

## Theory of Mind Test for Children: Content Validity

Tatiana Pontrelli Mecca<sup>1</sup>

Natália Martins Dias<sup>2</sup>

Patrícia Vieira de Oliveira<sup>1</sup>

Monalisa Muniz<sup>3</sup>

<sup>1</sup>Centro Universitário FIEO, Osasco, SP

<sup>2</sup>Universidade Federal de Santa Catarina, Florianópolis, SC

<sup>3</sup>Universidade Federal de São Carlos, São Carlos, SP

### Abstract

In the national context, there is a lack of instruments to evaluate Theory of Mind, especially with studies of their psychometric properties. This study aimed to investigate content validity evidence for the Theory of Mind Test for Children - TMEC. The development steps, analysis by expert judges and verification of the applicability of the instrument in children aged between 4 and 6 years are described. The TMEC was organized into 4 subtests, following the tasks model described in the literature. Next, 5 judges investigated the clarity of the instructions, the registration form and the score, the level of difficulty of the items and the need for changes. Items highlighted by the judges were reformulated, following the criterion of at least 80% agreement. The applicability of the instrument in a group of six preschool children was verified, which indicated the need to reword a few items for better understanding of the applicator and the child.

*Keywords:* Social cognition, psychological assessment, preschool children

### Teste de Teoria da Mente para Crianças: Validade de Conteúdo

#### Resumo

No contexto nacional, há uma escassez de instrumentos para avaliação de Teoria da Mente, especialmente com estudos de suas propriedades psicométricas. O presente estudo teve por objetivo investigar evidências de validade de conteúdo do Teste de Teoria da Mente para Crianças - TMEC. Foram descritas as etapas de elaboração, análise por juízes especialistas e verificação da aplicabilidade do instrumento em crianças entre 4 e 6 anos. O TMEC foi composto por quatro subtestes, seguindo o modelo de tarefas já descritas na literatura. Após, cinco juízes investigaram a clareza das instruções, da forma de registro e pontuação, nível de dificuldade dos itens e necessidade de mudanças. Foram reformulados itens conforme elencado pelos juízes, seguindo o critério de pelo menos 80% de concordância. Foi verificada a aplicabilidade do instrumento em um grupo de seis crianças pré-escolares, que indicou necessidade de reformulações em poucos itens para melhor compreensão do aplicador e da criança.

*Palavras-chave:* cognição social, avaliação psicológica, pré-escolares

### Test de Teoría de la Mente para Niños: Validez de Contenido

#### Resumen

En el contexto nacional existen pocos instrumentos para la evaluación de Teoría de la Mente, especialmente con estudios sobre las propiedades psicométricas. Este trabajo tuvo como objetivo investigar evidencias de validez de contenido del Test de Teoría de la Mente para Niños - TMEC. Se describen las etapas de elaboración, análisis por jueces especialistas y se verifica también la aplicación del instrumento en niños de entre 4 y 6 años. El TMEC se compone de 4 sub-tests, siguiendo el modelo de las tareas descritas en la literatura. Después, cinco jueces investigaron la veracidad de las instrucciones, la forma de registro y puntuación, el nivel de dificultad de los ítems y la necesidad de cambio. Según lo indicado por los jueces, algunos ítems se reformularon, siguiendo el criterio de por lo menos 80% de concordancia. Se verificó la aplicación del instrumento en un grupo de seis niños en edad pre-escolar, indicando la necesidad de reformulación en algunos ítems para la mejor comprensión del aplicador y del niño.

*Palabras-clave:* Cognición social, evaluación psicológica, pre-escolares

Theory of Mind (ToM) refers to the ability to comprehend mental states, such as thoughts, emotions, beliefs, memories, and intentions. It includes the realization that the behaviors of the individual result from these implied mental states, although they are not always compatible with reality and differ between individuals (Baron-Cohen, 1997).

