# Validity of Social Cognition Measures in the Clinical Services for Autism Spectrum Disorder

Maria Chiara Pino[1,2]*, Francesco Masedu[1], Roberto Vagnetti[1], Margherita Attanasio[1], Chiara Di Giovanni[2], Marco Valenti[1,2] and Monica Mazza[1,2]

[1] Department of Applied Clinical Sciences and Biotechnology, University of L'Aquila, L'Aquila, Italy, [2] Regional Centre for Autism, Abruzzo Region Health System, L'Aquila, Italy

The current study evaluated three social cognition (SC) tests for their clinical utility in aiding autism diagnosis. To do so, we compared the performance of 86 children with autism spectrum disorder (ASD) and 68 typically developing (TD) children, all aged from 4 to 10 years old, on three SC tasks [the Social Information Processing Interview (SIPI), the Comic Strip Task (CST), and the children's version of the Eyes Task] and calculated threshold scores that best differentiated the two groups. While difficulties in these abilities appear to represent the "central core" of ASD, services have largely ignored SC tests when supporting autism diagnoses. Therefore, this study attempted to validate and evaluate the diagnostic potential of these three tasks for children with ASD. To investigate the accuracy of these SC tests, we used the receiver operating characteristic (ROC) curve. As expected, the ASD group performed worse than the TD group on the SIPI and CST, but contrary to our prediction, the groups did not significantly differ on the Eyes Task. Specifically, the overall area under the curve (AUC) for the SIPI was 0.87, with a sensitivity of 73.5% and a specificity of 83.9% at the best cutoff point (score range 0–36; best cutoff = 31). The overall AUC for the CST was 0.75, with a sensitivity of 71.1% and a specificity of 77.0% at the best cutoff point (score range 0–15; best cutoff = 11). The overall AUC for the Eyes Task was 0.51, with a sensitivity of 50.3% and a specificity of 40.2% at the best cutoff point (score range 0–54; best cutoff = 45). In conclusion, the results showed that the SIPI test has good predictive power for classifying children with ASD. It should provide substantial supplementary clinical information and help to consolidate diagnostic procedures based on standard tools. Moreover, the results of the study have substantial implications for clinical practice: the better the knowledge of SC functioning in children with ASD, the more effective the intervention program for rehabilitation.

Keywords: autism spectrum disorder, diagnostic process, receiver operating characteristic (ROC) curve, social cognition, Theory of Mind, clinical utility

**Abbreviations:** ADOS-2, Autism Diagnostic Observation Schedule–Second Edition; ASD, autism spectrum disorder; AUC, area under the curve; CST, Comic Strip Task; DSM-5, *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*; ROC, receiver operating characteristic; SC, social cognition; SIPI, Social Information Processing Interview; TD, typically developing; ToM, Theory of Mind; TROG-2, Test for Reception of Grammar–Second Edition; VMA, verbal mental age.

# INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopment condition characterized by deficits in two domains: (1) social communication and social interaction and (2) restricted, repetitive patterns of behaviors, interests, or activities (American Psychiatric Association, 2013). A large body of research supports the hypothesis that difficulties in social interaction and communication can be explained by a deficit in social cognition (SC) abilities (Happé and Frith, 2014; Lai et al., 2014; Mazza et al., 2017; Pino et al., 2017). SC is a set of cognitive abilities involved in the processing and interpretation of the social world (Mazza et al., 2010; Bishop-Fitzpatrick et al., 2017; Pino et al., 2017). A main component of SC is the Theory of Mind (ToM), namely the ability to understand the mental and emotional states of other people (Mazza et al., 2014); it affects the development of social behavior from birth. A crucial development of ToM occurs around 3–4 years of age, when children acquire false belief attributions and realize that mental states, such as beliefs or the intentions of other people, may not be true (Mazza et al., 2017). Thus, ToM deficits are related to social communication and social interaction criteria of the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5).

Several studies (Shamay-Tsoory and Aharon-Peretz, 2007; Shamay-Tsoory et al., 2010; Mazza et al., 2014; Baron-Cohen et al., 2015; Pino et al., 2017) suggest that ToM is not a unitary construct; rather, it involves two distinct components: cognitive and affective. Specifically, the cognitive component of ToM includes the ability to understand what other people are thinking and make inferences about their beliefs, intentions, and motivations. The affective ToM component is the ability to understand what other people are feeling in a specific emotional context and comprehend their emotions. Understanding another person's cognitive or affective state is a crucial ability for development and production of adequate social behaviors (Krebs and Russell, 1981; Hoffman, 1984; Batson, 1987; Mazza et al., 2017).

According Happé and Frith (2014), social behavior develops around 5 years of age, when children are able to differentiate their own internal states form those of others (Mazza et al., 2017). Children with ASD show difficulties in understanding other people's mental state and their perspectives, and this deficit might compromise social behavior development (Frith and Happé, 1994; Happé, 1994; Frith and Frith, 2003; Jones et al., 2010; Mazza et al., 2014; Ziv et al., 2014).

The ToM hypothesis of ASD was first introduced by Baron-Cohen et al. (1985) three decades ago; it demonstrates difficulties for children with ASD in passing false belief tasks. Recent studies suggest that adults with ASD have difficulties in implicit mentalization tasks (measured by spontaneous looking patterns), despite the fact that they can pass classic explicit mentalizing tasks (direct questions about others mental states; Jones et al., 2018). Differentiation between the theoretical ToM components is crucial for future research in ASD (Altschuler et al., 2018).

