Intelligence 44 (2014) 78–92

Contents lists available at ScienceDirect

Intelligence

Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes"

Crystal A. Baker *, Eric Peterson, Steven Pulos, Rena A. Kirkland

University of Northern Colorado, School of Psychological Sciences, 501 20th Street, Greeley, CO 80639, United States

ARTICLE INFO

Article history:
Received 19 July 2013
Received in revised form 11 February 2014
Accepted 5 March 2014
Available online 11 April 2014

Keywords:
Mental state
Mentalizing
RMET
Social cognition
Theory of mind

ABSTRACT

Although the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al. 1997, 2001) has been used as a measure of mental state understanding in over 250 studies, the extent to which it correlates with intelligence is seldom considered. We conducted a meta-analysis to investigate whether or not a relationship exists between intelligence and performance on the RMET. The analysis of 77 effects sizes with 3583 participants revealed a small positive correlation ($r = .24$) with no difference between verbal and performance abilities. We conclude that intelligence does play a significant role in performance on the RMET and that verbal and performance abilities contribute to this relationship equally. We discuss these findings in the context of the theory of mind and domain-general resources literature.

1. Introduction

As socially intelligent beings, we make inferences regarding others' mental states during social interactions. Recently, many researchers interested in the study of mental state understanding have explored mechanisms across development that may contribute to individual differences among both healthy adults (e.g., Billington, Baron-Cohen, & Wheelwright, 2007) and relatively high-functioning individuals with disorders. The study of individual differences among adults, with and without disorders, has given rise to a demand for instruments with much greater sensitivity than traditional developmental tasks. One such instrument, the Reading the Mind in the Eyes Test (RMET; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), has been used in more than 250 studies (Kirkland, Peterson, Baker, Miller, & Pulos, 2013).

In the RMET, participants view photographs of eyes disembedded from the face and are asked to make a forced-choice as quickly as possible among four mental state descriptors. An important strength of the RMET is its sensitivity to individual differences in adults. For example, the RMET has been used to discriminate among parents of children on the autism spectrum (i.e., the broad autism phenotype) and IQ-matched typically developing parents (Baron-Cohen & Hammer, 1997; Hurley, Losh, Parlier, Reznick, & Piven, 2007). Recently, RMET performance has been associated with aspects of academic profile and professional identity. Students in humanities outperformed those in the physical sciences (Billington et al., 2007), and those choosing certain career paths (e.g., visual arts and law) scored above average on the RMET, whereas those in computer and information technology careers did not (Strong, Russell, Germine, & Wilmer, 2011). While a number of instruments (e.g., neuropsychological test batteries that screen for disorders) have adequate sensitivity for discriminating between disordered and typically-developing samples, the RMET has been especially successful as an individual differences instrument for use with normally developing adults. To date, the RMET has been used in over 250 studies in a wide variety of disciplines (e.g., business,

* Corresponding author. Tel.: +1 507 358 3110.
E-mail addresses: crystal.baker@unco.edu (C.A. Baker), eric.peterson@unco.edu (E. Peterson), steven.pulos@unco.edu (S. Pulos), rena.kirkland@gmail.com (R.A. Kirkland).

http://dx.doi.org/10.1016/j.intell.2014.03.001
0160-2896/

economics), translated into several languages, adapted for use with children (e.g., Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001; Hallerback, Lugnegard, Hjarthag, & Gillberg, 2009) and adapted for brain imaging (Adams et al., 2010). Despite its widespread use, the underlying cognitive processes mediating performance on the RMET have been minimally examined. We conducted a meta-analysis to explore the contribution of intelligence to performance on the RMET.

The goal of understanding the contribution of general intelligence on RMET performance is important for two reasons. From a methodological perspective, it is important to appreciate the degree to which performance differences reflect a specific social cognitive process rather than general intelligence. To date, the RMET has been used with many different clinical groups including, for example, autism and schizophrenia (Hallerback et al., 2009). Of course, such groups may differ from comparison participants on factors associated with general intelligence. If RMET performance relates substantively to general intelligence, studies should account for possible group differences when drawing conclusions about social cognitive impairment; however, not all previous studies have adequately controlled for this possibility (e.g., Bora, Sehitoglu, Aslier, Atabay, & Veznedaroglu, 2007; Demurie, De Corel, & Royers, 2011). A similar criticism is relevant in studies of normally developing individuals. For example, Bailey and Henry (2008) used the RMET to investigate social understanding among older and younger individuals without disorders and did not account for potential differences in general intelligence. A recent meta-analysis (Murphy & Hall, 2011) highlights the possibility that individual differences in psychosocial functioning may reflect general intelligence. From a more theoretical perspective, an exploration of the task demands of the RMET, specifically with respect to intelligence, may inform the degree to which the task taps a relatively implicit, automatic social–cognitive process. In introducing the current revised version of the RMET, Baron-Cohen, Wheelwright, Hill et al. (2001) did not obtain a significant correlation between intelligence and the RMET, "suggesting this is independent of general (nonsocial) intelligence" (p. 247), consistent with the view that performance reflects relatively "unconscious, rapid and automatic" (p. 241) judgments. Consistent with Baron-Cohen's original description, it makes intuitive sense that the RMET would be relatively free of general intelligence demands as compared to other tasks (e.g., Strange Stories) which, by design, involve explicit verbal reasoning. Evidence for such a differential loading of intelligence on the RMET relative to other instruments such as the Strange Stories Task (Happe, 1994) would be consistent with recent speculation about a dual route model of mental state understanding (e.g., Apperly, Samson, Chiavarino, Bickerton, & Humphreys, 2007; Sabbagh, 2004). This model proposes that mental state understanding involves both an implicit, automatic, and inflexible process that operates independently of an explicit cognitively demanding flexible process. The proposal that RMET performance is independent of general intelligence would suggest a process that is relatively more implicit and does not require explicit, cognitive demands.

A second issue of current theoretical interest concerns the degree to which there exists an independent, somewhat-insulated mechanism that mediates social understanding (e.g., Adolphs, 2006; Leslie, 1987). Clearly, the resolution of such a complex and theoretical question will require convergent support across a range of methodologies. This is particularly true because a somewhat modular process may "co-opt" more general mechanisms in the service of social understanding (Siegal & Varley, 2002). Likewise, an implicit social cognitive process may operate in conjunction with more explicit processes. Thus, our meta-analysis cannot provide a clear refutation of models of social cognition that posit either a somewhat modular mechanism or dissociable routes. However, meta-analytic evidence that RMET performance is relatively free of demands on general intelligence might be interpreted as support for a relatively implicit and somewhat modular mechanism. In our review that follows, we first consider the study of mental state understanding that gave rise to the RMET, and then we briefly review the subtopic of intelligence. Finally, we consider the intersection of the RMET and intelligence to frame our hypothesis.

1.1. The study of mental state understanding

Since the inception of the study of theory of mind (ToM; Premack & Woodruff, 1978), the "false belief task" has served as the gold standard for demonstrating ToM, a cognitive milestone of the preschool years (Wellman, Cross, & Watson, 2001). Specifically, the first-order false belief task, which requires understanding the belief another person holds was extended to the second-order false belief task where the participant demonstrates understanding of a belief someone holds about the belief of another person (e.g., Suzy believes that Ray believes that the cup of coffee is hot). However, as evidenced by adult participants on the autism spectrum who manifest pervasive social difficulties while passing false belief tasks, the study of individual differences among older children and adults, with or without disorders, requires instruments with greater sensitivity. ToM tasks designed to measure individual differences in adulthood have differed with respect to task characteristics. While some tasks clearly involve explicit, language-based reasoning, others may place a greater emphasis on perceptual processing (e.g., reading facial expression). For example, the Hinting Task (Corcoran, Mercer, & Frith, 1995) consists of scenarios in which one person hints to another person (e.g., "I want to wear that blue shirt but it's very creased"). The participant must explain the meaning underlying what the character in each story says or does. Similarly, in the Faux Pas Task (Stone, Baron-Cohen, & Knight, 1998) participants identify the social faux pas that occurred in the scenario. Alternatively, other tasks have been designed to capture the relatively more implicit (i.e., less linguistically mediated) process of quickly judging mental states based on brief exposure to perceptual information. Such tasks (e.g., Reading the Mind in the Voice Task, Profile of Nonverbal Sensitivity, RMET) involve decoding nonverbal behavior conveyed in facial expression, body movement, or voice.

In comparison to instruments like the Strange Stories Task or the Faux Pas Test, it makes sense to hypothesize that the RMET would rely relatively less on general intelligence given that it would seem to involve a more implicit social–perceptual analysis. Although one must select a verbal descriptor, we assume this is less linguistically demanding than inferring a mental state based on an analysis of sentence meaning. However, to date we are not aware of any evidence to support or refute this hypothesis.

1.2. General intelligence

Intelligence is generally thought to be a reasoning capacity that allows one to learn from experience and adapt to changing environments (Henry, Sternberg, & Grigorenko, 2005). Frequently it is operationalized through the assessment of intelligence (*g*), and practically it refers to the individual variation we find in mental competence (Hunt, 2005). Tests such as the Wechsler Adult Intelligence Scale include measures of verbal ability and performance ability to capture both crystallized knowledge and fluid reasoning skills, respectively. Intelligence correlates with many real-life outcomes throughout the lifespan (Wilhelm & Engle, 2005), such as income, educational attainment, occupational status, and personality characteristics (see Brand, 1987; Kline, 1991). In many research investigations, intelligence accounts for a significant portion of the variance in the dependent variable (Brand, 1987; Jensen, 1998; Strenze, 2007; Woodward & Fergusson, 2000), but there is controversy as to the extent that intelligence significantly contributes to all mental capacities (Henry et al., 2005; Jensen, 1998).