Social interactions require ToM. This is a guide for reciprocal social behavior, from interactions that

require simpler communications to more sophisticated behaviors such as empathy (Gamliel & Yirmiya, 2009). Therefore, ToM can be understood as a system of inferences (Wellman, Cross, & Watson, 2001) that allows the behavior of the other to be both comprehended and predicted (Astington & Barriault, 2001). In healthy individuals, it is associated with gratitude (Freitas et al., 2012) and pro-social motivation (Pavarini & Souza, 2010).

Theory of Mind develops early, especially between 3 and 5 years of age (Wellman et al., 2001). Its development is strongly associated with language (Milligan, Astington, & Dack, 2007) and inhibition (Carlson, Moses, & Breton, 2002). Environmental variables are also associated with ToM, such as the socioeconomic level (Santana & Roazzi, 2006), size of the family and number of siblings (Jenkins & Astington, 1996) and the frequency with which terms referring to mental states (e.g. to think, to feel) are used in the family setting (Dunn et al., 1991).

There are cases of impairment in the development of ToM, as observed in individuals with Autism Spectrum Disorder (ASD; Baron-Cohen, Leslie, & Frith, 1985), Schizophrenia (Brüne, 2005) and Bipolar Disorder (Kerr, Dunbar, & Bentall, 2003), among others. Theory of Mind deficits present in these conditions are associated with impairments in adaptive behavior. In this way, it can be verified that ToM is necessary for the individual to interact socially, making it possible to establish more healthy and adapted relationships. Thus, the relevance of its evaluation is observed not only in identifying deficits, but also in understanding the development of this ability and in planning interventions that contribute to its promotion.

In the majority of studies, the assessment of ToM is made from false-belief tasks (Osório et al., 2011). First developed by Wimmer and Perner (1983), these tasks consist of a story in which the character has a belief that is different to reality (false), with the necessary information presented to the examinee so that he/she can infer the mental state of the character or his action based on his belief. Other tasks have been developed based on this paradigm, as summarized in Wellman and Liu (2004).

Currently, there is a greater understanding of the assessment of ToM from a perspective regarding its development (Osório et al., 2011). A landmark study was carried out by Wellman and Liu (2004) who conducted a meta-analysis to compare the different types of ToM tasks and to verify the performance of children between 3 and 6 years of age in 7 tasks already described in the literature. It was found that 95% of the children responded correctly in the “diverse desires” task, in which they judged that two people had different desires about the same objects. Next, 84% responded correctly in the “diverse beliefs” task, in which they had to judge that two people had different beliefs without knowing which was true or false. In the “knowledge access” task, 73% of the children responded correctly

when they had to judge the knowledge of the other based on what they saw and not on their own knowledge. In addition, 59% of the children responded correctly to the “contents false belief” task, where they had to comprehend that the character had a belief that did not correspond to reality. Similarly, 57% responded correctly to the “explicit false belief” task, in which they had to infer the behavior of the character knowing that the person had a belief that was different from reality. The most difficult tasks were “belief emotion” with 52% of correct responses, in which the children had to judge the character’s emotion considering his false belief, and “real-apparent emotion”, with 32% correct responses, in which the children had to judge the emotion experienced by the character when the character wanted to demonstrate an emotion different from the one he was actually feeling.

It should be noted that there are tasks in which different levels of ability are required. Considering this perspective of the development of ToM, however, with the aim of providing an intervention for children with ASD, Howlin, Baron-Cohen and Hadwin (1999) described 5 distinct levels of attribution of mental state that progress according to the complexity. The first is understanding simple visual perspective, in which the child must understand that different people can see different objects and judge what the other can or cannot see. The second level requires the child to have a complex visual perspective, so as not only to judge what the other sees, but how this is done, that is, how the other can see the same object viewed by the child. At the third level it is necessary to understand the principle that seeing leads to knowledge, that is, people only know what they can see directly or indirectly, similar to the knowledge access task described by Wellman and Liu (2004). On the fourth level, the child must predict the action of the other based on the knowledge of this other, according to the true belief. The fifth and last level refers to false beliefs.

