metacognitive training program of mentalist skills. Psicothema. 2013; 25:31–37. [PubMed: 23336540]

Carlson SM. Developmentally Sensitive Measures of Executive Function in Preschool Children. Developmental Neuropsychology. 2005; 28(2):595–616. doi: http://dx.doi.org/10.1207/s15326942dn2802_3. [PubMed: 16144429]

Cassidy KW, Fineberg DS, Brown K, Perkins A. Theory of mind may becontagious, but you don't catch it from your twin. Child Development. 2005; 76:97–106. doi: http://dx.doi.org/10.1111/j.1467-8624.2005.00832.x. [PubMed: 15693760]

Cassidy J, Parke RD, Butkovsky L, Braungart JM. Family-peer connections: The roles of emotional expressiveness within the family and children's understanding of emotions. Child Development. 1992; 63(3):603–618. doi: http://dx.doi.org/10.2307/1131349. [PubMed: 1600825]

*. Cerruto A. The effects of training on theory of mind tasks with children who are deaf. Dissertation Abstract International: Section B: The Sciences and Engineering. 1999; 60(5-B):2383.

Chorpita, BF.; Daleiden, EL. 2007 Biennial report: Effective psychosocial interventions for youth with behavioral and emotional needs. Child and Adolescent Mental Health Division, Hawaii Department of Health; Honolulu, HI: 2007.

Chorpita BF, Daleiden EL, Weisz JR. Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. Mental health services research. 2005; 7(1):5–20. doi: http://dx.doi.org/10.1007/s11020-005-1962-6. [PubMed: 15832690]

Corcoran, R. Theory of mind in other clinical conditions: Is a selective 'theory of mind' deficit exclusive to autism?. In: Baron-Cohen, S.; Tager-Flusberg, H.; Cohen, D., editors. Understanding other minds: Perspectives from developmental cognitive neuroscience. 2nd. New York, NY: Oxford University Press; US; 2000. p. 391-421.

Cutting AL, Dunn J. Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. Child Development. 1999; 70:853–865. doi: http://dx.doi.org/10.1111/1467-8624.00061. [PubMed: 10446724]

Deleau M. Language and theory of mind. European Journal of Developmental Psychology. 2012; 9:295–312. doi: http://dx.doi.org/10.1080/17405629.2012.680303.

Devine RT, Hughes C. Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. Child Development. 2013; 84:989–1003. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/cdev.12017. [PubMed: 23199139]

Devine RT, Hughes C. Relations between false belief understanding and executive function in early childhood: A meta-analysis. Child Development. 2014; 85:1777–1794. [PubMed: 24605760]

de Rosnay M, Hughes C. Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? British Journal of Developmental Psychology. 2006; 24:7–37. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1348/026151005X82901.

de Villiers, JG. Can language acquisition give children a point of view?. In: Astington, JW.; Baird, JA., editors. Why language matters for theory of mind. New York: Oxford University Press; 2005. p. 186-219.

Doherty, MJ. Theory of mind: How children understand others' thoughts and feelings. New York, NY: Psychology Press; US; 2009.

Doudin, PA.; Martin, D.; Albanese, O. Metacognition et education: theorie et practique. Berne: Peter Lang; 2001.

Flavell JH. Theory-of-mind development: Retrospect and prospect. Merrill-Palmer Quarterly. 2004; 50:274–290. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1353/mpq.2004.0018.

Flavell JH, Flavell ER, Green FL. Development of the appearance–reality distinction. Cognitive Psychology. 1983; 15:95–120. doi: http://dx.doi.org/10.1016/0010-0285(83)90005-1. [PubMed: 6831859]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gopnik A, Wellmann HM. Why the child's theory of mind really is a theory. Mind & Language. 1992; 7:145–171. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/j.1468-0017.1992.tb00202.x.

Gopnik, A.; Wellmann, HM. The theory theory. In: Hirschfeld, LA.; Gelman, SA., editors. Mapping the mind: Domain specificity in cognition and culture. Cambridge: Cambridge University Press; 1994. p. 257-293.doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1017/CBO9780511752902.011

Grazzani I, Ornaghi V. How do use and comprehension of mental-state language relate to theory of mind in middle childhood? Cognitive Development. 2012; 27:99–111. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1016/j.cogdev.2012.03.002.

*. Guajardo NR, Watson AC. Narrative discourse and theory of mind development. The Journal of Genetic Psychology: Research and Theory on Human Development. 2002; 163(3):305–325. doi: http://dx.doi.org/10.1080/00221320209598686.

*. Hadwin J, Baron-Cohen S, Howlin P, Hill K. Can we teach children with autism to understand emotions, belief, or pretence? Development and Psychopathology. 1996; 8(2):345–365. doi: http://dx.doi.org/10.1017/S0954579400007136.

*. Hale CM, Tager-Flusberg H. The influence of language on theory of mind: A training study. Developmental Science. 2003; 6(3):346–359. doi: http://dx.doi.org/10.1111/1467-7687.00289. [PubMed: 16467908]

Harris, PL. The work of the imagination. Oxford: Blackwell; 2000.

Harris, PL. Conversation, Pretense, and Theory of Mind. In: Astington, JW.; Baird, J., editors. Why language matters for theory of mind. New York, NY: Oxford University Press; US; 2005. p. 70-83.

Harris PL, de Rosnay M, Pons F. Language and Children's Understanding of Mental States. Current Directions in Psychological Science. 2005; 14(2):69–73. doi: http://dx.doi.org/10.1111/j.0963-7214.2005.00337.x.

Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. New York, NY: Academic Press; 1985.

Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychological Methods. 1998; 3(4):486–504. doi: http://dx.doi.org/10.1037/1082-989X.3.4.486.