While tests of intelligence typically focus on more cognitive aspects of intelligence, a number of researchers have speculated that there may also be individual differences in social–emotional intelligence (Matthews, Zeidner, & Roberts, 2005). The concept of “emotional intelligence” has led to a proliferation of research on one's ability to perceive, express, and reason with emotion. A meta-analysis found a correlation of .22 between emotional intelligence and general mental ability (Van Rooy & Viswesvaran, 2004), and correlations have been reported between emotional intelligence and a number of real-life outcomes (for a review, see Mayer, Roberts, & Barsade, 2008). As a result, there is a growing appreciation of the importance of social ability relating to a variety of outcomes (e.g., McGlade et al., 2008).

In a meta-analysis with adults and adolescents, Murphy and Hall (2011) reported a correlation of .19 between intelligence and interpersonal sensitivity (i.e., the ability to decode the state or trait of unfamiliar others); however, this analysis included only 36 studies published between 1931 and 2006 using a wide range of “interpersonal sensitivity” tasks requiring participants to judge emotions, intentions, or thoughts being portrayed by actors through visual and auditory modalities (e.g., Profile of Nonverbal Sensitivity, Diagnostic Analysis of Nonverbal Accuracy, Interpersonal Perception Task). Studies using the RMET were not included in this analysis. While tasks such as the Profile of Nonverbal Sensitivity and Diagnostic Analysis of Nonverbal Accuracy are used in some current research (e.g., Ingersoll, 2010; Wynn, Sugar, Horan, Kern, & Green, 2010), the RMET has been used in hundreds of studies with both neurotypical and clinical samples and continues to be used in a wide variety of studies. Given the widespread acceptance of the RMET as a broad index of mental state understanding, an investigation of the degree to which the instrument involves intelligence is warranted.

1.3. Theory of mind, the RMET, and general intelligence

In past research, traditional ToM tasks have yielded correlations with intelligence varying from weak negative correlations to moderate positive correlations. For example, in schizophrenia research, first-order and second-order false belief tasks correlated at .30 and .24, respectively (Bora et al., 2007) and performance on the Hinting Task correlated with verbal ability at .54 (Bora, Eryavuz, Kayahan, Sungu, & Veznedaroglu, 2006). The Strange Stories Task (SST) correlated at .28 in a group with schizophrenia and .45 in relatives of individuals with schizophrenia; similarly, the Faux Pas Task correlated at .20 in a schizophrenia group and .65 in relatives (de Achaval et al., 2010). Among individuals with autism spectrum disorders in the normal range for IQ, SST performance correlated with intelligence at .26 (Adler, Nadler, Eviatar, & Shamay-Tsoory, 2010) and .19 (Dziobek et al., 2006). In typically developing samples, SST performance correlated at .18 (de Achaval et al., 2010) and .22 (Dziobek et al., 2006), while Faux Pas Task performance correlated at .24 (de Achaval et al., 2010), and reality known and unknown ToM tasks correlated at .40 and .34 (Bailey & Henry, 2008). Further muddying the literature regarding the relationship between intelligence and ToM task performance, some correlations have been reported with combined groups consisting of clinical and nonclinical groups, which introduces increased heterogeneity. For example, Hinting Task performance correlated at .54 in a combined normal/autism spectrum disorder/delusional sample (Craig, Hatton, Craig, & Bentall, 2004) and SST at .50 in a mixed control/autism spectrum disorder sample (David et al., 2008).

Clearly, a meta-analysis is needed in order to more precisely estimate the relationship between intelligence and ToM. A review of the literature also indicates there is a lack of task-specific meta-analyses. Given the diversity of tasks included under the ToM umbrella, it is important we examine the relationship of intelligence with each type of task independently. The insights gained from this task-specific approach may offer theoretical insight into the diverse processes that contribute to overall ToM ability. Further, this approach may have practical benefit for interpreting studies involving use of the RMET with atypical samples that are likely to differ on aspects of both social and nonsocial abilities.

Given that the relationship of intelligence with ToM tasks is presumed to involve more explicit processes (e.g., SST, Faux Pas Task), it makes sense that the RMET would also correlate to some extent with intelligence (though perhaps to a lesser extent than traditional ToM tasks). In a meta-analysis conducted in our lab, the RMET correlated with Strange Stories and Faux Pas Tasks ($r = .29$; Kirkland, Baker, Johnson, Peterson, & Pulos, 2012), indicating the tasks may be measuring some underlying common ability or share task demands or a combination of both. Some authors have assumed there is no correlation between intelligence and RMET performance. For example, referring to the RMET, Mar, Oatley, Hirsh, dela Paz, and Peterson (2006) wrote “scores on this test do not correlate with IQ” (p. 701). However, a number of studies have, in fact, found a relationship between the two. While it is not typically the primary investigation of authors (many do not even report the correlation in their publications), we have included 77 correlations between RMET performance and intelligence in this meta-analysis to estimate the extent of this relationship.

If RMET performance does correlate with intelligence, would verbal or performance ability contribute more to this relationship? While many studies have shown a relationship between theory of mind ability and language ability in children, this is often attributed to the linguistic task demands (e.g., Lewis & Osborne, 1990; Milligan, Astington, & Dack, 2007). However, Barrett, Lindquist, and Gendron (2007) presented evidence to support the hypothesis that language supports emotion perception in a context-rich, top-down process. Beck, Kumschick, Eid, and Klann-Delius (2012) demonstrated a moderate relationship between language ability and emotional competence in children through confirmatory factor analysis. In a study where children were asked to choose which drawing of face emotions best fits a situation evoking emotion (e.g., receiving a birthday gift), language ability explained 27% of the variance in emotion understanding in children after controlling for age (Pons, Lawson, Harris, & de Rosnay, 2003). In a recent study involving an adult sample, a significant correlation was found between RMET performance and verbal ability ($r = .49$) but not performance ability ($r = .18$, ns; Peterson & Miller, 2012). Thus, it makes sense to predict that verbal ability may play a larger role than performance ability.

1.4. Present study

We conducted a meta-analysis of the relationship between RMET performance and intelligence. Upon examining the literature, it is clear that the correlations reported between RMET performance and intelligence span a large range. Because of this variability among the correlations, a meta-analysis was necessary to precisely estimate this relationship in the general population and examine variables contributing to heterogeneity across studies. Our investigation focused on (a) the extent to which intelligence correlates with RMET performance; (b) the contribution of verbal ability relative to performance ability; (c) variables moderating the relationship between intelligence and RMET performance. We hypothesized that the RMET would have a relatively lower correlation with intelligence than more explicit ToM tasks, that verbal ability would influence RMET performance more relative to performance ability, and that moderators would not influence the relationship between intelligence and RMET performance.

First we examined whether studies using only tests of verbal ability or performance ability or both could be combined in the analysis. Next, we combined all studies and analyzed the overall relationship. Finally, we examined whether or not any study variables would influence the results or if the overall effect size would hold up across different types of studies. Studies varied on a number of factors, and these were examined as moderators when appropriate. Since the RMET is widely used, it has been translated into different languages, administered in several countries, and used with various populations. The purpose of the meta-analysis was to aggregate the correlation across these studies, but it is vital that we examine these potentially confounding factors for effects on RMET performance. When a modified version of the RMET is used, it is typically not substantiated with a psychometric study, so we can only assume these task characteristics have not been examined. Thus, we do not know if the language of administration or modifications to the RMET have any impact on performance. We recognize that these factors may influence the results and warrant examination. Thus, the following moderators were examined: (a) verbal ability versus performance ability; (b) Wechsler Intelligence Tests versus other tests of intelligence; (c) English version versus translated versions; (d) revised version versus altered versions (e.g., Bora et al. (2006) used 27 items only; Botting and Conti-Ramsden (2008) read items aloud to participants); (e) studies conducted in the UK, where the task originated (38%), versus other countries; (f) typically developing samples versus those with disorders (e.g., autism, schizophrenia); (g) adult performance versus performance in children; (h) year of publication.

2. Methods

2.1. Literature search

We located articles through the following databases: Academic Search Premier, Eric, PsycINFO, Medline, CINAHL (EBSCO Host), PAIS International (CSA), ProQuest Dissertation & Theses, Social Sciences Citation Index, Scirus, and Sherpa through November 1, 2010 using the search terms: Reading the Mind in the Eyes, RMET, and Eyes Task. We also examined the reference lists of major theory of mind studies and meta-analyses. Studies returned from the search were then examined for use of the RMET and to determine if they met the inclusion criteria (outlined below). To ensure we included every possible study, we examined each article for other articles and performed a simple Google search. This exhaustive literature search returned over 250 studies using the RMET. Out of these, 55 studies measured intelligence, and 22 reported correlations between RMET performance and intelligence. When IQ was measured in a study but no correlation with the RMET was reported (this happened often presumably due to intelligence being a secondary concern for researchers), an email was sent to the author in an effort to obtain the data. If no reply was received, a second attempt was made to retrieve the data. Data from 29 studies (representing 58 effect sizes) were retrieved in this manner and included in the analysis.

2.2. Coding system

Each study was coded by two independent coders for the following variables: (a) authors; (b) year of publication; (c) source of study; (d) country of publication; (e) participant characteristics (diagnoses, gender, age, education, ethnicities, and socioeconomic status); (f) sample size; (g) study design; (h) version of RMET used in the study; (i) test(s) of intelligence used in the study and whether they measured full scale, verbal, or performance IQ; (j) means and standard deviations on the RMET and IQ tests; (k) correlation coefficient reported for the relationship between RMET performance and intelligence; and (l) sample size reported for the correlation between intelligence and RMET performance. If more than one measure of intelligence was used, correlations were coded for each IQ test.

2.3. Inclusion criteria

Articles were included in our sample according to the following inclusion criteria: (a) a version of the RMET should be used; (b) an IQ test or similar test of intelligence should be used; (c) studies should be reported in English; (d) studies should be reported between 1997 and 2010 (the original version of the RMET was published in 1997, and we concluded our literature search in 2010); (e) studies should include a

Table 1
Difference analysis for studies reporting only verbal intelligence (VIQ) and performance intelligence (PIQ).