Due to the early development of ToM and the ease that some individuals, in spite of the diagnosis of ASD for example, presented in some tasks, researchers developed complex tasks, presenting situations which involve lying, sarcasm, pretense and gaffes, among others (Baron-Cohen et al., 1999; Corcoran, Mercer, & Frith, 1995; Happé, 1994). Brief stories are told and questions are asked to see whether there is comprehension of the reality and whether the subject attributes mental states to the characters. These are called “second order” tasks, in which the objective is to verify

whether the individual understands that it is possible to have false beliefs about the beliefs of others (Astington, Pelletier, & Homer, 2002).

Several studies have used the tasks analyzed by Wellman and Liu (2004), including national adaptations and versions (Maluf, Penna-Gallo, & Santos, 2011; Pavarini & Souza, 2010). However, a limitation in the studies is related to the number of items used. Wellman and Liu (2004) verified the level of difficulty of the tasks with only one item in each. Wellman et al. (2001) emphasize the need for a battery for the evaluation of ToM, with multiple tasks to reduce the chances of random errors, which is not possible with only a single item.

Considering the importance of measuring ToM, while identifying scarcity and problems in the instruments to evaluate this construct, the Theory of Mind Test for Children (Mecca & Dias, 2015) was created. Thus, the aim of the present work is to present the content validity study performed for this instrument. To date, no studies were found that reported evidence of content validity for the ToM instrument in the national context. In addition, little has been disclosed about the investigation of this form of evidence in validity studies. According to the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014), content validity refers to the representativeness and comprehensiveness of the items to evaluate the domain that the test intends to evaluate.

Evidence of content validity should be present in the development of an instrument, starting with the choice of items and their importance for the domain to be evaluated. An appropriate battery should contemplate items that exploit a broad, comprehensive and representative range of the domain and thus collect evidence about it (Primi, Muniz, & Nunes, 2009; Urbina, 2007). The theoretical basis and analysis of expert judges are examples of strategies for investigating content validity. The use of a theory that supports the development of the items is fundamental in the construction of an instrument. For Pasquali (2010), the items developed should be submitted to judges, who are specialists in the area in which the instrument is included, for theoretical analysis and verification of their relevance.

In view of the above, a test that presents evidence of validity of consistent content has a greater chance of finding other evidence of validity (in relation to other variables, response process, and internal and consequential structure), since there is no way to have

good evidence with poor quality items, which do not adequately represent and cover the construct. Thus, the study investigated evidence of content validity of the Theory of Mind Test for Children (TMEC).

## Method

The method was composed of three distinct steps. In the first, the development of the TMEC is described. The next step included the analysis by judges and, finally, the applicability of the protocol was verified. The aim of this was to investigate whether the children knew the stimuli used in the TMEC and whether the procedures for applying the instrument were adequate for the age group and clear to the child and applicator.

### *Development of the items of the Theory of Mind Test for Children - TMEC*

The TMEC was developed based on the assumptions of tasks already consolidated in the literature (Baron-Cohen et al., 1999; Happé, 1994; Howlin et al., 1999; O'Hare et al., 2009; Wellman & Liu, 2004). The TMEC subtests were developed considering the difficulty levels presented by Howlin et al. (1999) and Wellman and Liu (2004).

The first subtest, "Comprehending Perspective", was developed based on levels 1 and 2 of the mental state attributes proposed by Howlin et al. (1999) and the diverse desires (Wellman & Woolley, 1990) and diverse beliefs tasks (Wellman & Bartsch, 1989; Wellman et al., 1996). The latter two present higher success rates in children from 3 to 6 years of age (Wellman & Liu, 2004). They are considered tasks with lower levels of difficulty and for that reason they were included in the first subtest. The aim of this is to evaluate the child's comprehension of the perspective of the other, that is, to understand that different people can see things differently or have different views about the same object and may have different desires or beliefs, without knowing the true belief. The choice of stimuli for the composition of the items (pictures) was supported by the proportion of correct answers (values greater than 98%) of preschool children in a vocabulary test (Capovilla, Negrão & Damásio, 2011), guaranteeing the children's familiarity with the stimuli.