Some mentalizing tests, such as the Eyes Test (Baron-Cohen et al., 1997, 2001), require emotion recognition to infer mental states (Jones et al., 2018). This test should reflect the mentalizing process and the ability to understand other's mental states, such as emotions, thoughts, desires, beliefs, and goals (Peterson and Slaughter, 2009; Franco et al., 2014). Children and adults with ASD present lower performance on the Eyes Test (Baron-Cohen et al., 2001; Franco et al., 2014). Specifically, individuals with ASD have difficulties in processing information from the faces of others, such as facial expression and eye gaze, which play a significant role in SC (Hadjikhani et al., 2004; Pellicano et al., 2007; Ramachandran et al., 2010; Franco et al., 2014).

Deficits of social interaction in individuals with ASD are not related to general intellectual functioning. Rather, they are specific to the SC competences (Baron-Cohen et al., 1997; Ziv et al., 2014; Mazza et al., 2017). Ziv and Sorongon (2011), following Crick and Dodge's (1994) social information processing model, suggested that many mental steps occur before individuals implement a behavioral response to social cues, such as the encoding of social cues, interpretation of the cues, clarification of goals, generation of a behavioral response, response construction, response decision, and realization of the behavior response (Crick and Dodge, 1994; Ziv and Sorongon, 2011; Ziv et al., 2014; Mazza et al., 2017). According to this model, these internal processes include the ability to understand thoughts, intentions, and feelings of others (ToM) and select the adequate social responses (Crick and Dodge, 1994). Subsequently, Ziv et al. (2014) showed deficits in social information processing abilities in preschool children with ASD using the Social Information Processing Interview (SIPI), an instrument that allows one to evaluate social behavior and the pattern of social information processing based on Crick and Dodge's (1994) model. Ziv et al. (2014) demonstrated that children with ASD had a specific difficulty in social information processing; the ToM deficits were related to inadequate social behavior and poor social communication skills (Lerner et al., 2011; Ziv et al., 2014; Mazza et al., 2017). According to Mazza et al. (2017), social behavior is a consequence of how children process social cues. Considering that severe difficulties in social interaction are a defining feature of individuals with autism (Fletcher-Watson et al., 2014; Mazza et al., 2017), the SC assessment in ASD individuals, including psychometric evaluation of commonly used SC tasks, might help clinicians collect additional information and plan the best treatment in ASD research (National Advisory Mental Health, 2016; Morrison et al., 2019).

In ASD research, the SC construct is widely investigated, but it is rarely considered in the clinical practice due to a lack of well-validated tests with established psychometric data, as highlighted by Morrison et al. (2019). In contrast, the use of an SC test in ASD services might improve the diagnostic process and be exceedingly useful for prognoses and creating specific rehabilitation treatments for different age groups. Thus, the aim of the present study was to evaluate three SC tests for their clinical utility in aiding autism diagnosis. We compared performance by ASD and typically developing (TD) children on three SC tasks. Specifically, we chose to use the SIPI (Ziv and Sorongon, 2011; Ziv et al., 2014) for evaluation of social information process abilities, the Comic Strip Task (CST, Cornish et al., 2010; Sivaratnam et al., 2012) to assess the ToM sub-components (beliefs, emotions and intentions), and the children's version of the Eyes Task to evaluate

the ability to understand and infer mental and emotional states regardless of the child's language level. For each test, we calculated threshold scores that best differentiated the two groups using the receiver operating characteristic (ROC) curve.

## MATERIALS AND METHODS

### Participants

One hundred-fifty-four children participated in this study: 86 children with ASD (75 males and 11 females, from 4 to 10 years old, recruited by the Reference Regional Centre for Autism in L'Aquila in the Abruzzo Region, Italy) and 68 TD children (60 males and 8 females, from 4 to 10 years old). The TD children were recruited from a nursery (for 4- to 5-year-old children) and a primary school (for 6- to 10-year-old children) located in L'Aquila. We chose to match the two groups by verbal mental age (VMA), as assessed by the Test for Reception of Grammar (TROG-2; Bishop, 2003). Differences between the two groups emerged for chronological age, where ASD children (mean = 7.64 years, SD = 1.53) were older than TD children [mean = 6.62 years, SD = 1.79; $t(152) = 3.81$, $p < 0.001$] but did not differ in VMA [ASD: mean = 6.96 years, SD = 2.35; TD: mean = 7.52 years, SD = 2.47; $t(152) = 1.43$, $p = 0.15$]. The exclusion criterion was intellectual disability; the participants had an IQ > 80.

The ASD sample comprised children who came for a first-time diagnosis as well as those who came for a second evaluation. All previously diagnosed ASD children received special education through a support teacher. They also followed therapies provided by the National Health System: speech therapy, psychomotor intervention, and Applied Behavioral Analysis.

The clinical process for ASD diagnosis commences with an experienced neuropsychiatrist who observes the child and interviews caregivers. Thereafter, an experienced psychologist performs the Autism Diagnostic Observation Schedule–Second Edition (ADOS-2; Lord et al., 2012). Finally, they consult with one another to make the ASD diagnosis according to the DSM-5 (American Psychiatric Association, 2013) criteria and ADOS-2 outcomes. Clinicians directly involved in the clinical practice participated in the study. ASD participants were level 1, according to DSM-5 criteria (American Psychiatric Association, 2013): most of them showed a delayed language development. ADOS-2 comparison scores of our sample ranged from low to moderate autism-related symptoms. None of the participants had comorbidities with other disorders. All the children were native Italian speakers.