Authors	Year	Group	VIQ & PIQ correl.	VIQ & RMET correl.	PIQ & RMET correl.	Hedges *g*	df
Botting and Conti-Ramsden	2008	Nonclinical	0.25	0.32	0.43	0.099	121
Botting and Conti-Ramsden	2008	SLI	0.56	0.32	0.43	0.123	131
Carroll and Yung	2006	Nonclinical	−0.111	0.092	−0.153	−0.161	45
Golan et al.	2007	AS/HFA & nonclinical	0.382[a]	0.12	0.104	−0.014	69
Henry, Phillips, et al.	2009	Nonclinical	0.382[a]	0.34	0.36	0.018	27
Henry, Phillips, et al.	2009	MS	0.382[a]	0.28	0.23	−0.044	24
Lawrence et al.	2003	Nonclinical females	0.631	0.176	0.301	0.145	35
Lawrence et al.	2003	Turner's syndrome	0.589	0.072	0.062	0.011	42
Lawrence et al.	2003	Nonclinical males	0.454	0.183	0.374	0.177	16
McGlade et al.	2008	Schizophrenia & nonclinical	0.584	0.321	0.296	−0.029	147
Phillips et al.	2002	Nonclinical	0.213	0.042	0.100	0.045	57

Note: correl. refers to correlation.
[a] The average correlation was used when authors failed to report a correlation.

correlation coefficient between the RMET and measure of general, verbal, or performance ability (or requested from the authors); and (f) each sample should be independent of every other sample.

2.4. Statistical analyses

To begin our analysis, we categorized studies by type of IQ test: full scale, verbal, or performance IQ. We then conducted a difference analysis to determine whether the IQ tests could be combined. That is, since some studies reported only verbal or performance scores, we ran type of IQ test as a moderator variable to determine if we should meta-analyze the studies together. We then completed a mean weighted effect size analysis, moderator analysis, heterogeneity testing, and tests for publication bias.

All effect sizes are reported as Pearson product–moment correlation coefficients (r). Effect sizes were analyzed using Fisher's z' transformation to account for the non-normal distribution of r (Lipsey & Wilson, 2001). Inverse error variances were used for weighting. Effect size transformations and weighted mean effect sizes were calculated using MetaWin, version 2.1 (Rosenberg, Adams, & Gurevitch, 2000). To examine the representativeness of the weighted mean effect size, homogeneity of the effect sizes across studies was tested using the Q statistic, a chi-square test for significance (Lipsey & Wilson, 2001). We also examined the extent to which the studies are heterogeneous using the I^2 statistic and potential moderator variables.

Table 2
Type of intelligence test as a moderator.

Type of test used	k	n	ES	95% Confidence interval Lower	Upper
Full scale	21	36	.24	.16	.32
Verbal & performance combined	7	11	.22	.09	.35
Verbal	13	18	.24	.13	.34
Performance	10	12	.25	.11	.39
Overall	51	77	.24	.19	.29

Note: k = number of studies; n = number of effect sizes.

2.4.1. Moderator analysis

We conducted a moderator analysis to investigate whether certain variables could explain some heterogeneity among the effect sizes (Lipsey & Wilson, 2001), including: (a) test of intelligence — Wechsler versus other tests; (b) participant type — nonclinical versus clinical samples; (c) version of RMET — unmodified versus modified versions; (d) language in which the RMET is administered — English versus other languages; (e) age of participants — child versus adult samples; (f) country in which the studies were conducted — UK (where the RMET originated) versus other countries; and (g) year of publication — older studies versus more recent studies based on a median split. We chose these moderator variables because there were a sufficient number of studies per category to warrant investigation. For example, we chose to examine Wechsler Tests versus other tests because there was a large number of Wechsler Tests (36) while the next largest category consisted of only seven.

2.4.2. Assessment of publication bias

Finally, we tested for publication bias using Rosenberg's (2005) fail-safe N to estimate the number of studies with null findings that would result in a nonsignificant mean effect size. We chose to use Rosenberg's estimate because it overcomes many of the limitations of Rosenthal's fail-safe N calculation. We also examined funnel and normal quantile plots for asymmetry. A larger number of studies fell below the weighted mean effect size than above it, making trim and fill unnecessary.

3. Results

3.1. Preliminary analyses

The correlations included in this meta-analysis range from −.34 to .80 with 13 negative and 64 positive. Seven studies representing eleven independent samples reported both verbal and performance scores separately (verbal ranged from .042 to .321, performance ranged from −.153 to .430). We calculated the difference between the verbal and performance correlations with the RMET for each study and ran a meta-analysis to investigate whether these differences were meaningful. Aggregated, the mean weighted effect size for the difference between verbal and performance correlations was only .04, 95% CI [−.04, .12], $Q(10, N = 11) = 5.26$, $p = .87$, $I^2 = 0$ (see Table 1). Due to this nonsignificant effect size, we averaged the verbal and performance correlations for each study, and variances were adjusted for the combined correlations (Borenstein, Hedges, Higgins, & Rothstein, 2009).

In order to determine if there were differences among studies that used different tests of intelligence (i.e., verbal, performance, full scale, or both performance and verbal), we ran "type of intelligence test" as a moderator variable. We found no difference for type of intelligence test: $Q(3, N = 77) = .13$, $p = .99$, $I^2 = 0$ (see Table 2).

3.2. Meta-analysis

Seventy-seven effect sizes were included in the meta-analysis. For a listing of all studies included in the analysis, see Table 3. Mean effect sizes and their confidence intervals for all studies are included in forest plots in Figs. 1 and 2. The means of 64 studies were on the positive side, whereas the other 13 fell on the negative side. The confidence intervals from 52 of the studies crossed the line of no effect while the confidence intervals from the remaining 25 studies fell completely on the positive side.

3.2.1. Fixed effect model

The fixed effect model did not fit the sample of effect sizes ($Q[76, N = 77] = 156.61$, $p < .001$), and 51% of the heterogeneity ($I^2 = 51.45$) across studies was due to factors other than sampling error. Thus, a random effects model was conducted to account for some of the random variance.

3.2.2. Random effects model

The random effects model fit the data $Q(76, N = 77) = 75.56$, $p = .49$, $I^2 = 0$. The overall mean weighted effect size revealed a moderate positive correlation between intelligence and RMET performance: $r = .24$, 95% CI [.19, .29], $p < .001$ (see Table 4). A random effects model is more suitable for this analysis, as it accounts for the random error that exists beyond sampling error.

3.2.3. Moderator analysis

Results of the moderator analysis are included in Table 5. The only variable that significantly moderated the results was test of intelligence, with Wechsler tests correlating with the RMET significantly less than other tests of ability (.15 versus .32). Year of publication, participant type, age, language, country, and version of RMET were nonsignificant.

3.2.4. Publication bias

A correlation between variance and effect size indicated this finding should not be due to publication bias (Kendall's tau = −.07, $p = .36$). Rosenberg's (2005) fail-safe N for a random effects model estimated that 375 studies would be needed to reduce the effect size to nonsignificant. Given that we included unpublished studies in our searching efforts, it is unlikely there are unpublished studies lurking in file drawers numbering this high; thus, there is no evidence to suggest that publication bias is present. The funnel plot (see Fig. 3) suggests the studies were pulled from a common population, as the range of effect sizes narrows as standard error decreases. One study stands out in the bottom left-hand corner of the funnel plot. Castelli et al. (2010) had an unordinarily small sample size for this group, which accounts for the large standard error compared to the other studies. More studies fall below the mean effect size than above it, hence, it was not necessary to conduct a trim and fill analysis since no publication bias exists in favor of significant results. Finally, the normal quantile plot for this group of studies shows us additional evidence that publication bias is not present in this sample. All studies fall within the 95% confidence intervals (see Fig. 4; Wang & Bushman, 1998).

4. Discussion

This meta-analysis indicates performance on the RMET positively correlates with intelligence ($r = .24$); this relationship does not favor performance or verbal ability, as correlations were both .24. These correlations are considered small (Cohen, 1988), and the small standard error (.06) indicates this effect size is robust. We can be confident that this effect falls within the .18 to .30 range, indicating the effect is stable. This suggests that there is a real relationship between intelligence and RMET that needs to be taken into account when measuring performance on the RMET. In the original articles reporting on the RMET, Baron-Cohen et al. (1997) and Baron-Cohen, Wheelwright, Hill et al. (2001) concluded that intelligence did not contribute to the social cognitive process involved in RMET performance. The result of our analysis indicates there is a real relationship between performance on the RMET and intelligence. While this effect is small, it may be particularly important to consider when making conclusions about group differences in clinical studies (e.g., autism, schizophrenia). In addition, this effect size is very close to the correlation between ToM tests and the RMET ($r = .29$; Kirkland et al., 2012); thus, it is important to note that this correlation between measures of social cognition could be largely impacted by intelligence. This estimate of the effect must be considered conservative (i.e., the true correlation would be higher if the RMET had high reliability), since this meta-analysis does not correct for attenuating factors that may reduce the magnitude of the true effect size (Hunter & Schmidt, 2004). In particular, it is likely to be attenuated by the imperfect reliability of the tasks involved. While many intelligence tests have high reliability, the reliability of the RMET is not known, but when measured, it has been low (e.g., Mar et al., 2006; Meyer & Shean, 2006). Thus, our effect size likely underestimates the true effect (Hunter & Schmidt, 2004). Notably, there was no difference between verbal and performance intelligence in this meta-analysis. Therefore, although the foils consist of verbal labels, performance on the RMET did not appear to be influenced by verbal ability any more than performance ability.

We failed to find any statistically significant contribution of moderator variables to the heterogeneity of our effect sizes with one exception. Test of intelligence was found to be significant, indicating the intelligence test used may be contributing to some of the heterogeneity in this analysis.

Table 3
Studies included in the meta-analysis.