The subtest was composed of 9 items. Items 1 to 4 are formed by two questions (Example: 1a and 1b). In the first the children respond according to their own view and in the second according to that of the character. From item 5 the child must respond according

to the view, belief or desire of the character. Items 5, 6 and 9 are composed of one question while items 7 and 8 have 3 questions each, with the second sub-item (Example: 7b and 8b) being performed as a clue, when the child does not correctly respond to the preceding questions (7a and 8a). The subtest uses cards with pictures and toys such as dolls, ball, pencil and box, and the score ranges from 0 to 13 points.

The second subtest, "Attribution of Thought", was developed from levels 3 to 5 described in Howlin et al. (1999): comprehension of the knowledge of the other from what he sees, predicting his actions based on this knowledge and comprehending false beliefs. Also included were tasks based on the three intermediate levels described by Wellman and Liu (2004): knowledge access (adapted from Pratt & Bryant, 1990), contents false belief (adapted from Wellman et al., 2001) and explicit false belief (adapted from Wellman & Bartsch, 1989). In general, the tasks consist of stories in which the child must understand the action of the other on the basis of what the other thinks, even if it is a false belief.

"Attribution of thought" consists of 5 items. Item 1 is made up of 4 questions (Example: 1a to 1d), items 2, 4 and 5 of three questions (Example: 2a, 2b and 2c) and item 3 of one question. Despite the difference in the number of questions, each sub-item has a dichotomous score (zero or 1). The subtest uses cards with pictures and toys and the score ranges from 0 to 14 points.

The third subtest, "Attribution of Basic Emotions", differs from the previous subtest only in the type of attribution performed. In this case, the child must attribute emotions according to the belief, desire and reality of the characters, whether or not they have a false belief about a particular situation. The items were developed from the two tasks described by Wellman and Liu (2004) as more difficult: belief-based emotion (Harris et al., 1989) and real-apparent emotion (Harris et al., 1986). Level 5, identifications of belief-based emotions, described by Howlin et al., was also used (1999). This includes four types of situations: true belief and fulfilled desire, true belief and unfulfilled desire, false belief and fulfilled desire, and finally, false belief and unfulfilled desire. In these items a situation is presented (reality), a desire of the character, a belief (true or false) and the fulfillment or not of the desire, so that the child has to judge whether the character was happy or sad. This subtest was composed of 9 items, with 1, 3, 4 and 5 having two sub-items (questions), item 2 having

three sub-items and items 6 to 9 having four sub-items each. Some sub-items are clues if the child responds incorrectly to the previous sub-item, with these being worth 0.5 point. With the exception of item 1 that is presented using concrete materials (toys), the others are presented using short stories illustrated with cards. The score ranges from 0 to 27 points.

Finally, subtest 4, "Theory of Mind from Complex Situations and Emotions", was developed considering that children as young as 6 can perform the classic ToM tests and, therefore, there may be a ceiling effect in subtests 1 to 3. The vignettes that compose subtest 4 were developed based on existing instruments and studies in the area. The Hinting Task (Corcoran et al., 1995), the Faux Pas (Baron-Cohen et al., 1999) and Strange Stories (Happé, 1994) were used as models. No translations or adaptations of these instruments were carried out as they were only used as models for the construction of original tasks.

Analysis of the vignettes of the instruments mentioned showed that in the Hinting Task they refer to the implicit intention detection ability (Corcoran et al., 1995) and in the Faux Pas, to explicit inappropriate behaviors or gaffes (Baron-Cohen et al., 1999). A vignette was produced based on each model. In turn, the Strange Stories vignettes (Happé, 1994) relate to 12 different types of implicit intentions, emotions or behaviors (such as lies, jokes, sarcasm, deception, etc.). O'Hare et al. (2009) evaluated children aged 5 to 12 years with a version of Strange Stories, which allowed the identification of vignettes that were very complex for the younger children of their sample. Five types of vignettes were more adequate for the evaluation of this age group, namely: lies, misunderstandings, inverted emotions, pretense and double-bluffing. From these types, a vignette was created for each of these 5 aspects. In this way, subtest 4 is composed of 7 vignettes.