### Procedure

This study was performed in accordance with the recommendations of the Ethics Committee of Local Health Unit 1 (ASL1-Avezzano, Sulmona, L'Aquila), Abruzzo Region, L'Aquila, Italy. The Ethics Committee approved the protocol (number 186061/17) prior to the recruitment of participants, according to the principles established by the Declaration of Helsinki. Informed consent from the child and her or his parents was obtained before participation. Children with ASD were tested at the Reference Regional Centre for Autism, Abruzzo Region Health System, L'Aquila, Italy, whereas TD children were tested in their nurseries or schools. All children were tested individually by an expert psychologist in a quiet room according to the principles established by the Declaration of Helsinki.

### VMA Measure

According to recent literature (Pino et al., 2017), children with ASD show a delay in developing SC abilities based on chronological age, whereas VMA seems to be a good predictor of ToM abilities (Happé, 1994; Pino et al., 2017). Moreover, social difficulty does not appear to be based on the general IQ level, whereas VMA appears to be a more promising associated measure (Pino et al., 2017, 2018).

The literature suggests that children with ASD can use verbal strategies to support their reasoning during ToM tasks (Happé, 1995; Durrleman et al., 2019). Grammatical skills are particularly important during mentalizing (Fisher et al., 2005; de Villiers, 2007; Milligan et al., 2007). For these reasons, we chose to match two groups based on VMA, as assessed with the TROG-2 (Bishop, 2003), a standardized measure of receptive language that allows one to evaluate the ability to understand verbal language. The TROG-2 evaluates the comprehension of grammatical structures and contrasts grammatical indicated by suffixed, functional words, and order word. The test examines 20 syntactic constructions, each of which is examined with a block of four items. Participants select the picture–out of four presented choices–that corresponds to the sentence read by examiner. Standard and age-equivalent scores are made by the total number of blocks passed.

### SC Measures
#### SIPI

The SIPI (Ziv and Sorongon, 2011; Ziv et al., 2014) is a 20-min structured interview based on a storybook-easel that depicts a series of vignettes in which a protagonist is either rejected by two other peers or provoked by another peer. Each type of vignette is combined with each type of peer intent to generate four stories: (1) a non-hostile peer-entry rejection story, (2) an ambiguous peer-entry rejection story, (3) an accidental provocation story, and (4) an ambiguous provocation story. According to Ziv et al. (2014), the scores correspond to four of the five mental steps of social information-processing proposed by Crick and Dodge's (1994) model: (1) encoding, (2) interpretation of cues, (3) response construction, and (4) response evaluation.

An example of a SIPI story is the following: Michael is watching the other children playing. Michael walks up to the other children and asks them: "Can I play with you?" The child says: "Sorry. The teacher said only two can play in the block area" (for details, see Ziv et al., 2014).

The Encoding component evaluates the level of detail that the child recalls across the four stories. Thus, the examiner asks the child: "Tell me what happened in the story, from the beginning to the end." A code of 0 is given to children who recall no correct details from the stories and a code of 1 to children who correctly recall all the details in all the stories. An overall score is then calculated (ranging from 0 to 4).

The Interpretation component evaluates hostile attribution to others' behavior (the question is: "Do you think the other children who didn't let Michael play are mean or not mean?"). The answers are coded with 0 or 1, and an overall score (0–4) is then calculated, with higher scores representing higher levels of hostile attribution bias. Scores for this component are inversely encoded compared with the other SIPI components; that is, a higher score indicates a major tendency to consider the behavior of other children as hostile.

The Response Generation score is derived from the child's responses to the open-ended question: "Pretend that you ask your friends if you can play with them and they say that only two can play in the block area. What would you do?" For each story, the examiner encodes the response as competent or non-competent and assigns a code of 1 if the child's response is classified as competent and of 0 if the answer is classified as non-competent. An overall score (from 0 to 8) is then calculated.

The Response Evaluation items examine the way in which the child assesses the behavior of other people as being right or wrong. This score is obtained by combining the 36 response evaluation questions (4 stories × 3 presented responses × 3 questions per presented response). The three response variables for these steps are: (1) a competent response (e.g., Michael could say, "Then can I play next?"); (2) an aggressive response (e.g., Michael could kick apart the blocks and say to the other children, "If I can't play, then you can't play either!"); and (3) an avoidant or inappropriate response (e.g., Michael could cry and say, "It's not fair"; Pino et al., 2018). The total number of non-competent responses (aggressive and avoidant responses) are subtracted from the total number of competent responses and adjusted for negative scores in order to obtain a score (from 0 to 36).

For the purpose of this study, we also calculated a total score. In our analysis, we did not include the Encoding subscale because one item showed poor psychometric properties (Ziv and Sorongon, 2011). Instead, we used the three main SIPI scores as reported by Ziv and Sorongon (2011): Interpretation, Response Generation, and Response Evaluation. A higher score on the Interpretation subscale (range 0–4) represents hostile attribution. Therefore, we first converted this scale into a non-hostile attribution scale (called Positive Interpretation) by calculating its complementary scale using the following formula: 4 — the number of hostile responses. Next, we summed the Positive Interpretation, Response Generation, and Response Evaluation scores to obtain a total SIPI score.

We decided to use the SIPI because it can evaluate the social cue processing that is closely related to the ability to understand and recognize the intentions, beliefs, and emotions of other people (ToM). According to Mazza et al. (2017), if a child has difficulties in processing social cues within a context, she or he will show difficulties in the ability to evaluate whether other people's social behavior is right or wrong and she or he will respond inadequately in social situations. This phenomenon will impair social relations with others. The test is coded by considering different aspects of the social information process, including the hostile style of attribution and the generation of

socially competent, avoidant, or hostile responses. This factor represents an added value in the diagnostic evaluation; in fact, during the assessment, some behavioral problems may arise that should be considered for future intervention or evaluation. Indeed, Ziv and Sorongon (2011) demonstrated that preschoolers with aggressive tendencies evaluate aggressive responses as better ones. However, future research should deepen this aspect in the clinical setting for details see **Supplementary Material**.