Authors (Year)	Group[a]	N	Intelligence test[b]	Type of test[c]	Effect size (*r*)	RMET version[d]	Country[e]	Language	Reported in article or sent data[f]
Adler et al. (2010)	AS	15	WAIS-III BD	P	−0.13	ET-R	IL	Hebrew	S
Adler et al. (2010)	Normals	20	WAIS-III BD	P	0.15	ET-R	IL	Hebrew	S
Bailey and Henry (2008)	Normals younger & older	69	RPM	P	0.38	ET-R	AU	English	R
Baron-Cohen et al. (1997)	HFA/AS	16	WAIS-R	FS	−0.08	ET-O	UK	English	R
Baron-Cohen, Wheelwright, Hill et al. (2001)	15 ASD/239 normals	254	WAIS-R Brief	FS	0.09	ET-R	UK	English	R
Bora et al. (2006)	Sz	50	WAIS-R Verbal	V	0.25	ET-R – 27 items	TR	Turkish	R
Bora et al. (2007)	Sz	58	WAIS-R IT	V	0.09	ET-R	TR	Turkish	R
Botting and Conti-Ramsden (2008)	Normals	124	WISC-III/CELF-R	FS	0.38	ET-C aloud	UK	English	R
Botting and Conti-Ramsden (2008)	SLI	134	WISC-III/CELF-R	FS	0.38	ET-C aloud	UK	English	R
Brent et al. (2004)	ASD children	20	WISC-III	FS	0.16	ET-C – 27 items	UK	English	R
Brent et al. (2004)	Normals – children	20	WISC-III	FS	−0.07	ET-C – 27 items	UK	English	R
Camargo (2007)	Normals – male	48	AA – Vocab	V	0.41	ET-R	US	English	R
Camargo (2007)	Normals – female	189	AA – Vocab	V	0.27	ET-R	US	English	R
Carroll and Yung (2006)	Normals	48	WASI BD & Vocab	FS	−0.03	ET-R	UK	English	R
Castelli et al. (2010)	Normals – male	6	RPM	P	−0.34	ET-R – 24 items	IT	Italian	R
Castelli et al. (2010)	Normals – female	18	RPM	P	0.09	ET-R – 24 items	IT	Italian	R
Chapman et al. (2006)	Normals – children	76	WASI	FS	0.15	ET-C	UK	English	R
de Achaval et al. (2010)	Normals	40	ACE	FS	0.29	ET-R	AR	Spanish	S
de Achaval et al. (2010)	Sz	20	ACE	FS	0.44	ET-R	AR	Spanish	S
de Achaval et al. (2010)	Normals – Sz relatives	20	ACE	FS	0.8	ET-R	AR	Spanish	S
Demurie et al. (2011)	ASD	13	WISC-III	FS	0.19	ET-C	BE	Dutch	S
Demurie et al. (2011)	ADHD	13	WISC-III	FS	−0.09	ET-C	BE	Dutch	S
Dorris et al. (2004)	Normals	54	BPVS-II	V	0.41	ET-C	UK	English	R
Dziobek et al. (2006)	Normals	20	Shipley (WAIS)	FS	−0.17	ET-R – 24 items	US	English	R
Dziobek et al. (2006)	AS	19	Shipley (WAIS)	FS	0.28	ET-R – 24 items	US	English	R
Ferguson and Austin (2010)	Normals	153	QTB	V	−0.14	ET-R	UK	English	R
Garrido et al. (2009)	Normals	18	WASI	FS	0.32	ET-R	UK	English	S
Garrido et al. (2009)	PA	15	WASI	FS	−0.29	ET-R	UK	English	S
Golan et al. (2007)	50 AS/22 normals	72	WAIS	FS	0.11	ET-R	UK	English	R
Harrison et al. (2009)	20 normal/20 ANX	40	NART	V	0.26	ET-R	UK	English	S
Hassenstab et al. (2007)	Normals	38	Shipley (WAIS)	FS	0.21	ET-R – 24 items	US	English	S
Havet-Thomassin et al. (2006)	17 TBI & 17 normals	34	WAIS-R	FS	−0.07	ET-R	FR	French	S
Hefter et al. (2005)	SDD	26	WAIS	FS	0.3	ET-R	US	English	R
Henry, Phillips, et al. (2009)	Normals	30	Shipley/SEFCI	FS	0.35	ET-R	AU/UK	English	R
Henry, Phillips, et al. (2009)	MS	27	Shipley/SEFCI	FS	0.26	ET-R	AU/UK	English	R
Henry, Rendell, et al. (2009)	AD	20	ACE-R	FS	0.25	ET-R	AU	English	R
Hirao et al. (2008)	Sz	20	WAIS-R BD	P	−0.07	ET-R	JP	Jap.	R
Kaland et al. (2008)	Normals	20	WISC-III	V	−0.2	ET-R	DK	Danish	S
Kaland et al. (2008)	AS	21	WISC-III	V	0.43	ET-R	DK	Danish	S
Lau (2006)	Normals	36	C-MMSE	FS	0.39	ET-R	CN	Chinese	R
Lawrence et al. (2003)	Normals – male*	19	WAIS	FS	0.28	ET-R	UK	English	S
Lawrence et al. (2003)	Normals – female	38	WAIS	FS	0.24	ET-R	UK	English	R
Lawrence et al. (2003)	TS	45	WAIS	FS	0.07	ET-R	UK	English	R
Lawrence et al. (2004)	Normals	48	NART	V	0.39	ET-R	UK	English	R

Lysaker et al. (2010)	Sz	88	WAIS-III PA	P	0.44	ET-R	US	English	R
Mar et al. (2006)	Normals	94	WAIS-MR	P	0.001	ET-R	CA	English	R
McGlade et al. (2008)	73 Sz/77 normals	150	WAIS	FS	0.31	ET-R	IE	English	S
Montgomery (2007)	AS	25	WASI	V	0.04	ET-R	CA	English	R
Oldershaw et al. (2010)	Normals	46	NART	V	0.17	ET-R	UK	English	S
Phillips et al. (2002)	Normals	60	WAIS-III	FS	0.07	ET-O	UK	English	S
Plesa Skewerer et al. (2006)	WS	43	KBIT	FS	0.12	ET-R – 32 items	US	English	S
Plesa Skewerer et al. (2006)	LD	39	KBIT	FS	0.4	ET-R – 32 items	US	English	S
Plesa Skewerer et al. (2006)	Normals	46	KBIT	FS	0.55	ET-R – 32 items	US	English	S
Richell et al. (2003)	19 Ps/18 normals – male	37	RPM	P	0.17	ET-R	UK	English	R
Riveros et al. (2010)	15 Sz/32 normals	47	RPM	P	0.65	ET-R	CL	Spanish	R
Roca et al. (2010)	34 PD/35 normals	69	RPM	P	0.36	ET-O – 15 items	AR	Spanish	R
Roca et al. (2008)	12 MS/12 normals	24	RPM	P	0.12	ET-R – 17 items	AR	Spanish	R
Russell et al. (2009)	Anorexia	22	NART	V	0.26	ET-R	UK	English	R
Schwartz et al. (2010)	HFA/AS	20	WAIS-R	FS	0.23	ET-R – 24 items	DE	German	S
Schwartz et al. (2010)	Normals	20	WAIS-R	FS	−0.02	ET-R – 24 items	DE	German	S
Sharp (2008)	Normals – children	79	WISC	FS	0.03	ET-C	UK	English	R
Shaw et al. (2005)	Normals	91	NART	V	0.13	ET-R	UK	English	R
Shaw et al. (2005)	TBI temporal	54	NART	V	0.38	ET-R	UK	English	R
Shaw et al. (2005)	TBI frontal	31	NART	V	0.47	ET-R	UK	English	R
Slessor et al. (2007)	Normals – younger	40	Mill Hill	V	0.44	ET-R – 25 items	UK	English	S
Slessor et al. (2007)	Normals – older	40	Mill Hill	V	0.14	ET-R – 25 items	UK	English	S
Szily and Keri (2009)	MDD at risk for psychosis	26	WAIS-R	FS	0.21	ET-R	HU	Hung.	S
Szily and Keri (2009)	MDD	42	WAIS-R	FS	0.25	ET-R	HU	Hung.	S
Szily and Keri (2009)	Normals	50	WAIS-R	FS	0.31	ET-R	HU	Hung.	S
Tso et al. (2010)	Sz – male	22	WRAT3-R	FS	0.03	ET-R	US	English	S
Tso et al. (2010)	Sz – female	11	WRAT3R	FS	0.77	ET-R	US	English	S
Tso et al. (2010)	Normals – male	23	WRAT3R	FS	0.3	ET-R	US	English	S
Tso et al. (2010)	Normals – female	10	WRAT3R	FS	0.56	ET-R	US	English	S
Turkstra (2008)	Normals	19	KBIT	FS	0.19	ET-R	US	English	S
Turkstra (2008)	TBI	19	KBIT	FS	0.71	ET-R	US	English	S
Wang et al. (2008)	MDD	52	WAIS-R	FS	0.2	ET-R – 34 items	CN	Chinese	R
Wigan (2007)	Normals	60	MAB-II	FS	0.37	ET-R	UK	English	R

[a] AD = Alzheimer disease, AS = Asperger syndrome, ANX = anorexia, ASD = autism syndrome disorders, HFA = high-functioning autism, LD = learning disability, MDD = major depressive disorder, MS = multiple sclerosis, PA = prosopagnosia, PD = Parkinson disease, Ps = psychopathy, SDD = social developmental disorder, SLI = specific language impairment, Sz = schizophrenia, TBI = traumatic brain injury, TS = Turner syndrome, WS = William syndrome.