For the correction, a criterion was chosen that allows the understanding of reality and the understanding of the mental state to be contrasted, according to the model presented by Shah (2003). After each vignette, a question is asked to verify the understanding of reality and then a second question that assesses the comprehension of the mental state. In this way, each item has two questions, each one being scored as a correct response (one point) or an error (zero). No materials are used for the application except the vignettes themselves and the score ranges from 0 to 14 (7 points for the comprehension of reality and 7 points for the comprehension of the mental state).

All items were theoretically based, with the next step being the analysis by judges to investigate the relevance of the item in relation to the construct evaluated. This step, operationalization of the construct, aimed to develop items with comprehensiveness and representativeness of ToM, an important element as highlighted by Pasquali (2010), Primi, Nunes and Muniz (2009), AERA, APA and NCME (2014) and Urbina (2007). However, in agreement with Pasquali (2010), it is necessary for these items to be evaluated by experts.

#### *Analysis of expert judges*

##### *Participants*

The 4 subtests of the TMEC were submitted to the analysis of judges, in order to evaluate the relevance of the items. The judges were selected according to their area of expertise. Participants were four psychologists and one neuropediatrician, characterized as follows: J1: Psychologist. Master's degree in Psychology with emphasis on Psychological Evaluation. Doctoral degree in Medical Psychology. Experience in clinical care for individuals with Psychiatric Disorders, specifically conditions with ToM alterations; J2: Psychologist. Master's degree in Developmental Disorders. Specialist in Applied Behavior Analysis. Experience in the development and adaptation of tasks for the evaluation of social cognition of children with ASD; J3: Psychologist. Master's and Doctoral degrees in Developmental Disorders. Knowledge in the area of Neuroscience and Social Cognition. Develops work in the area of Social Decision Making; J4: Psychologist. Specialization and Doctoral degree in Clinical Psychology. Postdoctoral degree with emphasis on Psychology of Human Development and Neurosciences. Performs research in the area of neuroscience and social cognition, development of ToM and its alterations in Developmental Disorders; J5: Neuropediatrician. Has clinical experience with children with Developmental Disorders. Performs research with an emphasis on the evaluation of individuals with ASD, using development and adaptation tasks for the evaluation of social cognition and ToM. Although 5 judges participated, each TMEC subtest was sent to and analyzed by 3 judges.

##### *Instruments*

- *Theory of Mind Test for Children (TMEC)*. This test is composed of 4 subtests (total of 30 items) that evaluate, respectively: perspective comprehension (9 items), attribution of thought (5 items), attribution of basic

emotions (9 items) and ToM from complex situations and emotions (7 items). The majority of the items are dichotomous (score is zero or one). Some items (subtests 1 and 3) have clues so that participants can get 0.5 points. There is no criterion for interruption. The application is individual and takes approximately 45 minutes.

- *Protocol of evaluation of the subtests by the judges*: form composed of 12 items, of which 8 are Likert-type. Each item refers to a specific valuation criterion to be verified for each item. Each evaluation form contained instructions for the judges, the main theoretical concepts that guided the construction of the items and the references used by the authors. For the TMEC subtests 1 to 3, seven criteria for the evaluation of the instrument were answered according to the scale: 0 - no; 1 - yes, with reservations; 2 - yes. These referred to: 1) relevance for the age group; 2) clarity of the application description; 3) clarity for assigning points; 4) clarity of the instructions for the child; 5) need for clue for application; 6) consistency of task type; and 7) representativeness of the construct. The eighth question, also of the Likert-type, is a scale of 0 to 10 points regarding the degree of difficulty of the item. The protocol included 4 more open questions related to the quantity, need for insertion, exclusion or reformulation of the item, as well as a space for the judge's observations regarding the subtest. For subtest 4, items 4, 5 and 8 were modified to: 4) clarity of the story/vignette; 5) clarity of the questions; and 8) clarity of the content implicit in the vignette.