## CST

The CST (Cornish et al., 2010; Sivaratnam et al., 2012) is a 21-item measure that was developed to assess three aspects of ToM: understanding Beliefs, Intentions, and Emotions. There are five items in each component, and each comprises a five-picture comic strip that illustrates everyday social scenarios involving interpersonal interactions that are familiar to young children. Each component has a maximum score of 5, with a total test score range of 0–15 (higher scores correspond to better ToM). We used the CST because it does not require verbal abilities, a factor that allows one to measure ToM deficits *per se.* Moreover, the CST is suitable for a wide swath of the ASD population; it was designed for 4- to 8-year-old children, but it can be used in both younger and older children (Philpott et al., 2013). We also suppose that the use of comics might attract the attention of children, and the formal administration is very brief (10–15 min; Sivaratnam et al., 2012).

## Eyes Task–Children's Version

The Eyes Task (Franco et al., 2014) consists of a series of black and white photos of children's eyes; they portray either mental states or primary emotions. The expressions selected as primary emotions were happy and surprised (positive/neutral valence) and sad and angry (negative valence), while excited and thinking (positive/neutral valence) and worried and shy (negative valence) were selected to represent mental states (for further details, see Franco et al., 2014; Pino et al., 2017). A total of 56 images are presented to the child; each represents one of the stimuli described above with two possible responses. If the child responds correctly, the item is coded as 1; otherwise, it is coded as 0. A total score is then calculated by adding the correct responses to the primary emotions and mental states. Total scores range from 0 to 56 (with higher scores indicating better ToM performance). We used the version by Franco et al. (2014) because stimuli are derived from naturalistic pictures of children rather than posed adults like the version of Baron-Cohen et al. (2001). Moreover, the Eyes Task (Franco et al., 2014) requires fewer cognitive demands because it shows one eye picture with two possible responses. This design is suitable even for low-functioning autism. Score calculations for each test are shown in **Table 1**.

## Data Analysis
### Descriptive Analysis
Demographic parameters and total scores for the SIPI, the CST, and the Eyes Task were recorded for both groups (ASD and TD).

**TABLE 1 |** Score construction.

| | Score Construct | Count (#) | Range |
|---|---|---|---|
| *SIPI* | **Interpretation** | | |
| | $Score_{I-}$ (Negative interpretation) | #(Hostile responses) | 0–4 |
| | $Score_{I+}$ (Positive interpretation)* | 4-#(Hostile responses) | 0–4 |
| | **Response generation** | | |
| | $Score_G$ | #(Competent responses) + [4-#(Non-competent responses)] | 0–8 |
| | **Response evaluation** | | |
| | $Score_E$ | #(Competent responses) + [24-#(Non-competent responses)] | 0–36 |
| | **Total SIPI Score** | | |
| | $Score_{SIPI}$ | $Score_{I+} + Score_G + Score_E$ | 0–48 |
| *Comic Strip Task* | **Intention** | | |
| | $Score_i$ | #(Correct responses) | 0–5 |
| | **Beliefs** | | |
| | $Score_b$ | #(Correct responses) | 0–5 |
| | **Emotions** | | |
| | $Score_e$ | #(Correct responses) | 0–5 |
| | **Total CST Score** | | |
| | $Score_{CST}$ | $Score_i + Score_b + Score_e$ | 0–15 |
| *Eyes Task* | **Primary emotions** | | |
| | $Score_P$ | #(Correct responses) | 0–28 |
| | **Mental states** | | |
| | $Score_M$ | #(Correct responses) | 0–28 |
| | **Total Eyes Task Score** | | |
| | $Score_{ET}$ | $Score_P + Score_M$ | 0–56 |

*This score is used to calculate the total SIPI score.*

## Reliability and Internal Consistency

We assessed the internal consistency and reliability, in relation to the overall measure, for each ToM measure (the SIPI, the CST and the Eyes Task) using Cronbach's α.

## ROC Analysis

The overall goal of the ROC analysis was to estimate the cutoff points for the ToM measures that could distinguish between the two groups. ROC analysis is used to assess the diagnostic properties of tests, specifically, to assess the way in which various measures generally discriminate between categories of subjects. In order to do this, a cutoff point must be established. Based on the cutoff point, we can determine whether a person with a certain score belongs to one category or another (e.g., normal/non-clinical or clinical group). ROC analysis can also be used when comparing the diagnostic performance of two or more tests (Westin, 2001).

In a ROC curve, the true-positive rate (sensitivity) is plotted as a function of the false-positive rate (100 - specificity) for various cutoff points. The obtained area under the curve (AUC) signifies how well a parameter distinguishes between two groups. In order to establish a diagnostic threshold and corresponding test sensitivity and specificity, we established the cutoff as the

value where the highest percentage of true positives was correctly classified as positive and true negatives was correctly classified as negative (Cleves, 1999). In our study, ROC curve analysis was performed to evaluate the accuracy of the total score of ToM measures (the SIPI, CST, and children's version of the Eyes Task) in discriminating between ASD and TD children, using ADOS-2 and DSM-5 criteria as the *gold standard*. The analysis was performed using STATA version 14 statistical software (StataCorp, 2015).