[b] AA = Army Alpha, ACE = Adult Cognitive Exam = Addenbrooke's Cognitive Examination, BD = Block Design, BPVS = British Picture Vocabulary Scale, CELF-R = Clinical Evaluation of Language Fundamentals – Revised, IT = Information Test, KBIT = Kaufman Brief Intelligence Test, MMSE = Mini Mental State Examination, MR = Matrix Reasoning, NART = National Adult Reading Test, PA = Picture Arrangement, QTB = Gf/Gc Quickie Test Battery, RPM = Raven's Progressive Matrices, SEFCI = Screening Examination for Cognitive Impairment, WAIS = Wechsler Adults Intelligence Scale, WASI = Wechsler Abbreviated Scale of Intelligence, WISC = Wechsler Intelligence Scale for Children, WRAT = Wide Range Achievement Test, VIQ = Verbal IQ, PIQ = Performance IQ.

[c] FS = full scale, V = verbal, P = performance.

[d] ET-O = Original version of RMET, ET-R = Revised version of RMET, ET-C = Children's version of RMET.

[e] AR = Argentina, AU = Australia, BE = Belgium, CA = Canada, CL = Chile, CN = China, DE = Germany, DK = Denmark, FR = France, IE = Ireland, IL = Israel, IT = Italy, JP = Japan, HU = Hungary, TR = Turkey, UK = United Kingdom.

[f] R = reported in article, S = sent by author.

* This group not included in published study.

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Baron-Cohen, Jolliffe, et al.	-0.080	-0.554	0.433	-0.289	0.773
Baron-Cohen, Wheelwright, et al.	0.090	-0.033	0.211	1.430	0.153
Botting & Conti-Ramsden*	0.376	0.249	0.491	5.483	0.000
Botting & Conti-Ramsden* b	0.376	0.239	0.498	5.104	0.000
Brent, Rios, et al.	0.160	-0.304	0.563	0.665	0.506
Brent, Rios, et al. b	-0.070	-0.497	0.384	-0.289	0.773
Carroll & Yung	-0.031	-0.222	0.163	-0.312	0.755
Chapman, Baron-Cohen	0.150	-0.078	0.363	1.291	0.197
de Achaval, Costanzo, et al.	0.291	-0.023	0.552	1.823	0.068
de Achaval, Costanzo, et al. b	0.442	-0.001	0.740	1.957	0.050
de Achaval, Costanzo, et al. c	0.800	0.553	0.918	4.530	0.000
Demurie, De Corel, & Roeyers	0.194	-0.400	0.673	0.621	0.534
Demurie, De Corel, & Roeyers b	-0.094	-0.613	0.482	-0.298	0.766
Dziobek, Fleck, et al.	-0.173	-0.572	0.292	-0.721	0.471
Dziobek, Fleck, et al. b	0.278	-0.202	0.650	1.142	0.253
Garrido, Furl, et al.	0.316	-0.177	0.682	1.267	0.205
Garrido, Furl, et al. b	-0.285	-0.696	0.266	-1.015	0.310
Golan, Baron-Cohen, & Rutherford	0.112	-0.083	0.299	1.125	0.261
Hassentab, Dziobek, et al.	0.215	-0.113	0.500	1.291	0.197
Havet-Thomassin, et al.	-0.070	-0.399	0.275	-0.390	0.696
Hefter, Manoach, & Barton	0.297	-0.102	0.614	1.467	0.142
Henry, Phillips, Beatty, et al.*	0.350	0.052	0.591	2.284	0.022
Henry, Phillips, Beatty, et al.* b	0.255	-0.072	0.532	1.537	0.124
Henry, Rendell, et al.	0.250	-0.216	0.624	1.053	0.292
Lau	0.390	0.070	0.637	2.366	0.018
Lawrence, Campbell, et al.	0.240	-0.052	0.495	1.614	0.107
Lawrence, Campbell, et al. b	0.067	-0.200	0.324	0.488	0.625
Lawrence, Campbell, et al. c	0.281	-0.128	0.608	1.355	0.175
McGlade, Behan, et al.	0.309	0.174	0.433	4.347	0.000
Phillips, MacLean, & Allen	0.071	-0.130	0.266	0.691	0.490
Plesa Skewerer et al.*	0.116	-0.191	0.402	0.737	0.461
Plesa Skewerer et al.* b	0.401	0.098	0.636	2.549	0.011
Plesa Skewerer et al.* v	0.548	0.306	0.723	4.036	0.000
Schwartz, Bente, et al.	0.226	-0.241	0.608	0.948	0.343
Schwartz, Bente, et al. b	-0.019	-0.458	0.427	-0.078	0.938
Sharp	0.030	-0.192	0.249	0.262	0.794
Szily & Keri	0.210	-0.193	0.552	1.022	0.307
Szily & Keri b	0.250	-0.058	0.515	1.595	0.111
Szily & Keri c	0.310	0.035	0.542	2.198	0.028
Tso, Grove, & Taylor	0.026	-0.400	0.443	0.113	0.910
Tso, Grove, & Taylor b	0.774	0.325	0.938	2.914	0.004
Tso, Grove, & Taylor c	0.299	-0.129	0.633	1.379	0.168
Tso, Grove, & Taylor d	0.559	-0.109	0.879	1.670	0.095
Turkstra	0.190	-0.289	0.593	0.769	0.442
Turkstra b	0.710	0.378	0.880	3.549	0.000
Wang, Wang, et al.	0.197	-0.080	0.446	1.397	0.162
Wigan	0.366	0.124	0.567	2.898	0.004
	0.233	0.173	0.293	7.345	0.000

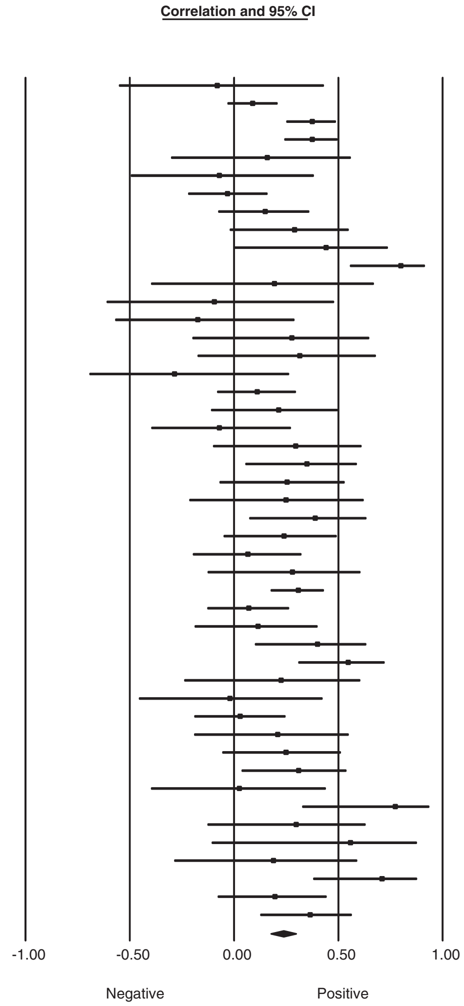

Fig. 1. Full scale and combined Forrest plot.

Assessment for publication bias suggests no bias is present; 375 more studies with null results would be needed to reduce the mean effect size to nonsignificant, and the funnel plot showed equal representation of studies around the mean effect size. Most authors were not primarily interested in the correlation between RMET and intelligence; thus, the secondary nature of the relationships reported reduces the likelihood of bias (Eagly & Wood, 1991).

When examined as a moderator variable, studies using Wechsler scales report a significantly lower correlation with the RMET ($r = .15$) than studies using other tests of intelligence ($r = .32$). This result is difficult to interpret due to the heterogeneity among studies that comprise the Wechsler group. For example, the Wechsler category includes many Wechsler IQ Tests (e.g., WAIS, WASI, and WISC). Even with the discrepancy between Wechsler and all other intelligence tests, both mean effect sizes indicate a significant relationship (small and moderate effects, respectively) between RMET performance and intelligence. While the recruiting procedures and IQ ranges did not indicate any clear differences between the groups, the Wechsler group included more non-normal samples than the other group (69% versus 37%, respectively). What drives this discrepancy is not clear, although it is typical in a meta-analysis that a construct will be measured using different instruments across studies (e.g., Murphy & Hall, 2011). In the current study, 36 effect sizes used Wechsler Intelligence Tests as their measure of intelligence, whereas 41 effect sizes were obtained from 12 different instruments. Examples include Addenbrooke's Cognitive Examination (e.g., de Achaval et al., 2010), Shipley's Institute of Living Scale (e.g., Henry et al., 2009), Raven's Colored Progressive Matrices (e.g., Riveros et al., 2010), and Kaufman Brief Intelligence Test (e.g., Plesa Skewerer, Verbalis, Schofield, Faja, & Tager-Flusberg, 2006). The most widely used instruments were Raven's Progressive Matrices and the National Adult Reading Test (NART; each was used in seven studies). These instruments have been shown to have construct validity in measuring intelligence. For example, in a cross-validation regression analysis, the NART (similar to a Wechsler verbal IQ