##### *Procedure*

Each judge received a copy of the instrument and the protocol for evaluation. The items judged to be "adequate", that is, those that received 2 points by at least 80% of the judges were not modified. Items that received less than 80% agreement regarding their adequacy (items where there were reservations - 1 point - or were not adequate - score 0) were modified according to the suggestions of the judges. After changes in the instrument from the suggestions of the experts, another step was carried out to ascertain the clarity and applicability of the instrument, considering the instructions and registration of the answers in the TMEC application protocol.

##### *Applicability of the protocol*

##### *Participants*

In order to verify the applicability of the instrument and the accuracy of the registration of the points,

a female child, six years of age, was evaluated by five professionals trained in the instrument, these being teachers and students of the Graduate Program where the study was conducted. In a second step, 6 children, aged 4 to 6 years, participated, being 3 girls and 3 boys. The children were selected by convenience from a public school. Children who did not present developmental delays or school difficulties, according to teachers and school records, were included.

#### *Instrument*

In this step, the second version of the TMEC was applied, according to the changes and notes made by the judges, as described in the results section. After the evaluation of the judges, the same number of items remained in each subtest. The items were also kept in the same order, there was agreement regarding their increasing difficulty and complexity. According to the results presented, changes were made in the description of some items to make them clearer, in relation to the presentation of the material, instruction and allocation of points.

#### *Procedure*

The applications were carried out after the approval of the study by the Research Ethics Committee (CAAE 41550315.0.0000.5435), and authorization from those responsible for the children, through the signing of the consent form. The first application of the instrument was carried by one researcher and the other four researchers watched the application and independently scored the questions. With the other 6 children the collection occurred in the school, at a time previously scheduled with the directors. Each application was also scored by two or three different researchers, who also noted observations throughout the course of the application for further analysis.

### **Results and Discussion**

Initially, the results obtained through the analysis of the expert judges will be described. Tables 1 through 4 illustrate the agreement of the judges for each assessment criterion in each item of the four TMEC subtests.

According to the analysis of the judges, subtest 1 proved to be relevant for the age group and clear regarding the allocation of points, as well as showing item-ability coherence and good representativeness of the construct. The judges highlighted a lack of clarity in two items, specifically regarding the form of

presentation of the stimuli and position of the examiner in relation to the stimuli and the child. The instructions for the child were adequate, except for item 1, and the need for a clue was highlighted by some (but not by a majority) of the judges in four items. In relation to the level of difficulty (from 0 to 10) of the items, there was greater diversity in the responses, in the sense that different judges evaluated the same items with different degrees of difficulty. However, there was consistency with regard to the increase in complexity from the first to the last item.

Subtest 2 was assessed as relevant for the age group. With the exception of only one item, all the judges agreed on the clarity of the description for the application and for allocating points. In two items there was a need for clarification of the instructions for the child and in one item the need for a clue was highlighted. Regarding item-ability coherence and representativeness, there were reservations in the evaluation of the judges only for item 1. Regarding the level of difficulty of the items, some of the judges judged the first items as easy and the last ones as moderately difficult, while others judged the first as moderate and the last as very difficult. Some also judged the first item as slightly more difficult than the others.

In subtest 3, 100% agreement between the judges was obtained in relation to the relevance of the items for the age group, item-ability coherence and representativeness of the construct in the item. Some reservations were highlighted in the description for the application for item 8, the description for allocating points in items 1, 2 and 6; and in the instruction for the child in items 2, 6 and 7. The need for a clue was highlighted in items 1, 6, 7 and 8. Regarding the degree of difficulty, the majority of the judges indicated the first 5 items as easier than the last 4 items. In general, items 3 to 5 were judged to be easier than item 2. The level of difficulty did not achieve a minimum of 80% agreement of the judges, however, in general, a gradation of the complexity of the items was observed in subtests 1 to 3. Therefore, the sequence of items in the three subtests was maintained according to the version delivered to the judges. In relation to the difficulty of the subtests (using the arithmetic mean of the classification of the level of difficulty of each item in each subtest), the theoretically expected sequence was obtained: Subtest 1 ( $M=3.26$ ); Subtest 2 ( $M=3.5$ ); and Subtest 3 ( $M=4.45$ ).