*. Howard AA. Mental verb input and children's theory of mind: A training study. Dissertation Abstract International: Section B: The Sciences and Engineering. 2009; 69(10):6451.

Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or Isuperscript 2 index? Psychological Methods. 2006; 11(2):193–206. doi:http://dx.doi.org/10.1037/1082-989X.11.2.193. [PubMed: 16784338]

Hughes C, Graham A. Measuring executive functions in childhood: Problems and solutions? Child and Adolescent Mental Health. 2002; 7(3):131–142. doi: http://dx.doi.org/10.1111/1475-3588.00024.

Hughes C, Jaffee S, Happe F, Taylor A, Caspi A, Moffitt T. Origins of individualdifferences in theory of mind: From nature to nurture? Child Development. 2005; 76:356–370. doi: http://dx.doi.org/10.111/j.1467-8624.2005.00850_a.x. [PubMed: 15784087]

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clinical Trials. 1996; 17:1–12. doi: http://dx.doi.org/10.1016/0197-2456(95)00134-4.

Jones SM, Brown JL, Aber J. Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. Child Development. 2011; 82(2):533–554. doi: http://dx.doi.org/10.1111/j.1467-8624.2010.01560.x. [PubMed: 21410922]

Jones SM, Brown JL, Hoglund WL, Aber J. A school-randomized clinical trial of an integrated social-emotional learning and literacy intervention: Impacts after 1 school year. Journal of Consulting and Clinical Psychology. 2010; 78(6):829–842. doi: http://dx.doi.org/10.1037/a0021383. [PubMed: 21114343]

*. Kloo D, Perner J. Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. Child Development. 2003; 74(6):1823–1839. doi: http://dx.doi.org/10.1046/j.1467-8624.2003.00640.x. [PubMed: 14669898]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lecce S, Bianco F, Demicheli P, Cavalini E. Training preschoolers on first-order false belief understanding: Transfer on advanced ToM skills and metamemory. Child Development. 2014; 85:2404–2418.10.1111/cdev.12267 [PubMed: 25040788]

Lecce S, Bianco F, Devine RT, Hughes C, Banerjee R. Promoting theory of mind during middle childhood: A training program. Journal of Experimental Child Psychology. 2014; 126:52–67.10.1016/j.jecp.2014.03.002 [PubMed: 24892882]

Lecce S, Caputi M, Hughes C. Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement? Journal of Experimental Child Psychology. 2011; 110(3):313–331. doi: http://dx.doi.org/10.1016/j.jecp.2011.04.011. [PubMed: 21624614]

Lillard AS, Kavanaugh RD. The contribution of symbolic skills to the development of an explicit theory of mind. Child Development. 2014; 85:1535–1551. [PubMed: 24502297]

Lillard AS, Lerner MD, Hopkins EJ, Dore RA, Smith ED, Palmquist CM. The impact of pretend play on children's development: A review of the evidence. Psychological Bulletin. 2012; 39:1–34. doi: http://dx.doi.org/10.1037/a0029321. [PubMed: 22905949]

Liu D, Wellman HM, Tardif T, Sabbagh MA. Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. Developmental Psychology. 2008; 44:523–531. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1037/0012-1649.44.2.523. [PubMed: 18331141]

Lohmann H, Tomasello M. The Role of Language in the Development of False Belief Understanding: A Training Study. Child Development. 2003; 74(4):1130–1144. doi: http://dx.doi.org/10.1111/1467-8624.00597. [PubMed: 12938709]

*. Lu H, Su Y, Wang Q. Talking about others facilitates theory of mind in Chinese preschoolers. Developmental Psychology. 2008; 44(6):1726–1736. doi: http://dx.doi.org/10.1037/a0013074. [PubMed: 18999334]

McAlister A, Peterson C. A longitudinal study of child siblings and theory of mind development. Cognitive Development. 2007; 22(2):258–270.

McDowell DJ, O'Neil R, Parke RD. Display rule application in a disappointing situation and children's emotional reactivity: Relations with social competence. Merrill-Palmer Quarterly: Journal of Developmental Psychology. 2000; 46(2):306–324.

Meins E, Fernyhough C, Wainwright R, Gupta MD, Fradley E, Tuckey M. Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. Child Development. 2002; 73:1715–1726. doi: http://dx.doi.org/10.111/1467-8624.00501. [PubMed: 12487489]

Milligan K, Astington JW, Dack LA. Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. Child Development. 2007; 78(2):622–646. doi: http://dx.doi.org/10.1111/j.1467-8624.2007.01018.x. [PubMed: 17381794]

*. Nash HM. Social skills training, training to task, and theory of mind performance among young children. Dissertation Abstract International: Section B: The Sciences and Engineering. 2002; 62(10-B):4797.

Onishi KH, Baillargeon R. Do 15-month-old infants understand false beliefs? Science. 2005; 308:255–258. doi: http://dx.doi.org/10.1126/science.1107621. [PubMed: 15821091]

*. Ornaghi V, Brockmeier J, Gavazzi IG. The role of language games in children's understanding of mental states: A training study. Journal of Cognition and Development. 2011; 12(2):239–259. doi: http://dx.doi.org/10.1080/15248372.2011.563487.

*. Ozonoff S, Miller JN. Teaching theory of mind: A new approach to social skills training for individuals with autism. Journal of Autism and Developmental Disorders. 1995; 25(4):415–433. doi: http://dx.doi.org/10.1007/BF02179376. [PubMed: 7592252]

Paynter J, Peterson CC. Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. Research in Autism Spectrum Disorders. 2013; 7:344–348.10.1016/j.rasd.2012.10.001

Perner J, Ruffman T, Leekam SR. Theory of mind is contagious: You catch it from your sibs. Child development. 1994; 65(4):1228–1238.