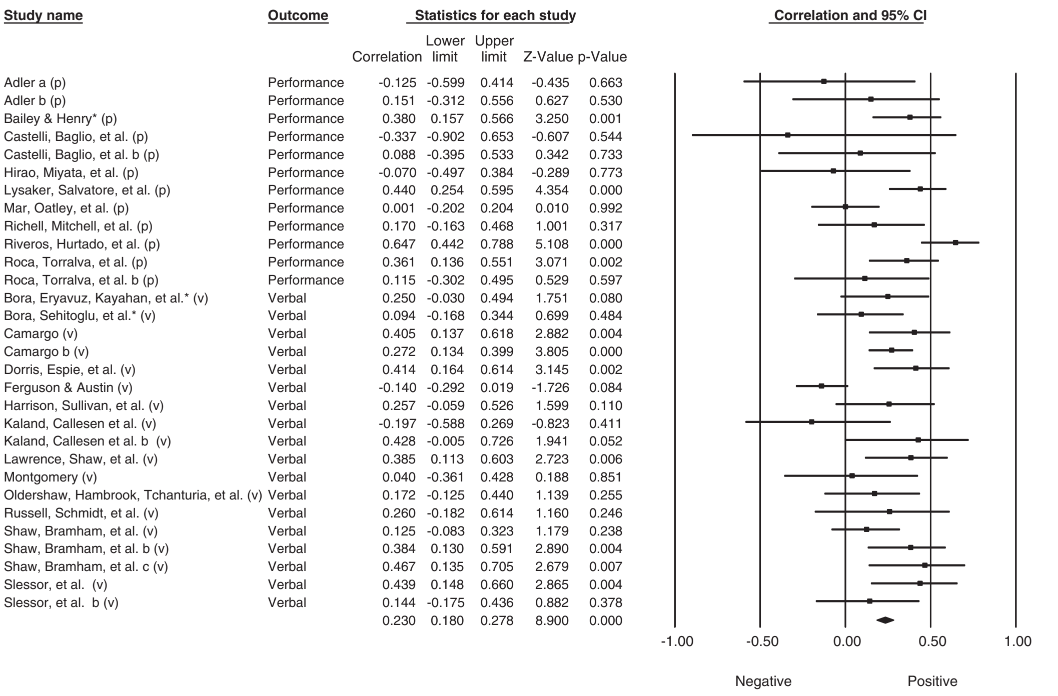

Fig. 2. Verbal and performance Forrest plot.

subtest) predicted 63% of the variance in WAIS Verbal IQ and 57% in full scale IQ (Crawford, Parker, Stewart, Besson, & De Lacey, 1989). Raven's Progressive Matrices (similar to a Wechsler performance IQ subtest) also correlated highly with general cognitive ability (Raven, 2000). While most instruments did not have enough studies to warrant a separate category as a moderator, we did run a follow-up analysis on the Raven's and NART due to the size of their groups. The NART and the Raven's had similar correlations to RMET performance (.29 and .33, respectively), which did not differ from the overall correlation.

Although many authors (e.g., Moran, 2013) report that the RMET is a rapid-processing measure relatively free from the constraints of intelligence, this analysis indicates performance on the RMET is significantly related to intelligence, verbal ability, and performance ability. The degree to which some theory of mind processes may be somewhat modular, operating independently from general intelligence resources, remains a theoretical question that drives current debate. Similarly, some authors have argued in favor of a dual route model in which social understanding in everyday situations involves both implicit, early-emerging, relatively automatic and inflexible processes and more explicit cognitively demanding flexible processes. From this perspective, performance in social perceptual processes such as reading face emotion should load relatively less on intelligence than performance in tasks like Strange Stories that require explicit analysis of meaning conveyed in discourse. It is important to emphasize that our analysis cannot provide a clear refutation of either a somewhat modular perspective or a dual route model. In either framework, we might expect real world performance in social cognitive tasks to involve a range of different component processes including general intelligence. However, while we cannot provide evidence against such models, we can say that the meta-analytic results across many studies warrant an abandonment of the position that RMET performance reflects a process that is "independent of general (nonsocial) intelligence" (Baron-Cohen, Wheelwright, Hill et al., 2001, p. 247).

At face value, the RMET does seem to require less reasoning than a false belief task involving a story to follow with multiple characters; however, this hypothesis remains to be tested. A meta-analysis between other specific ToM tasks and measures of intelligence is required to make a comparison between the differential loadings of intelligence on different ToM tasks and the RMET. Of note, a separate

Table 4
Weighted mean effect size for the correlation between RMET performance and intelligence.

Model	n	ES *r*	95% Confidence interval Lower	Upper	*Z*	*p*	Heterogeneity Q	df	*p*	I^2
Fixed	77	.23	.20	.25	14.12	<.001	156.61	76	<.001	51.45
Random	77	.24	.19	.29	9.31	<.001	75.56	76	.49	0

Table 5
Random effects moderator analysis.

	Category 1			Category 2						
Variable	Group 1	*r* 1	N 1	Group 2	*r* 2	N 2	*Q*	df	*p*	I^2
IQ test[a]	Wechsler	.15	36	Other	.32	41	13.03	1	<.001	92.33
Participant type[b]	Nonclinical	.22	42	Clinical	.26	35	.53	1	.466	0
Language	English	.24	52	Other	.24	25	.01	1	.910	0
Age	Adult	.24	66	Child	.22	11	.07	1	.800	0
Version of ET[c]	Unmodified	.23	57	Modified	.26	20	.22	1	.640	0
Country	UK	.19	29	Other	.27	48	2.97	1	.083	66.33
Year of publication	1997–2007	.21	35	2008–2010	.28	42	1.67	1	.200	40.12

[a] Two independent samples were included from Botting and Conti-Ramsden (2008) with the performance test coming from WISC-III and the verbal test being Clinical Evaluation of Language Fundamentals — Revised. Due to this unique situation, one sample was coded as other tests (n = 134) and the other was coded as Wechsler (n = 124).

[b] Baron-Cohen, Wheelwright, Hill et al.'s (2001) sample included 15 individuals with autism and 239 normals and was coded as nonclinical; Golan, Baron-Cohen, Hill, and Rutherford's (2007) sample included 50 AS/HFA and 22 normals and was coded as clinical; Havet-Thomassin, Allain, Etcharry-Bouyx, and Le Gall's (2006) sample included 17 individuals with traumatic brain injury and 17 normals and was coded as nonclinical; McGlade et al.'s (2008) sample included 73 individuals with schizophrenia and 78 normals and was coded as nonclinical.

[c] Modified is defined as any modification from the intended administration of the task. For example, some studies only administered 24 items instead of the full 36 or read items aloud to participants rather than having them read on their own. A different language was not considered a modification, as these were analyzed with language as a moderator variable.

meta-analysis (Kirkland et al., 2012) found a mean correlation between the RMET and both the Strange Stories Task (SST) and the Faux Pas Test at $r = .29$. Perhaps intelligence is driving the association between RMET and SST and Faux Pas performance; however, at this time this hypothesis remains untested.

Although the RMET was originally purported to be a relatively implicit task, our results indicate it draws on general cognitive resources. While the specific role of intelligence in ToM tasks remains unknown, it may be fruitful to conduct within subject analyses of the contribution of general intelligence to a range of tasks that might be expected to require differing degrees of explicit processing such as verbal reasoning. It should be noted that there is insufficient evidence to determine whether the RMET in particular loads on processing speed relative to other kinds of theory of mind tasks. Prior to beginning the RMET, participants are instructed: "You should try to do the task as quickly as possible, but you will not be timed." In our own observation of more than 200 students taking this task in paper-and-pencil format, students appear to be attentive but relaxed during the task and do not appear to be considering time constraints. It may be that in computerized versions of the task, the task setup encourages faster performance. In one recent experiment with 86 participants (Experiment 1, Kidd & Castano, 2013), time spent on each item in a computerized format increased performance on the RMET. Further analyses of time spent per item would enable a more refined examination of the potential role of processing speed and perhaps other factors. But, again, such questions could only be addressed with data from computerized versions of the task, and it may well be that the computer format elicits a different, faster approach from participants. To date, we are not

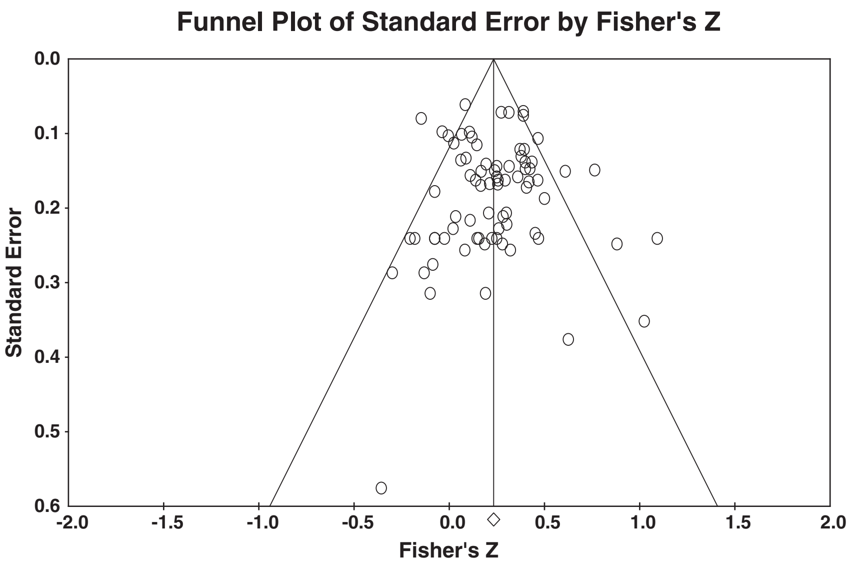

Fig. 3. Funnel plot.

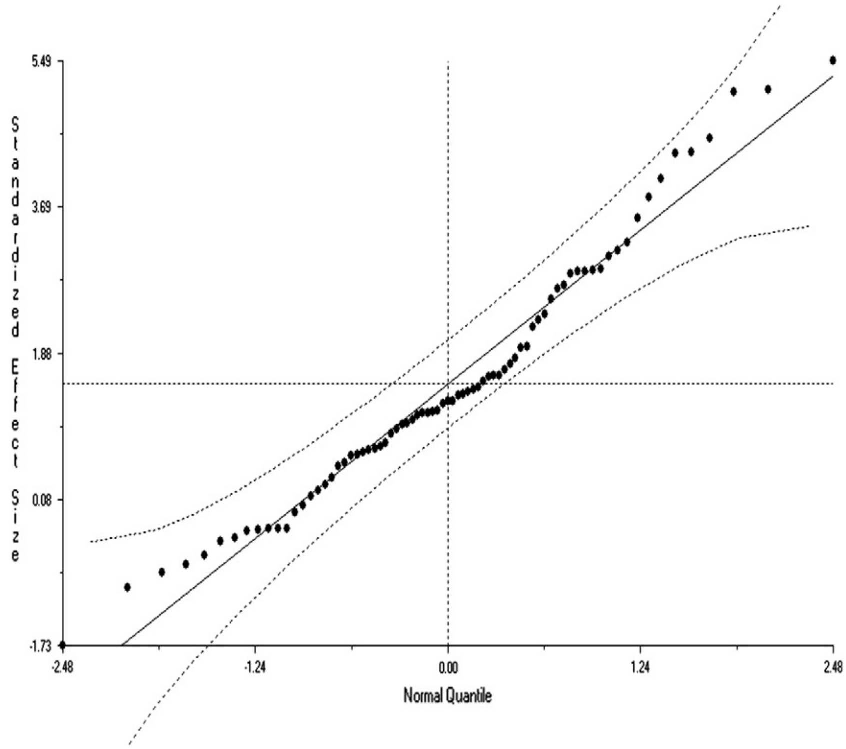

Fig. 4. Normal quantile plot.

aware of any other studies that record time spent per trial; thus, we can only speculate based on this one finding. To the degree that time spent on each item increases performance, this finding calls into question the degree to which RMET is a relatively more implicit task.