For subtest 4, the criteria in which greater concordance among the judges was observed for all the items were item-ability coherence and representativeness,

Table 1  
*Agreement of the judges in relation to each evaluation criterion for the TMEC items - Subtest 1*

| Main criteria   | Score | Agreement of the judges in relation to each evaluation criterion for the TMEC items (%) * |      |      |      |      |      |      |      |      |
|---|-------|---|------|------|------|------|------|------|------|------|
|   |       | Subtest 1   |      |      |      |      |      |      |      |      |
|   |       | 1   | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    |
| Relevant for children aged from 4 to 6 years?   | 0     |   |      |      |      |      |      |      |      |      |
|   | 1     |   |      |      |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 100  | 100  | 100  | 100  | 100  |
| Is the application description clear?   | 0     |   |      | 33.4 | 33.4 |      |      |      |      |      |
|   | 1     |   |      |      |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 66.6 | 66.6 | 100  | 100  | 100  | 100  | 100  |
| Is the description for the allocation of points clear?                                | 0     |   |      |      |      |      |      |      |      |      |
|   | 1     |   |      |      |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 100  | 100  | 100  | 100  | 100  |
| Are the instructions for the child clear and concise?                                 | 0     |   |      |      |      |      |      |      |      |      |
|   | 1     | 66.6  |      |      |      |      |      |      |      |      |
|   | 2     | 33.4  | 100  | 100  | 100  | 100  | 100  | 100  | 100  | 100  |
| Is there a need for a clue for the child in the questions scored?                     | 0     | 100   | 100  | 100  | 100  | 100  | 66.6 | 66.6 | 66.6 | 66.6 |
|   | 1     |   |      |      |      |      | 33.4 |      | 33.4 | 33.4 |
|   | 2     |   |      |      |      |      |      | 33.4 |      |      |
| Is there consistency between the type of task and the ability it intends to evaluate? | 0     |   |      |      |      |      |      |      |      |      |
|   | 1     |   |      |      |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 100  | 100  | 100  | 100  | 100  |
| Is the ability to be evaluated well represented in the item?                          | 0     |   |      |      |      |      |      |      |      |      |
|   | 1     |   |      |      |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 100  | 100  | 100  | 100  | 100  |
| On a scale of 0 to 10, how difficult is the item?                                     | M     | 1.67  | 1.67 | 3    | 3    | 3.34 | 3.67 | 4    | 4.67 | 4.34 |

0 - no; 1 - yes, with reservations; 2 - yes.

in which only item 5 received reservations from the judges. Relevance for the age received reservations by the judges for 2 items and clarity of the story for 3 items. The description for application was assessed as adequate by the majority of the judges, however, there were reservations in all the items, as well as the criterion of clarity of the questions for the children. The clarity of the implicit content in the item was adequate for 3

items, with total agreement among the judges, except for reservations in 4 items. Finally, the criterion of clarity for allocating points was the topic with greater variability in the responses, with the first 5 items being evaluated by the majority as adequate; the last 2 receiving reservations and all being evaluated by some of the judges as inadequate, requiring reformulation. For all the subtests, all the judges agreed on the number of