*. Peskin J, Astington JW. The effects of adding metacognitive language to story texts. Cognitive Development. 2004; 19(2):253–273. doi: http://dx.doi.org/10.1016/j.cogdev.2004.01.003.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pons, F.; Doudin, PA.; Harris, PL.; de Rosnay, M. Métaémotion et intégration scolaire. In: Lafortune, L.; Mongeau, P., editors. L'affectivité dans l'apprentissage. Sainte-Foy: Presses de l'Université du Québec; 2002. p. 7-28.

Pons F, Harris PL. Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. Cognition and Emotion. 2005; 19(8):1158–1174. doi: http://dx.doi.org/10.1080/02699930500282108.

Qu L, Chee YY, Chene L. Teachers's theory-of-mind coaching and children's executive function predict the training effect of sociodramatic play on children's theory of mind. Social Development. 2015; 24:716–733.10.1111/sode.12116

Repacholi, B.; Slaughter, V. Individual differences in theory of mind: Implications for typical and atypical development. New York, NY: Psychology Press; US; 2003.

Rieffe C, Oosterveld P, Miers AC, Terwogt MM, Ly V. Emotion awareness and internalising symptoms in children and adolescents: The Emotion Awareness Questionnaire revised. Personality and Individual Differences. 2008; 45(8):756–761. doi: http://dx.doi.org/10.1016/j.paid.2008.08.001.

Rieffe C, Terwogt MM, Petrides K, Cowan R, Miers AC, Tolland A. Psychometric properties of the Emotion Awareness Questionnaire for children. Personality and Individual Differences. 2007; 43(1):95–105. doi: http://dx.doi.org/10.1016/j.paid.2006.11.015.

Rosenthal, R. Meta-analytic procedures for social research. rev. Thousand Oaks, CA: Sage Publications Inc; 1991.

Rosenthal, R. Meta-analytic procedures for social research. Newbury Park, CA: Sage Publications; 1993.

Rosenthal R, Rubin DB. Comment: Assumptions and procedures in the file drawer problem. Statistical Science. 1988; 3:120–125. doi: http://dx.doi.org/10.1214/ss/1177013014.

Rostan C, Sidera F, Serrano J, Amadó A, Vallés-Majoral E, Esteban M, Serrat E. Fostering theory of mind development. Short and medium-term effects of training false belief understanding. Journal for the Study of Education and Development. 2014; 37:498–529.10.1080/02103702.2014.965464

Ruffman T, Perner J, Naito M, Parkin L, Clements W. Older (but not younger) siblings facilitate false belief understanding. Developmental Psychology. 1998; 34:161–174. doi: http://dx.doi.org/10.1047/0012-1649.34.1.161. [PubMed: 9471013]

Ruffman T, Perner J, Parkin L. How parenting style affects false belief understanding. Social Development. 1999; 8:395–411. doi: http://dx.doi.org/10.1111/1467-9507.00103.

Sabbagh MA, Xu F, Carlson SM, Moses LJ, Lee K. The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. Psychological Science. 2006; 17:74–81. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/j.1467-9280.2005.01667.x. [PubMed: 16371147]

Sellabona ES, Sánchez CR, Vallés-Majoral E, Guitart ME, Caballero FS, Ortiz JS. Labelling improves false belief understanding. A training study. The Spanish Journal of Psychology. 2013; 16 DOI: http://dx.doi.org/10.1017/sjp.2013.8.

*. Slaughter V. Children's understanding of pictorial and mental representations. Child Development. 1998; 69(2):321–332. doi: http://dx.doi.org/10.2307/1132167. [PubMed: 9586208]

*. Slaughter V, Gopnik A. Conceptual coherence in the child's theory of mind: Training children to understand belief. Child Development. 1996; 67(6):2967–2988. doi: http://dx.doi.org/10.2307/1131762. [PubMed: 9071768]

Slaughter V, Dennis MJ, Pritchard M. Theory of mind and peer acceptance in preschool children. British Journal of Developmental Psychology. 2002; 20(4):545–564. doi: http://dx.doi.org/10.1348/026151002760390945.

*. Smith H. The effects of a drama-based language intervention on the development of theory of mind and executive function in urban kindergarten children. Dissertation Abstract International: Section B: The Sciences and Engineering. 2010; 72(5-B):3125.

Sprung M. Unwanted intrusive thoughts and cognitive functioning in kindergarten and young elementary school-age children following Hurricane Katrina. Journal of Clinical Child and Adolescent Psychology. 2008; 37(3):575–587. doi: http://dx.doi.org/10.1080/15374410802148236. [PubMed: 18645748]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sprung M. Clinically relevant measures of children's Theory of Mind and knowledge about thinking: Non-standard and advanced measures. Child and Adolescent Mental Health. 2010; 15(4):204–216. doi: http://dx.doi.org/10.1111/j.1475-3588.2010.00568.x.

Sprung M, Harris PL. Intrusive thoughts and young children's knowledge about thinking following a natural disaster. Journal of Child Psychology and Psychiatry. 2010; 51(10):1115–1124. doi: http://dx.doi.org/10.1111/j.1469-7610.2010.02273.x. [PubMed: 20840501]

*. Sprung, M.; Lohmann, H.; Zauner, P.; Bauer, N. Unpublished Manuscript, Department of Psychology. University of Vienna; Austria: 2011. Theory of Mind and discourse based interventions: Can language structure training really improve theory of mind understanding?.

*. Sprung, M.; Ristic, S.; Fritz, M.; Markova, G. Unpublished Manuscript, Department of Psychology. University of Vienna; Austria: 2012. Theory-of-Mind and discourse-based intervention: What aspects of language facilitate theory-of-mind understanding.