## Optimizing Diagnostic Performance

To improve diagnostic performance, we constructed a test based on a linear combination of the SIPI, CST, and Eyes Task scores. A multivariate logistic regression was performed to obtain the respective logit scores. The logit model allowed us to assess the marginal diagnostic advantage of the SIPI, CST, and Eyes Task and test their statistical significance. Their marginal diagnostic gain can be viewed in terms of the AUC of the ROC curve of the new logit score.

# RESULTS

## Descriptive Analysis

Compared with TD children, children with ASD scored significantly lower on the SIPI [$t(152) = 9.19$, $p < 0.001$] and the CST [$t(152) = 5.59$, $p < 0.001$], but they recorded similar scores on the Eyes Task [$t(152) = 0.43$, $p = 0.66$]. The results are shown in **Table 2**.

## Internal Consistency Results

The results for the CST demonstrated high internal consistency (Cronbach's α = 0.80), the results for SIPI demonstrated good internal consistency (Cronbach's α = 0.76), and the results for Eyes Task demonstrated high internal consistency (Cronbach's α = 0.80).

## ROC Analysis

For the SIPI, the overall $AUC_{SIPI}$ was 0.87 ($SE = 0.02$). The optimal cutoff value was 31 (correctly classified = 79.3%), which corresponded to a sensitivity of 73.5% and a specificity of 83.9%. For the CST, the overall $AUC_{CST}$ was 0.75 ($SE = 0.03$), and the optimal cutoff value was 11 (correctly classified = 71.1%). This value corresponded to a sensitivity of 63.0% and a specificity of 77.0%. For the ET, the overall $AUC_{ET}$ was 0.51 ($SE = 0.04$). The optimal cutoff value was 45 (correctly classified = 50.3); this value corresponded to a sensitivity of 63.24% and a specificity of 40.2%. The analysis revealed a significant difference between AUC measures ($\chi^2 = 60.9$, $p < 0.001$). The results are reported in **Table 3**, and the ROC curves are displayed in **Figure 1**.

## Diagnostic Performance Optimization

The logistic model showed that the SIPI (β = 0.26, $SE = 0.04$, $z = 5.23$, $p < 0.001$) and Eyes Task (β = -0.10, $SE = 0.03$, $z = -3.12$, $p < 0.001$) were statistically significant diagnostic predictors, while the CST (β = 0.18, $SE = 0.11$, $z = 1.66$, $p = 0.09$) was not. When merging the two tests into one new test (hereafter referred

**TABLE 2 |** Between-group differences for demographic data, clinical information, and social cognition measures.

| | ASD (N = 86) Mean (SD) | TD (N = 68) Mean (SD) | t (df = 152) | p |
|---|---|---|---|---|
| Chronological age (in years) | 7.64 (1.53) | 6.62 (1.79) | 3.81 | <0.001* |
| Verbal mental age (in years) | 6.96 (2.35) | 7.52 (2.47) | 1.43 | 0.15 |
| ADOS-Social communication and social interaction | 8.34 (3.50) | – | – | – |
| ADOS-Repetitive and stereotyped behaviors | 1.26 (1.12) | – | – | – |
| ADOS total scores | 9.78 (3.62) | – | – | – |
| *Social cognition measures (total score)* | | | | |
| SIPI | 22.3 (9.22) | 34.3 (6.13) | 9.19 | <0.001* |
| CST | 9.01 (2.48) | 11.2 (2.22) | 5.59 | <0.001* |
| Eyes Task | 44.3 (8.01) | 43.6 (12.0) | 0.43 | 0.66 |

*Significant difference for p < 0.05.*

**TABLE 3 |** ToM measures' AUCs and cut-offs with respective sensitivity and specificity.

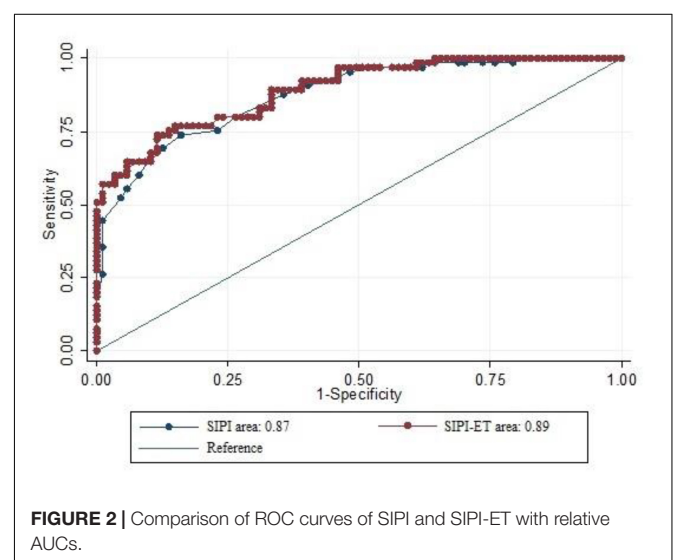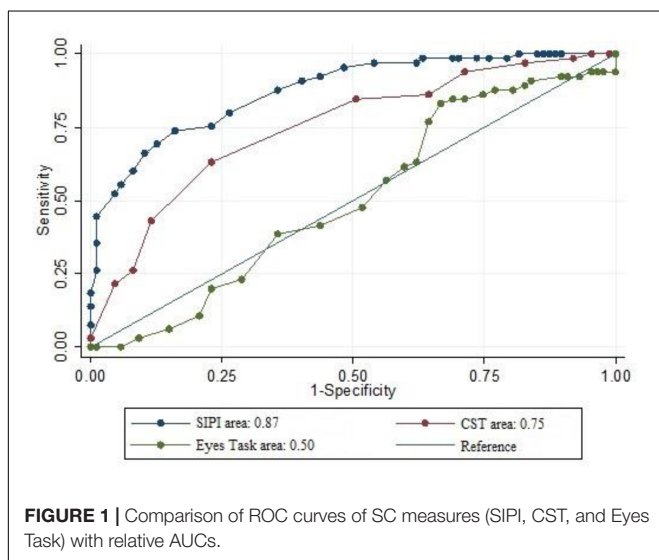| Social cognition measures | AUC* | SE | 95% CI | Cutoff | Sensitivity(%) | Specificity (%) | Correctly Classified (%) |
|---|---|---|---|---|---|---|---|
| SIPI | 0.87 | 0.02 | 0.81–0.92 | 31 | 73.5 | 83.9 | 79.4 |
| CST | 0.75 | 0.03 | 0.67–0.82 | 11 | 63.1 | 77.0 | 71.1 |
| Eyes Task | 0.51 | 0.04 | 0.42–0.60 | 45 | 63.2 | 40.2 | 50.3 |