This meta-analysis did not find a significant difference in the relations between RMET performance with verbal and performance intelligence. Stone and Gerrans (2006) suggest that future research should examine the relationship between performance on both verbal and nonverbal measures of ToM (e.g., Apperly, Samson, Chiavarino, & Humphreys, 2004) and differential aspects of intelligence. The RMET is often used in studies discriminating disordered populations from typically developing populations, and some of these studies use only the RMET to indicate problems in social cognition (e.g., Tso, Grove, & Taylor, 2010). Many such disordered groups (e.g., autism, schizophrenia, acquired brain injury) may differ on general intelligence factors leading to difficulty of interpretation. In the case of autism, the long held expectation of a predicted IQ asymmetry (i.e., performance IQ greater than verbal IQ) has received some support in an epidemiological study (Charman et al., 2011), making the issue of matching participants on intelligence more complex. Future studies comparing clinical and nonclinical groups should account for possible group differences in intelligence when making inferences based on RMET performance.

In addition to controlling for IQ, future studies using the RMET should address a few issues. First, the computerized version of the task can be used to examine the potential role of response style in performance on the RMET. At least some evidence (Kidd & Castano, 2013) suggests that increased time spent per trial may contribute to higher scores. If, indeed, deliberation correlates with increased performance, the original notion that RMET reflects predominantly implicit processes seems less compelling. Second, future studies should make direct comparisons between the relative contributions of IQ on the RMET to tasks such as the Strange Stories Task that very clearly require rich linguistic processing. Third, a more refined examination of the potential underlying cognitive processes that may be driving the RMET and IQ association would further our understanding. Future studies may examine the relative contributions of such processes as working memory, novel problem solving, processing speed, and vocabulary knowledge.

In conclusion, a small mean effect size correlation was found in this meta-analysis examining the relationship between intelligence and RMET performance with no difference between verbal and performance IQ. Given the RMET's success at identifying individual differences among adult samples, the instrument will undoubtedly continue to be used in numerous studies. The current study begins to disambiguate the contribution of constructs aggregated in RMET performance, and further research on the cognitive underpinnings will assist researchers in interpreting their findings using this instrument.

Acknowledgments

We would like to thank all of the authors who kindly responded to our request for data to include in this analysis. We would also like to thank Stephanie Miller and Cynthia Johnson for all of their coding efforts, and Robin Peterson for her contributions.

References

Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., & Ambady, N. (2010). Cross-cultural Reading the Mind in the Eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97–108.

Adler, N., Nadler, B., Eviatar, Z., & Shamay-Tsoory, S. G. (2010). The relationship between theory of mind and autobiographical memory in high functioning autism and Asperger syndrome. *Psychiatry Research*, *178*, 214–216.

Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Research*, *1079*, 25–35.

Apperly, I. A., Samson, D., Chiavarino, C., Bickerton, W. L., & Humphreys, G.W. (2007). Testing the domain-specificity of a theory of mind deficit in brain-injured patients: Evidence for consistent performance on non-verbal, "reality-unknown" false belief and false photograph tasks. *Cognition*, *103*, 300–321.

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, *16*(10), 1773–1784.

Bailey, P. E., & Henry, J. D. (2008). Growing less empathic with age: Disinhibition of the self-perspective. *Journal of Gerontology*, *63B*, P219–P226.

Baron-Cohen, S., & Hammer, J. (1997). Parents of children with Asperger syndrome: What is the cognitive phenotype? *Journal of Cognitive Neuroscience*, *9*(4), 548–554.

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's syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822.

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

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome. *Journal of Developmental and Learning Disorders*, *5*, 47–78.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Science*, *11*(8), 327–332.

Beck, L., Kumschick, I. R., Eid, M., & Klann-Delius, G. (2012). Relationship between language competence and emotional competence in middle childhood. *Emotion*, *12*(3), 503–514.

Billington, J., Baron-Cohen, S., & Wheelwright, S. (2007). Cognitive style predicts entry into physical sciences and humanities: Questionnaire and performance tests of empathy and systemizing. *Learning and Individual Differences*, *17*, 260–268.

Bora, E., Eryavuz, A., Kayahan, B., Sungu, G., & Veznedaroglu, B. (2006). Social functioning, theory of mind and neurocognition in outpatients with schizophrenia; mental state decoding may be a better predictor of social functioning than mental state reasoning. *Psychiatry Research*, *145*, 95–103.

Bora, E., Sehitoglu, G., Aslier, M., Atabay, I., & Veznedaroglu, B. (2007). Theory of mind and unawareness of illness in schizophrenia: Is poor insight a mentalizing deficit? *European Archives of Psychiatry and Clinical Neuroscience*, *257*(2), 104–111.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* UK: John Wiley & Sons, Ltd.

Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the function social outcomes of young adolescents with and without a history of SLI. *British Journal of Developmental Psychology*, *26*, 281–300.

Brand, C. (1987). The importance of general intelligence. In S. Modgil, & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 251–265). Philadelphia, PA: The Falmer Press.

Brent, E., Rios, P., Happe, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism*, *8*, 283–299.

Camargo, M. A. (2007). Hypothesized fitness indicators and mating success. (Unpublished master thesis.) University of New York, New Paltz, New York.

Carroll, J. M., & Yung, C. K. (2006). Sex and discipline differences in empathizing, systemizing and autistic symptomatology: Evidence from a student population. *Journal of Autism and Developmental Disorders*, *36*, 949–957.

Castelli, I., Baglio, F., Blasi, V., Alberoni, M., Falini, A., Liverta-Sempio, O., & Marchetti, A. (2010). Effects of aging on mindreading ability through the eyes: An fMRI study. *Neuropsychologia*, *48*, 2586–2594.

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the empathy quotient (EQ) and the "Reading the Mind in the Eyes" Test. *Social Neuroscience*, *1*(2), 135–148.

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: Data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*, 619–627.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: L. Eralbuam Associates.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research*, *17*, 5–13.

Craig, J. S., Hatton, C., Craig, F. B., & Bentall, R. P. (2004). Persecutory beliefs, attributions and theory of ind: Comparison of patients with paranoid delusions, Asperger's syndrome and health controls. *Schizophrenia Research*, *69*, 29–33.

Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A. O., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology*, *28*, 267–273.

David, N., Gawronski, A., Santos, N. S., Huff, W., Lehnhardt, F. G., Newen, A., & Vogeley, K. (2008). Dissociation between key processes of social cognition in autism: Impaired mentalizing but intact sense of agency. *Journal of Autism and Developmental Disorders*, *38*, 593–605.

de Achaval, D., Costanzo, E. Y., Vilarreal, M., Jauregui, I. O., Chiodi, A., Castro, M. N., Fahrer, R. D., Leiguarda, R. C., Chu, E. M., & Guinjoan, S.M. (2010). Emotion processing and theory of mind in schizophrenia patients and their unaffected first-degree relatives. *Neuropsychologia*, *48*, 1209–1215.

Demurie, E., De Corel, M., & Royers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorder and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders*, *5*(1), 126–134.

Dorris, L., Espie, C. A. E., Knott, F., & Salt, J. (2004). Mind-reading difficulties in the siblings of people with Asperger's syndrome: Evidence for a genetic influence in the abnormal development of a specific cognitive domain. *Journal of Child Psychology and Psychiatry*, *45*, 412–418.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*, 623–636.

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin*, *17*, 306–315. http://dx.doi.org/10.1177/0146167291173011.

Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on theory of mind tasks in an adult sample. *Personality and Individual Differences*, *49*(5), 414–418.

Garrido, L., Furl, N., Draganski, B., Weiskopf, N., Stevens, J., Tan, C. G., & Duchaine, B. (2009). Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. *Brain*, *132*, 3443–3455.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' Test — Revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism Developmental Disorders*, *37*, 1096–1106.

Hallerback, M. U., Lugnegard, T., Hjarthag, F., & Gillberg, C. (2009). The Reading the Mind in the RMET: Test–retest reliability of a Swedish version. *Cognitive Neuropsychiatry*, *14*, 127–143.

Happe, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154.

Harrison, A., Sullivan, S., Tchanturia, K., & Treasure, J. (2009). Emotion recognition and regulation in anorexia nervosa. *Clinical Psychology and Psychotherapy*, *16*, 348–356.

Hassenstab, J., Dziobek, I., Rogers, K., Wolf, O. T., & Convit, A. (2007). Knowing what others know, feeling what others feel: A controlled study of empathy in psychotherapists. *The Journal of Nervous and Mental Disease*, *195*(4), 277–281.

Havet-Thomassin, V., Allain, P., Etcharry-Bouyx, F., & Le Gall, D. (2006). What about theory of mind after severe brain injury? *Brain Injury*, *20*(1), 83–91.