Stanzione C, Schick B. Environmental language factors in theory of mind development: Evidence from children who are deaf/hard-of-hearing or who have specific language impairment. Topics in Language Disorders. 2014; 34:296–312. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1097/TLD.0000000000000038.

*. Steerneman P, Huskens B. The development of a social cognition training for autistic children. Psychologie in Erziehung und Unterricht. 1996; 43(4):291–301. Retrieved from http://www.reinhardt-verlag.de/de/zeitschriften/peu/.

Taylor M, Carlson SM, Maring BL, Gerow L, Charley CM. The characteristics and correlates of fantasy in school-age children: Imaginary companions, impersonation, and social understanding. Developmental Psychology. 2004; 40:1173–1187. doi: http://dx.doi.org/10.1037/0012-1649.40.6.1173. [PubMed: 15535765]

*. Taylor M, Hort B. Can children be trained in making the distinction between appearance and reality? Cognitive Development. 1990; 5(1):89–99. doi: http://dx.doi.org/10.1016/0885-2014%2890%2990014-K.

Wang Y, Su Y. False belief understanding: Children catch it from classmates of different ages. International Journal of Behavioral Development. 2009; 33:331–337. doi: http://dx.doi.org/10.1177/0165025409104525.

Watson AC, Nixon CL, Wilson A, Capage L. Social interaction skills and theory of mind in young children. Developmental Psychology. 1999; 35(2):386–391. doi: http://dx.doi.org/10.1037/0012-1649.35.2.386. [PubMed: 10082009]

Waugh C, Peskin J. Improving the social skills of children with HFASD: An intervention study. Journal of Autism and Developmental Disorders. 2015; 45:2961–2980. [PubMed: 25956127]

Weisz JR, Chorpita BF, Palinkas LA, Schoenwald SK, Miranda J, Bearman SK, Daleiden EL, Ugueto AM, Ho A, Martin J, Gray J, Alleyne A, Langer DA, Southam-Gerow MA, Gibbons RD. the Research Network on Youth Mental Health. Testing standard and modular designs for psychotherapy with youth depression, anxiety, and conduct problems: A randomized effectiveness trial. Archives of General Psychiatry. 2012; 69:274–282. doi: http://dx.doi.org/10.1001/archgenpsychiatry.2011.147. [PubMed: 22065252]

Wellman, HM. Making minds: How theory of mind develops. New York: Oxford University Press; 2014.

Wellman HM, Cross D, Watson J. Meta-analysis of theory-of-mind development: The truth about false belief. Child Development. 2001; 72:655–684. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/1467-8624.00304. [PubMed: 11405571]

Wellman HM, Fuxi F, Peterson CC. Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. Child Development. 2011; 82:780–792. doi: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/j.1467-8624.2011.01583.x. [PubMed: 21428982]

Wellman HM, Liu D. Scaling of theory of mind tasks. Child Development. 2004; 72:655–684. doi: http://dx.doi.org/10.1111/j.1467-8624.2004.00691.x. [PubMed: 11405571]

Wellman HM, Peterson CC. Deafness, thought bubbles, and theory-of-mind development. Developmental Psychology. 2013; 49:2357–2367. doi: http://dx.doi.org/10.1037/a0032419. [PubMed: 23544856]

Author Manuscript

Wimmer H, Perner J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition. 1983; 13(1):103–128. doi: http://dx.doi.org/10.1016/0010-0277(83)90004-5. [PubMed: 6681741]

Xiao X, Yang N, Qian L, Zhou S. Pretend playing training improves theory of mind in children with autism. Psychology. 2014; 22:742–745.

Youngblade LM, Dunn J. Individual differences in young children's pretend play with mother and sibling: Links to relations and understanding of other people's feelings and beliefs. Child Development. 1995; 66:1472–1492. doi: http://dx.doi.org/10.2307/1131658. [PubMed: 7555225]

*. Zhang LJ, Wu N, Zheng Y. The training effects of different linguistic interference on young Chinese performance on a theory of mind task. Acta Psychologica Sinica. 2008; 40(7):819–827. doi: http://dx.doi.org/10.3724/SP.J.1041.2008.00819.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