*Comparison between AUC show a significant difference ($\chi^2$ = 60.9, p < 0.001).*

to as SIPI-ET), we observed an improvement in overall diagnostic performance ($AUC_{SIPI-ET}$ = 0.89, $SE$ = 0.02). However, there was no statistically significant difference between $AUC_{SIPI-ET}$ and $AUC_{SIPI}$ ($\chi^2$ = 2.39, $p$ = 0.12), a finding that indicates that there was no statistically significant improvement. **Figure 2** shows $AUC_{SIPI}$ versus $AUC_{SIPI-ET}$.

## DISCUSSION

This study highlights the utility of including a SC battery of tests to improve the quality and quantity of information collected during procedures for diagnosing ASD. According to Lai et al. (2014), social difficulties in children with autism have

been reported since 1985, when it was first highlighted by Baron-Cohen and collaborators. This impaired ability is believed to play a central role in the social communication and interaction deficits (the first diagnostic criterion in DSM-5; American Psychiatric Association, 2013) of ASD individuals. In fact, this criterion requests clinicians to evaluate abilities as "reduced sharing of interest, emotion or affect" (criterion A1/DSM-5; American Psychiatric Association, 2013); "deficits in social–emotional reciprocity" (criterion A1/DSM-5; American Psychiatric Association, 2013), "deficits in non-verbal communicative behaviors used for social interaction" (criterion A2/DSM-5; American Psychiatric Association, 2013), and "deficits in developing, maintaining, and understanding relationships" (criterion A3/DSM-5; American Psychiatric Association, 2013).



**FIGURE 1 |** Comparison of ROC curves of SC measures (SIPI, CST, and Eyes Task) with relative AUCs.



**FIGURE 2 |** Comparison of ROC curves of SIPI and SIPI-ET with relative AUCs.

All of these competences are part of the complex cognitive construct of SC. Despite the significant role exerted by SC components, such as ToM, in ASD diagnoses, assessment of these competences is neglected in Italian clinical services. Indeed, the use of ToM tests is limited to the research field. For this reason, we evaluated the accuracy of SC measures–using an ROC curve–to discriminate ASD from TD children in a small Italian sample. Additionally, we determined the best cutoff point for the three SC measures used: the SIPI, CST, and Eyes Task.

The results of the ROC analysis showed that the SIPI had good predictive power in terms of accurately classifying children with ASD. On the other hand, the CST showed moderate predictive power, while the Eyes Task showed no ability to correctly distinguish between ASD and TD.

Regarding the Eyes Task, Franco et al. (2014) found that ASD were less accurate compared to TD children, but based on our results, the difference between the groups would not allow us to characterize the ASD individuals during the diagnostic process. In fact, ASD children around 5–6 years old can recognize simple emotional and mental states (i.e., happy, sad, angry, and worried). Thus, there were no distinguishing characteristics in their performance compared to their TD peers. For the CST, the original authors administered the test to 4- to 8-year-old children with high functioning ASD (Sivaratnam et al., 2012). They performed significantly worse compared to controls on the overall two-subscale CST (belief- and intention-understanding). There were no group differences in the emotion understanding subscale performance (Cornish et al., 2010; Sivaratnam et al., 2012). In our study, unlike the authors of CST, we matched subjects by VMA. This method reduced differences in mentalizing ability due to delayed development based on chronological age. Additionally, the participants in our research presented a wider age range compared to Sivaratnam et al. (2012). On the basis of these results, the SIPI represents a useful instrument to support the ASD diagnosis. Specifically, the SIPI assesses the ability to correctly interpret the presented social scenarios (interpretation), "put oneself in another's shoes" (response generation), and determine whether other people's social behaviors are right or wrong (response evaluation).

Our results regarding the SIPI are consistent with a previous study that demonstrated differences between ASD and TD children on this task (Pino et al., 2018). Additionally, Mazza et al. (2017) showed that mentalizing ability plays a key role in the development of social abilities, and the lack of ToM competences in children with ASD impairs their competent social behavior (Mazza et al., 2017). Thus, these components are closely related and improved mentalizing ability might also enhance social behavior.

Collection of the data examined in this study should allow clinicians to plan a treatment focused on social abilities to improve the relationship with other people and avoid isolation and the emergence of other clinical symptomatology, such as depression or anxiety disorder. Furthermore, the systematic use of SC measures in clinical evaluations might help monitor improvements related to treatment and therapy.