Hefter, R. L., Manoach, D. S., & Barton, J. J. S. (2005). Perception of facial expression and facial identity in subjects with social developmental disorders. *Neurology*, *65*, 1620–1625.

Henry, J. D., Phillips, L. H., Beatty, W. W., McDonald, S., Longley, W. A., Joscelyne, A., & Rendell, P. G. (2009). Evidence for deficits in facial affect recognition and theory of mind in multiple sclerosis. *Journal of the International Neuropsychological Society*, *15*(2), 277–285.

Henry, J. D., Rendell, P. G., Scicluna, A., Jackson, M., & Phillips, L. H. (2009). Emotion experience, expression, and regulation in Alzheimer's disease. *Psychology and Aging*, *24*(1), 252–257.

Henry, P. J., Sternberg, R. J., & Grigorenko, E. L. (2005). Capturing successful intelligence through measures of analytic, creative, and practical skills. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 295–311). Thousand Oaks, CA: Sage Publications.

Hirao, K., Miyata, J., Fujiwara, H., Yamada, M., Namiki, C., Shimizu, M., & Murai, T. (2008). Theory of mind and frontal lobe pathology in schizophrenia: A voxel-based morphometry study. *Schizophrenia Research*, *105*, 165–174.

Hunt, E. (2005). Information processing and intelligence: Where we are and where we are going. In R. J. Sternberg, & J. E. Pretz (Eds.), *Cognition & intelligence: Identifying the mechanisms of the mind* (pp. 1–25). Cambridge, UK: Cambridge University Press.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA: Sage Publications, Inc.

Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders*, *37*, 1679–1690.

Ingersoll, B. (2010). Broader autism phenotype and nonverbal sensitivity: Evidence for an association in the general population. *Journal of Autism and Developmental Disorders*, *40*(5), 590–598.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kaland, N., Callesen, K., Moller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced theory of mind tasks. *Journal Autism Developmental Disorders*, *38*(6), 1112–1123.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*(6156), 377–380.

Kirkland, R. A., Baker, C. A., Johnson, C., Peterson, E., & Pulos, S. (May 2012). *Meta-analysis reveals a moderate relationship between tests of theory of mind and the RMET. Poster presented at the Association for Psychological Science annual conference.* (IL Chicago).

Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis reveals adult female superiority in "Reading the Mind in the Eyes Test". *North American Journal of Psychology*, *15*(1), 121–146.

Kline, P. (1991). *Intelligence: The psychometric view.* New York, NY: Routledge.

Lau, J. Y. H. (2006). *Age-related effect on emotion recognition.* China: The University of Hong Kong Unpublished master thesis.

Lawrence, K., Campbell, R., Swettenham, J., Terstegge, J., Akers, R., Coleman, M., & Skuse, D. (2003). Interpreting gaze in Turner syndrome: Impaired sensitivity to intention and emotion, but preservation of social cueing. *Neuropsychologia*, *41*, 894–905.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological Medicine*, *34*, 911–924.

Leslie, A. M. (1987). Pretense and representation: The origins of theory of mind. *Psychological Review*, *94*, 412–426.

Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development*, *61*(5), 1514–1519.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications, Inc.

Lysaker, P. H., Salvatore, G., Grant, M. L. A., Procacci, M., Olesek, K. L., Buck, K. D., & Dimaggio, G. (2010). Deficits in theory of mind and social anxiety as independent paths to paranoid features in schizophrenia. *Schizophrenia Research*, *124*, 81–85.

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*, *40*, 694–712.

Matthews, G., Zeidner, M., & Roberts, R. D. (2005). Emotional intelligence: An elusive ability? In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 79–99). Thousand Oaks, CA: Sage Publications.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507–536.

McGlade, N., Behan, C., Hayden, J., O'Donoghue, T., Peel, R., Haq, F., & Donohoe, G. (2008). Mental state decoding v. mental state reasoning as a mediator between cognitive and social function in psychosis. *The British Journal of Psychiatry*, *193*, 77–78.

Meyer, J., & Shean, G. (2006). Social–cognitive functioning and schizotypal characteristics. *The Journal of Psychology*, *140*(3), 199–207.

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

Montgomery, J. (2007). Asperger syndrome and emotional intelligence. (Unpublished doctoral dissertation.) University of Saskatchewan, Saskatoon.

Moran, J. M. (2013). Lifespan development: The effects of typical aging on theory of mind. *Behavioural Brain Research*, *237*, 32–40.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and interpersonal sensitivity: A meta-analysis. *Intelligence*, *39*, 54–63.

Oldershaw, A., Hambrook, D., Tchanturia, K., Treasure, J., & Schmidt, U. (2010). Emotional theory of mind and emotional awareness in recovered anorexia nervosa patients. *Psychosomatic Medicine*, *72*(1), 73–79.

Peterson, E., & Miller, S. F. (2012). The RMET as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology: Personality Science and Individual Differences*, *3*, 1–6.

Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *Journal of Gerontology*, *57*, 526–530.

Plesa Skewerer, D. P., Verbalis, A., Schofield, C., Faja, S., & Tager-Flusberg, H. (2006). Social–perceptual abilities in adolescents and adults with Williams' syndrome. *Cognitive Neuropsychology*, *23*(2), 338–349.

Pons, F., Lawson, J., Harris, P. L., & de Rosnay, M. (2003). Individual differences in children's emotion understanding: Effects of age and language. *Scandinavian Journal of Psychology*, *44*, 347–353.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral Brain Science*, *4*, 515–526.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*, 1–48.

Richell, R. A., Mitchell, D. G. V., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. J. R. (2003). Theory of mind and psychopathy: Can psychopathic individuals read the 'language of the eyes'? *Neuropsychologia*, *41*, 523–526.

Riveros, R., Manes, F., Hurtado, E., Escobar, M., Martin Reyes, M., Cetkovich, M., & Ibanez, A. (2010). Context-sensitive social cognition is impaired in schizophrenia patients and their healthy relatives. *Schizophrenia Research*, *116*(2), 297–298.

Roca, M., Torralva, T., Gleichgerrcht, E., Chade, A., Arevalo, G. G., Gershanik, O., & Manes, F. (2010). Impairments in social cognition in early medicated and unmedicated Parkinson disease. *Cognitive & Behavioral Neurology*, *23*(3), 152–158.

Roca, M., Torralva, T., Meli, F., Fiol, M., Calcagno, M. L., Carpintiero, S., & Correale, J. (2008). Cognitive deficits in multiple sclerosis correlate with changes in fronto-subcortical tracts. *Multiple Sclerosis Journal*, *14*(3), 364–369.

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464–468.

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (2000). *MetaWin: Statistical software for meta-analysis: Version 2.0.* Sinauer Associates.

Russell, T. A., Schmidt, U., Doherty, L., Young, V., & Tchanturia, K. (2009). Aspects of social cognition in anorexia nervosa: Affective and cognitive theory of mind. *Psychiatry Research*, *168*(3), 181–185.

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition*, *55*, 209–219.

Schwartz, C., Bente, G., Gawronski, A., Schilbach, L., & Vogeley, K. (2010). Response to nonverbal behaviour of dynamic virtual characters in high-functioning autism. *Journal of Autism and Developmental Disorders*, *40*(1), 100–111.

Sharp, C. (2008). Theory of mind and conduct problems in children: Deficits in reading the emotions of the eyes. *Cognition and Emotion*, *22*(6), 1149–1158.

Shaw, P., Bramham, J., Lawrence, E. J., Morris, R., Baron-Cohen, S., & David, A.S. (2005). Differential effects of lesions of the amygdala prefrontal cortex on recognizing facial expression of complex emotions. *Journal of Cognitive Neuroscience*, *17*(9), 1410–1419.

Siegal, M., & Varley, R. (2002). Neural systems involved in 'theory of mind'. *Nature Reviews Neuroscience*, *3*, 463–471.

Slessor, G., Phillips, L. H., & Bull, R. (2007). Exploring the specificity of age-related differences in theory of mind tasks. *Psychology and Aging*, *22*, 639–643.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656.

Stone, V. E., & Gerrans, P. (2006). What's domain-specific about theory of mind? *Social Neuroscience*, *1*(3–4), 309–319.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*, 401–426.

Strong, E., Russell, R., Germine, L., & Wilmer, J. (2011). Face processing abilities relate to career choice. *Journal of Vision*, *11*(11) (article 621).

Szily, E., & Keri, S. (2009). Anomalous subjective experience and psychosis risk in young depressed patients. *Psychopathology*, *42*, 229–235.
Tso, I. F., Grove, T. B., & Taylor, S. F. (2010). Emotional experience predicts social adjustment independent of neurocognition and social cognition in schizophrenia. *Schizophrenia Research*, *122*, 156–163.
Turkstra, L. (2008). Conversation-based assessment of social cognition in adults with traumatic brain injury. *Brain Injury*, *22*(5), 397–409.
Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*, *65*, 71–95.
Wang, M. C., & Bushman, B. J. (1998). Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods*, *3*(1), 46–54.
Wang, Y., Wang, Y., Chen, S., Zhu, C., & Wang, K. (2008). Theory of mind disability in major depression with or without psychotic symptoms: A componential view. *Psychiatry Research*, *161*, 153–161.
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.
Wigan, E. (2007). Is general intelligence or social intelligence related to social network size? (Unpublished undergraduate thesis.) The University of Edinburgh, Scotland.
Wilhelm, O., & Engle, R. W. (2005). Intelligence: A diva and a workhorse. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 1–9). Thousand Oaks, CA: Sage Publications.
Woodward, L. J., & Fergusson, D. M. (2000). Childhood peer relationship problems and later risks of educational underachievement and unemployment. *Journal of Child Psychology and Psychiatry*, *41*(2), 191–201.
Wynn, J. K., Sugar, C., Horan, W. P., Kern, R., & Green, M. (2010). Mismatch negativity, social cognition, and functioning in schizophrenia patients. *Biological Psychiatry*, *67*(10), 940–947.