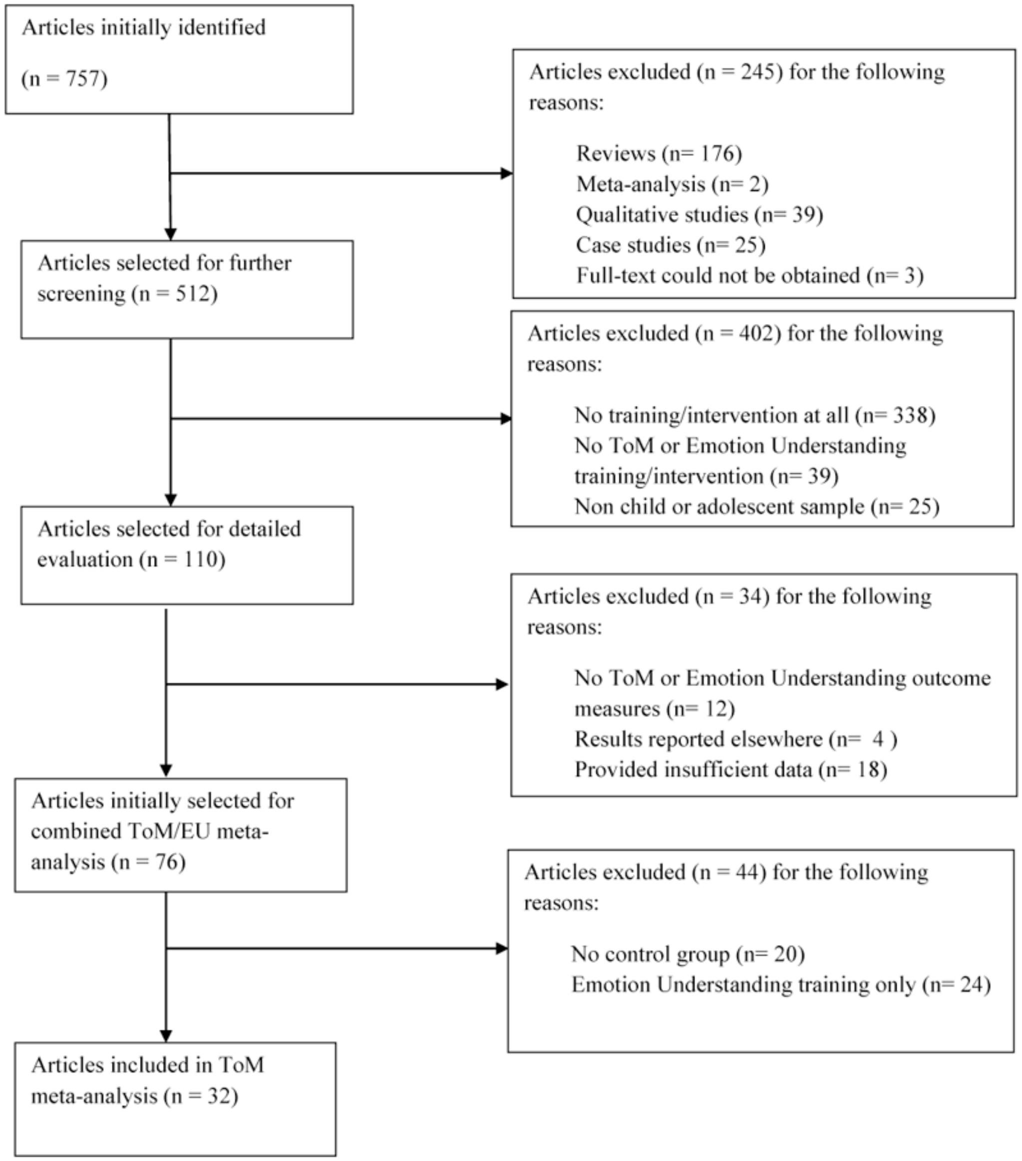

Figure 1. Flow diagram of the study selection process

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

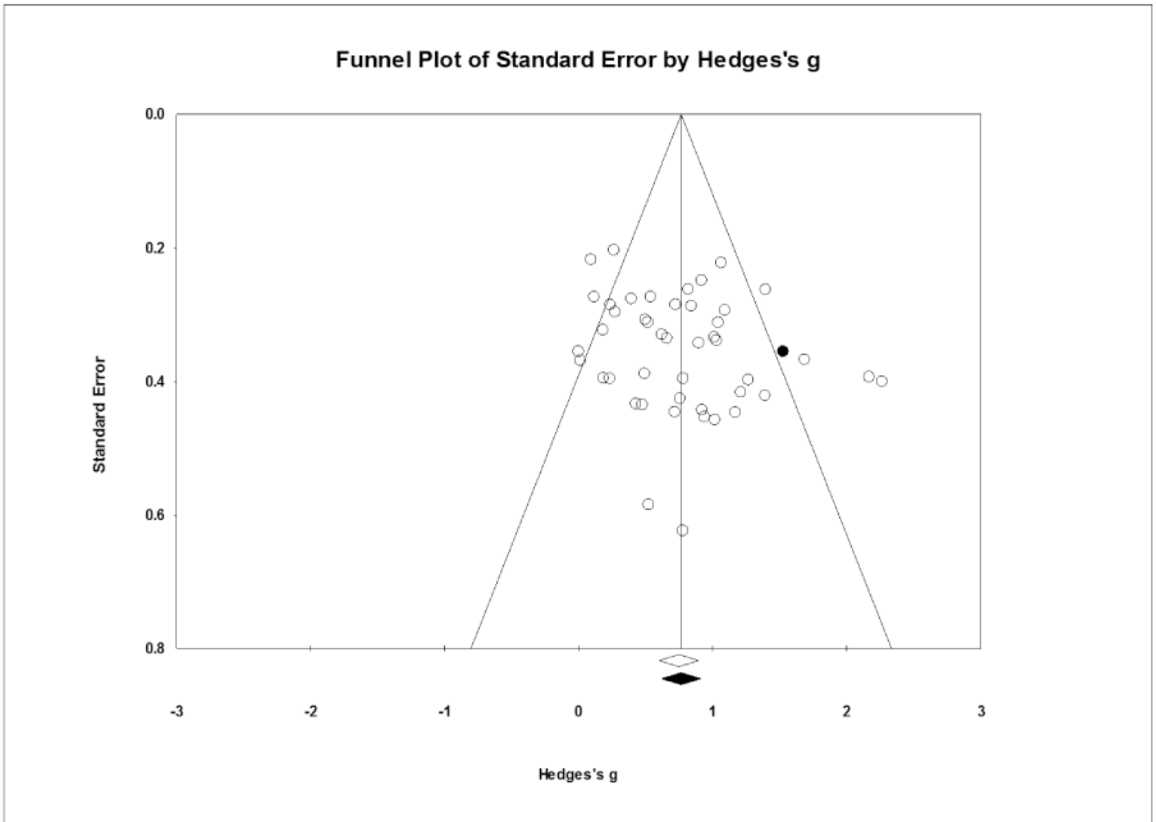

Figure 2. Funnel plot of precision by Hedges's *g*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Description of studies