In conclusion, we think that the data provided in this study are valuable because they emphasize the utility of incorporating SC measures into diagnostic processes in ASD clinical practice. In particular, the SIPI showed valid accuracy in distinguishing between ASD and TD children. These findings indicate that this test can be implemented into the diagnostic procedure. Additionally, the data provided by our work suggest the cutoff points for each of the examined SC tests (**Table 3**); these data should allow examiners to use these tests with normative values.

We are aware that the present study has some limitations. (1) Our two samples differed in chronological age. However, we stress that the development of SC competencies, particularly mentalizing ability, is related more to mental rather than to chronological age (Pino et al., 2018). (2) This study is also limited by the small Italian sample size; future studies are needed to demonstrate the generalizability of our results. (3) Performance would also need to be compared to other clinical conditions to determine whether these tasks adequately discriminate autism from competing diagnoses. Given that other clinical conditions also present with impairments in social performance, it is necessary to investigate the utility of these tasks for selectively aiding an ASD diagnosis.

# DATA AVAILABILITY STATEMENT

The datasets generated during the current study are not publicly available because the data were obtained in the course of mental health care. However, they are available from the corresponding author on reasonable request.

# ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee approved the protocol (number 186061/17) prior to the recruitment of participants, according to the principles established by the Declaration of Helsinki. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

# AUTHOR CONTRIBUTIONS

MM and MP designed the research. RV, MA, and CD collected the data. FM, MP, and MV analyzed the data. All authors contributed to writing the manuscript.

# SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020. 00004/full#supplementary-material

# REFERENCES

Altschuler, M., Sideridis, G., Kala, S., Warshawsky, M., Gilbert, R., Carroll, D., et al. (2018). Measuring individual differences in cognitive, affective, and spontaneous theory of mind among school-aged children with autism spectrum disorder. *J. Autism. Dev. Disord.* 48, 3945–3957. doi: 10.1007/s10803-018-3663-1

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. Virginia: American Psychiatric Association.

Baron-Cohen, S., Bowen, D. C., Holt, R. J., Allison, C., Auyeung, B., Lombardo, M. V., et al. (2015). The "reading the mind in the eyes" test: complete absence of typical sex difference in˜ 400 men and women with autism. *PloS One* 10:e0136521. doi: 10.1371/journal.pone.0136521

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and 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. *J. Child Psychol. Psychiatr. Allied Disciplines* 42, 241–251. doi: 10.1037/xlm0000745

Baron−Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *J. Child Psychol. Psychiatry* 38, 813–822. doi: 10.1111/j.1469-7610.1997.tb01599.x

Batson, C. D. (1987). "Prosocial motivation: Is it ever truly altruistic?," in *Advances in Experimental Social Psychology*, Vol. 20, (Cambridge, MA: Academic Press. ), 65–122. doi: 10.1016/s0065-2601(08)60412-8

Bishop, D. (2003). *The Test for Reception of Grammar (TROG2)—Version 2*. London: The Psychological Corporation.

Bishop-Fitzpatrick, L., Mazefsky, C. A., Eack, S. M., and Minshew, N. J. (2017). Correlates of social functioning in autism spectrum disorder: the role of social cognition. *Res. Autis. Spectr. Disord.* 35, 25–34. doi: 10.1016/j.rasd.2016.11.013

Cleves, M. A. (1999). "sg120: Receiver Operating Characteristic (ROC) analysis," in *Stata Technical Bulletin* 52, 19–33. In *Stata Technical Bulletin Reprints* 9, 212–229. College Station, TX: Stata Press.

Cornish, K., Rinehart, N., Gray, K., and Howlin, P. (2010). *Comic Strip Task*. Melbourne: Monash University Developmental Neuroscience.

Crick, N. R., and Dodge, K. A. (1994). A review and reformulation of social information: processing mechanisms in children's social adjustment. *Psychol. Bull.* 115, 74–101. doi: 10.1093/deafed/enw030

de Villiers, J. (2007). The interface of language and theory of mind. *Lingua* 117, 1858–1878. doi: 10.1016/j.lingua.2006.11.006

Durrleman, S., Burnel, M., De Villiers, J. G., Thommen, E., and Delage, H. (2019). The impact of grammar on mentalizing: a training study including children with autism spectrum disorder and developmental language disorder. *Front. Psychol.* 10:2478. doi: 10.3389/fpsyg.2019.02478

Fisher, N., Happé, F., and Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *J. Child Psychol. Psychiatry* 46, 409–419. doi: 10.1111/j.1469-7610.2004.00371.x

Fletcher−Watson, S., McConnell, F., Manola, E., and McConachie, H. (2014). Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). *Cochrane Database Syst. Rev.* 21, CD008785. doi: 10.1002/14651858.CD008785

Franco, F., Itakura, S., Pomorska, K., Abramowski, A., Nikaido, K., and Dimitriou, D. (2014). Can children with autism read emotions from the eyes? The eyes test revisited. *Re. Dev. Disabil.* 35, 1015–1026. doi: 10.1016/j.ridd.2014.01.037

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 358, 459–473.

Frith, U., and Happé, F. (1994). Autism: beyond "theory of mind". *Cognition* 50, 115–132. doi: 10.1016/0010-0277(94)90024-8

Hadjikhani, N., Joseph, R. M., Snyder, J., Chabris, C. F., Clark, J., Steele, S., et al. (2004). Activation of the fusiform gyrus when individuals with autism spectrum disorder view faces. *Neuroimage* 22, 1141–1150. doi: 10.1016/j.neuroimage.2004.03.025

Happé, F., and Frith, U. (2014). Annual research review: towards a developmental neuroscience of atypical social cognition. *J. Child Psychol. Psychiatry* 3, 553–577. doi: 10.1111/jcpp.12162

Happé, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Dev.* 66, 843–855.