Study	Primary domain targeted in training	No. of training sessions	Theory-of-Mind training (n)	Comparison condition (n)	Total sample size	Total sample mean age (months)	Sample type	False Belief	Appearance-Reality	Combined ToM
Begeer et al., 2011	Nonspecific	16	Social cognition + perspective-taking (19)	Wait-list (17)	36	123	ASD	None	None	Muris battery
Begeer et al., 2015	Nonspecific	8	Social cognition + perspective-taking (45)	Wait-list (52)	97	115.2	ASD	None	None	Muris battery
*Brockmeyer, 2010 Comparison a	Nonspecific	12	Storytelling enriched with mental states + Story-acting (12)	Wait-list (14)	26	56	Typical developing	Deception comprehension task; Active deception task	Object distinction task	Wellman battery
*Brockmeyer, 2010 Comparison b	Nonspecific	12	Storytelling without mental states + Story-acting (15)	Wait-list (14)	19	56	Typical developing	Deception comprehension task; Active deception task	Object distinction task	Wellman battery
Carbonero et al., 2013	Nonspecific	12	Social cognition + perspective-taking (10)	Wait-list (10)	20	62.5	Typical developing	Unexpected contents task; Unexpected transfer task	None	None
*Cerruto, 1999	FB	2	False belief tasks + corrective feedback (19)	Number conservation task (19)	38	88.44	Hearing Impaired	Unexpected contents task	None	None
Guajardo & Watson, 2002 Study 1	Nonspecific	12-15	Narrative Training highlighting mental states (18)	Wait-list (19)	37	46	Typical developing	Unexpected contents task; Unexpected transfer task; Deception task	Object distinction task	None
Guajardo & Watson, 2002 Study 2	Nonspecific	13-15	Narrative Training highlighting mental states (28)	Wait-list (26)	54	41	Typical developing	Unexpected contents task; Unexpected transfer task; Deception task	None	None
Hadwin et al, 1996	FB	8	Discussion + role-play involving false belief + corrective feedback (10)	Free play + modeling + verbal guidance (10)	20	108.28	ASD	Simple perspective taking; Complex perspective taking; Seeing/Knowing; True belief/Action prediction; False Belief	None	None
Hale & Tager-Flusberg, 2003 Comparison a	FB	2	False belief tasks + corrective feedback (20)	Relative clause training (20)	40	47	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None
Hale & Tager-Flusberg, 2003 Comparison b	Nonspecific	2	Sentential complements (20)	Relative clause training (20)	40	47	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None
* Howard, 2008	AR	4	Videos enriched with mental state terms + discussion about misleading appearance (12)	Wait-list (12)	24	42	Typical developing	None	None	Wellman battery
Kloo & Perner, 2003	FB	2	False belief tasks + corrective feedback (15)	Number conservation or relative clause training (15)	30	45.1	Typical developing	Unexpected transfer task	None	None
Lecce et al., 2014	Nonspecific		Narrative Training highlighting mental states (33)	Narrative training highlighting	62	52	Typical developing	Deception Comprehension task	None	None

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Study	Primary domain targeted in training	No. of training sessions	Theory-of-Mind training (n)	Comparison condition (n)	Total sample size	Total sample mean age (months)	Sample type	False Belief	Appearance-Reality	Combined ToM
				physical features in the story (29) physical features in the story (29)						
Lecce et al., 2012	Nonspecific		Narrative Training highlighting mental states (45)	Narrative training highlighting physical features in the story (46)	91	112	Typical developing	Unexpected contents task; Unexpected transfer task; Deception task	None	None
Lu et al., 2008	Nonspecific	4	Narrative training highlighting characters in the story (25)	Narrative training highlighting physical features in the story (26)	31	43.9	Typical developing	Unexpected contents task; Unexpected transfer task; Deception task	None	None
*Nash, 2002	FB	9	Videos enriched with mental state terms + discussion about false belief (11)	Number conservation task (10)	21	47.3	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None
Ornahgi et al, 2011	Nonspecific	16	Storytelling enriched with mental states (18)	Free play (18)	36	46.3	Typical developing	Unexpected contents task; Unexpected transfer task; False Belief explanation	None	None
Ozonoff & Miller, 1995	Nonspecific	14	Perspective-taking training through role-play scenarios highlighting mental states (4)	Wait-list (5)	9	164.27	ASD	Unexpected contents task, 2nd order false belief; Advanced false belief	None	None
Paynter & Peterson, 2013	Nonspecific	1-3	Thought bubble training + false belief tasks +corrective feedback (17)	Wait-list (7)	24	84	ASD	Unexpected contents task	None	Wellman battery
Peskin & Astington, 2004	Nonspecific	20	Storytelling enriched with mental state verbs (24)	Storytelling without mental state verbs (24)	48	54	Typical developing	Unexpected contents task; Unexpected transfer task; False Belief explanation	None	None
Qu et al., 2015 Comparison (a)	Nonspecific	4	Sociodramatic play (25)	Free play (20)	45	60	Typical Developing	Unexpected contents task; Unexpected transfer task; False Belief explanation	None	None
Qu et al., 2015 Comparison (b)	Nonspecific	4	Sociodramatic play + mental state terms +adult intervention (26)	Free play (20)	46	60	Typical developing	Unexpected contents task; Unexpected transfer task; False Belief explanation	None	None
Rostan et al., 2014 Comparison (a)	AR	3	Appearance-Reality task + corrective feedback (26)	Non deceptive objects + labeling (26)	52	44.4	Typical developing	Unexpected contents task; deception task	Object distinction task	None
Rostan et al., 2014 Comparison (b)	AR	3	Appearance-Reality task + labeling (26)	Non deceptive objects + labeling (26)	52	44.4	Typical developing	Unexpected contents task; deception task	Object distinction task	None
Sellabona et al., 2013 Comparison (a)	AR	3	Appearance-Reality task + corrective feedback (26)	Non deceptive objects + labeling (26)	52	44.4	Typical developing	Unexpected contents task; deception task	Object distinction task	None
Sellabona et al., 2013 Comparison (b)	AR	3	Appearance-Reality task + labeling (26)	Non deceptive objects + labeling (26)	52	44.4	Typical developing	Unexpected contents task; deception task	Object distinction task	None

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Study	Primary domain targeted in training	No. of training sessions	Theory-of-Mind training (n)	Comparison condition (n)	Total sample size	Total sample mean age (months)	Sample type	False Belief	Appearance-Reality	Combined ToM
Slaughter, 1998	FB	2	False belief tasks + corrective feedback (10)	Number conservation task (10)	20	45	Typical developing	Unexpected contents task; Unexpected transfer task; Deception task	None	None
Slaughter & Gopnik, 1996 Study 1	FB	2	False belief tasks + corrective feedback (11)	Number conservation task (11)	22	44	Typical developing	Unexpected contents task	None	None
Slaughter & Gopnik, 1996 Study 1	Nonspecific	2	Perception & desire tasks + corrective feedback (11)	Number conservation task (11)	22	44	Typical developing	Unexpected contents task	None	None
Slaughter & Gopnik, 1996 Study 2 Comparison a	FB	2	False belief tasks + corrective feedback (13)	Number conservation task (13)	26	45	Typical developing	Unexpected contents task	Object distinction task	None
Slaughter & Gopnik, 1996 Study 2 Comparison b	Nonspecific	2	Perception & desire tasks + corrective feedback (13)	Number conservation task (13)	26	45	Typical developing	Unexpected contents task	Object distinction task	None
*Smith, 2010	Nonspecific	13	Drama-based language intervention (41)	Standard language instruction (42)	83	72.8	Typical developing	None	None	NEPSY-II Theory-of-Mind subtest
+Sprung et al, 2011 Study 1	Nonspecific	3	Sentential complements (21)	General language: WH-questions (20)	41	43	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None
+Sprung et al, 2011 Study 2	Nonspecific	3	Sentential complements (21)	General language: yes/no questions (21)	42	46	Typical developing	Unexpected contents task; Unexpected transfer task; Deception task	None	None
+Sprung et al, 2012 Comparison a	Nonspecific	5	Sentential complements (35)	Neutral storybook reading (35)	70	61	Typical developing	None	None	Wellman battery
+Sprung et al, 2012 Comparison b	Nonspecific	5	Perspective-taking (36)	Neutral storybook reading (35)	70	61	Typical developing	None	None	Wellman battery
Steerneman & Huskens, 1996 Study 1	Nonspecific	27	Social cognition + perspective-taking (5)	Wait-list (5)	10	108	ASD & PDD-NOS	Unexpected contents task; Unexpected transfer task	None	None
Taylor & Hort, 1990 Study 2	AR	1	Appearance-Reality task + corrective feedback (12)	Task without appearance-reality content (12)	24	40.8	Typical developing	None	Color distinction task; Object distinction task	None
Waugh & Peskin, 2015	Nonspecific	10	Narrative Training highlighting mental states (19)	Children's friendship training (11)	49	108	ASD	Unexpected contents task; Unexpected transfer task	None	None
Wellman & Peterson, 2013	Nonspecific	6	Thought bubble training + false belief tasks + corrective feedback (13)	Art project program (14)	43	108	Hearing Impaired	Unexpected contents task; Unexpected transfer task	None	Wellman battery

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Study	Primary domain targeted in training	No. of training sessions	Theory-of-Mind training (n)	Comparison condition (n)	Total sample size	Total sample mean age (months)	Sample type	False Belief	Appearance-Reality	Combined ToM
Xiao et al., 2014	Nonspecific	32	Sociodramatic Play (10)	Wait-list (10)	20	58.9	ASD	Unexpected contents task; Unexpected transfer task	None	None
Zhang et al, 2008 Comparison a	FB	14	Sentential complements + deceptive experience + mental state verbs (20)	Wait-list (19)	39	53.9	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None
Zhang et al, 2008 Comparison b	Nonspecific	14	Sentential complements (20)	Wait-list (19)	39	53.9	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None
Zhang et al, 2008 Comparison c	FB	14	Deceptive experience (18)	Wait-list (19)	37	53.9	Typical developing	Unexpected contents task; Unexpected transfer task	Object distinction task	None

Note:

* Dissertation;

+ Unpublished;