Hoffman, M. L. (1984). "Interaction of affect and cognition in empathy," in *Emotions, Cognition, and Behavior*, Chap. Cambridge, eds C. E. Izard, J. Kagan, and R. B. Zajonc, (Cambridge University Press), 103–131.

Jones, A. P., Happé, F. G., Gilbert, F., Burnett, S., and Viding, E. (2010). Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *J. f Child Psychol. Psychiatry* 51, 1188–1197. doi: 10.1111/j.1469-7610.2010.02280.x

Jones, C. R., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J., Tregay, J., et al. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Res.* 11, 95–109. doi: 10.1002/aur.1873

Krebs, D., and Russell, C. (1981). "Role-taking and altruism: when you put yourself in the shoes of another, will they take you to their owner's aid?," in *Altruism and Helping Behavior: Social, Personality, and Developmental Perspectives*, eds J. P. Rushton, and R. M. Sorrentino (New Jersey: Lawrence Erlbaum), 137–165.

Lai, M. C., Lombardo, M. V., and Baron-Cohen, S. (2014). Autism. *Lancet* 383, 896–910. doi: 10.1016/S0140-6736(13)61539-1

Lerner, M. D., Hutchins, T. L., and Prelock, P. A. (2011). Brief report: preliminary evaluation of the theory of mind inventory and its relationship to measures of social skills. *J. Autism Dev. Disord.* 41, 512–517. doi: 10.1007/s10803-010-1066-z

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., and Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule (ADOS 2): Manual,* 2nd Edn. Los Angeles: CA: Western Psychological Services.

Mazza, M., Lucci, G., Pacitti, F., Pino, M. C., Mariano, M., Casacchia, M., et al. (2010). Could schizophrenic subjects improve their social cognition abilities only with observation and imitation of social situations? *Neuropsychol. Rehabil.* 20, 675–703. doi: 10.1080/09602011.2010.486284

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

Mazza, M., Pino, M. C., Mariano, M., Tempesta, D., Ferrara, M., De Berardis, D., et al. (2014). Affective and cognitive empathy in adolescents with autisms spectrum disorder. *Front. Hum. Neurosci.* 7:791. doi: 10.3389/fnhum.2014.00791

Milligan, K., Astington, J. W., and Dack, L. A. (2007).Language and theory of mind: meta−analysis of the relation between language ability and false−belief understanding. *Child Dev.* 78, 622–646. doi: 10.1111/j.1467-8624.2007

Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., and Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. *Autism Res.* 12, 766–778. doi: 10.1002/aur.2084

National Advisory Mental Health (2016). *Council Workgroup on Tasks and Measures for Research Domain Criteria. Behavioral Assessment Methods for RDoC Constructs*. Bethesda, MA: National Advisory Mental Health.

Pellicano, E., Jeffery, L., Burr, D., and Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Curr. Biol.* 17, 1508–1512. doi: 10.1016/j.cub.2007.07.065

Peterson, C. C., and Slaughter, V. (2009). Theory of mind (ToM) in children with autism or typical development: links between eye-reading and false belief understanding. *Re. Autism Spectr. Disord.* 3, 462–473. doi: 10.1016/j.rasd.2008.09.007

Philpott, A. L., Rinehart, N. J., Gray, K. M., Howlin, P., and Cornish, K. (2013). Understanding of mental states in later childhood: an investigation of theory of mind in autism spectrum disorder and typical development with a novel task. *Int. J Dev. Disabil.* 59, 108–117. doi: 10.1179/2047387713y.0000000015

Pino, M. C., Mariano, M., Peretti, S., D'Amico, S., Masedu, F., Valenti, M., et al. (2018). When do children with autism develop adequate social behaviour? Cross-sectional analysis of developmental trajectories. *Eur. J. Dev. Psychol.* 17, 71–78. doi: 10.1080/17405629.2018.1537876

Pino, M. C., Mazza, M., Mariano, M., Peretti, S., Dimitriou, D., Masedu, F., et al. (2017). Simple mind reading abilities predict complex theory of mind: developmental delay in autism spectrum disorders. *J. Autism Dev. Disord.* 47, 2743–2743. doi: 10.1007/s10803-017-3194-1

Ramachandran, R., Mitchell, P., and Ropar, D. (2010). Recognizing faces based on inferred traits in autism spectrum disorders. *Autism* 14, 605–618. doi: 10.1177/1362361310372777

Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., and Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex* 46, 668–677. doi: 10.1016/j.cortex.2009.04.008

Sivaratnam, C. S., Cornish, K., Gray, K. M., Howlin, P., and Rinehart, N. J. (2012). Brief report: Assessment of the social-emotional profile in children with autism spectrum disorders using a novel comic strip task. *J. Autism Dev. Disord.* 42, 2505–2512. doi: 10.1007/s10803-012-1498-8

StataCorp, (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP. doi: 10.1007/s10803-012-1498-8

Westin, L. (2001). *Receiver Operating Characteristic (ROC) Analysis: Evaluating Discriminance Effects Among Decision Support systems. Technical Report.* Sweden: Umea University.

Ziv, Y., Hadad, B. S., Khateeb, Y., and Terkel-Dawer, R. (2014). Social information processing in preschool children diagnosed with autism spectrum disorder. *J. Autism Dev. Disord.* 44, 846–859. doi: 10.1007/s10803-013-

Ziv, Y., and Sorongon, A. (2011). Social information processing in preschool children: relations to sociodemographic risk and problem behaviour. *J. Exp. Child Psychol.* 3, 412–429. doi: 10.1016/j.jecp.2011.02.009