FB = False Belief; AR = Appearance Reality

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Overall effect sizes

Study	Hedges's *g*	95% CI	*p*
Begeer et al., 2011	0.660	[.002, 1.318]	0.05
Begeer et al, 2015	0.264	[-.134, .662]	0.19
*Brockmeyer, 2010 (a)	1.211	[.395, 2.027]	0.00
*Brockmeyer, 2010 (b)	1.267	[.487, 2.046]	0.00
Carbonero et al 2013	1.017	[.120, 1.93]	0.03
*Cerruto, 1999	1.029	[.365, 1.694]	0.00
Guajardo & Watson, 2002 (Study 1)	0.183	[-.450, .816]	0.57
Guajardo & Watson, 2002 (Study 2)	0.537	[.002, 1.073]	0.05
Hadwin et al, 1996	0.429	[-.420, 1.279]	0.32
Hale & Tager-Flusberg, 2003 (a)	2.265	[1.480, 3.049]	0.00
Hale & Tager-Flusberg, 2003 (b)	2.168	[1.396, 2.939]	0.00
*Howard, 2008	0.186	[-.589, .960]	0.64
Kloo & Perner, 2003	0.000	[-.696, .696]	1.00
Lecce et al 2012	1.064	[.628, 1.500]	0.00
Lecce et al 2014	0.820	[.307, 1.333]	0.00
Lu et al., 2008	0.724	[.165, 1.282]	0.01
*Nash, 2002	0.922	[.054, 1.790]	0.04
Ornahgi et al, in press	0.897	[.225, 1.568]	0.01
Ozonoff & Miller, 1995	0.778	[-.445, 2.001]	0.21
Paynter & Peterson 2013	0.718	[-.156, 1.592]	0.11
Peskin & Astington, 2004	0.236	[-.323, .796]	0.41
Qu et al, 2015 (a)	0.273	[-.308, .853]	0.36
Qu et al, 2015 (b)	1.042	[.430, 1.653]	0.00
Rostan et al 2014 (a)	0.117	[-.419, .652]	0.67
Rostan et al 2014 (b)	1.091	[.516, 1.666]	0.00
Sellabona et al 2013 (a)	0.394	[-.147, .936]	0.15
Sellabona et al 2013 (b)	0.843	[.280, 1.405]	0.00
Slaughter & Gopnik, 1996, Study 1 (a)	1.170	[.294, 2.045]	0.01
Slaughter & Gopnik, 1996, Study 1 (b)	0.757	[-.078, 1.591]	0.08
Slaughter & Gopnik, 1996, Study 2 (a)	0.493	[-.268, 1.255]	0.20
Slaughter & Gopnik, 1996, Study 2 (b)	0.781	[.005, 1.556]	0.05
Slaughter, 1998	0.940	[.051, 1.828]	0.04
*Smith, 2010	0.094	[-.333, .520]	0.67
+Sprung et al, 2011, Study 1	0.518	[-.093, 1.129]	0.10
+Sprung et al, 2011, Study 2	0.498	[-.105, 1.101]	0.11
+Sprung et al, 2011, Study 3 (a)	1.394	[.880, 1.909]	0.00
+Sprung et al, 2011, Study 3 (b)	0.917	[.429, 1.405]	0.00
Steerneman & Huskens, 1996, Study 1	0.524	[-.622, 1.670]	0.37
Taylor & Hort, 1990, Study 2	0.233	[-.542, 1.009]	0.56

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Study	**Hedges's *g***	**95% CI**	***p***
Waugh & Peskin, 2015	0.014	[-.708, .737]	0.97
Wellman & Peterson, 2013	1.391	[.565, 2.217]	0.00
Xiao et al, 2014	0.476	[-.377, 1.329]	0.27
Zhang et al, 2008 (a)	0.621	[-.025, 1.268]	0.06
Zhang et al, 2008 (b)	1.687	[.967, 2.407]	0.00
Zhang et al, 2008 (c)	1.012	[.357, 1.667]	0.00
Overall Effect Size	0.747	[.599, .894]	0.00

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Moderator analyses

Variable	***B* value**	***SE***	***p***	***R^2 analog***
Number of sessions	-0.02	0.01	.07	0.01
Length of each session	-0.01	0.00	.02*	0.03
Length of training period	0.01	0.00	.03*	0.03
Post-test delay	0.00	0.00	.53	0.00
Total sample mean age	0.00	0.00	.21	0.00
Gender	-0.91	0.60	.06	0.00
Jadad 1	0.33	0.22	.13*	0.00
Jadad 2	-0.16	0.19	.39	0.00
Jadad 3	-0.01	0.10	.95	0.00
Jadad 4	0.09	0.07	.23	0.00
Publication Status	-.05	.19	.77	0.00
Year	0.01	0.00	.29	0.00

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

PRISMA Checklist of items to include when reporting a systematic review or meta-analysis

Section/topic	Item No.	Checklist item	Reported on page No.
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both	1
Abstract			
Structured Summary	2	Provide a structured summary including, as applicable, background, objectives, data sources, study eligibility criteria, participants, interventions, study appraisal and synthesis methods, results, limitations, conclusions and implications of key findings, systematic review registration number	3
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known	3-8
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	8
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (such as web address), and, if available, provide registration information including registration number	8
Eligibility Criteria	6	Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale	8-9
Information sources	7	Describe all information sources (such as databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	9
Study selection	9	State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	9
Data collection process	10	Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	10-11
Data items	11	List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made	12
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias on individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	12
Summary measures	12	State the principal summary measure (such as risk ratio, difference in means)	13
Synthesis of results	12	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I^2 statistic) for each meta-analysis	14
Risk of bias across studies	12	Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)	14

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Section/topic	Item No.	Checklist item	Reported on page No.
Additional analyses	13	Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	14
Results			
Study selection	13	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusion at each stage, ideally with a flow diagram	13-15
Study characteristics	11	For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide the citations	13-15
Risk of bias within studies	19	Present data on risk of bias of each study and, if applicable, any outcome-level assessment (see item 12)	14
Results of individual studies	20	For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot	Table 1
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency	13, Table 1
Risk of bias across studies	12	Present results of any assessment of risk of bias across studies (see item 15)	Figure 2
Additional analyses	23	Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression) (see item 16)	11-12
Discussion			
Summary of evidence	24	Summarize the main findings including strength or evidence for each main outcome; consider their relevance to key groups (such as health care providers, users, and policy makers)	15-21
Limitations	25	Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)	19-21
Conclusions	26	Provide general interpretation of the results in the context of other evidence, and implications for future research	21
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (such as supply or data) and role of funders for the systematic review	22