Table 2

*Agreement of the judges in relation to each evaluation criterion for the TMEC items - Subtest 2*

| Main criteria   | Score | Agreement of the judges in relation to each evaluation criterion for the TMEC items (%) * |      |      |      |      |
|---|-------|---|------|------|------|------|
|   |       | Subtest 2   |      |      |      |      |
|   |       | 1   | 2    | 3    | 4    | 5    |
| Relevant for children aged from 4 to 6 years?   | 0     |   |      |      |      |      |
|   | 1     |   |      |      |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 100  |
| Is the application description clear?   | 0     |   |      |      |      |      |
|   | 1     |   | 33.4 |      |      |      |
|   | 2     | 100   | 66.6 | 100  | 100  | 100  |
| Is the description for the allocation of points clear?                                | 0     |   |      |      |      |      |
|   | 1     | 33.4  |      |      |      |      |
|   | 2     | 66.6  | 100  | 100  | 100  | 100  |
| Are the instructions for the child clear and concise?                                 | 0     |   |      |      |      |      |
|   | 1     | 33.4  | 33.4 |      |      |      |
|   | 2     | 66.6  | 66.6 | 100  | 100  | 100  |
| Is there a need for a clue for the child in the questions scored?                     | 0     | 100   |      | 100  | 100  | 100  |
|   | 1     |   | 33.4 |      |      |      |
|   | 2     |   | 66.6 |      |      |      |
| Is there consistency between the type of task and the ability it intends to evaluate? | 0     |   |      |      |      |      |
|   | 1     | 33.4  |      |      |      |      |
|   | 2     | 66.6  | 100  | 100  | 100  | 100  |
| Is the ability to be evaluated well represented in the item?                          | 0     |   |      |      |      |      |
|   | 1     | 33.4  |      |      |      |      |
|   | 2     | 66.6  | 100  | 100  | 100  | 100  |
| On a scale of 0 to 10, how difficult is the item?                                     | M     | 3   | 2.5  | 3.34 | 4.34 | 4.34 |

0 - no; 1 - yes, with reservations; 2 - yes.

items. Thus, there were no suggestions for adding or withdrawing items.

After the analyses of the judges, changes were made to the items that did not present at least 80% agreement of the evaluators in each criterion. Prior to the subtests, general instructions were included, such as materials, structure required for the application, positioning of the examiner in relation to the child, absence of interruption rule, estimated time for the application,

number of repetitions that can be made in each item and one general instruction to be given to the children in relation to the task they will perform. In subtests 1 to 3, where relevant, information was provided on how to present the item to the child (position of stimulus cards or materials), order of presentation of each stimulus and exact moment in the course of instruction of the item in which each stimulus must be presented.



Table 3

*Agreement of the judges in relation to each evaluation criterion for the TMEC items - Subtest 3*

| Main criteria   | Score | Agreement of the judges in relation to each evaluation criterion for the TMEC items (%) * |      |     |      |      |      |      |      |      |
|---|-------|---|------|-----|------|------|------|------|------|------|
|   |       | Subtest 3   |      |     |      |      |      |      |      |      |
|   |       | 1   | 2    | 3   | 4    | 5    | 6    | 7    | 8    | 9    |
| Relevant for children aged from 4 to 6 years?   | 0     |   |      |     |      |      |      |      |      |      |
|   | 1     |   |      |     |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100 | 100  | 100  | 100  | 100  | 100  | 100  |
| Is the application description clear?   | 0     |   |      |     |      |      |      |      |      |      |
|   | 1     |   |      |     |      |      |      |      | 33.4 |      |
|   | 2     | 100   | 100  | 100 | 100  | 100  | 100  | 100  | 66.6 | 100  |
| Is the description for the allocation of points clear?                                | 0     |   |      |     |      |      |      | 33.4 |      |      |
|   | 1     | 66.6  | 33.4 |     |      |      |      |      |      |      |
|   | 2     | 33.4  | 66.6 | 100 | 100  | 100  | 66.6 | 100  | 100  | 100  |
| Are the instructions for the child clear and concise?                                 | 0     |   |      |     |      |      |      |      |      |      |
|   | 1     |   | 33.4 |     |      |      |      | 33.4 | 33.4 |      |
|   | 2     | 100   | 66.6 | 100 | 100  | 100  | 66.6 | 33.4 | 100  | 100  |
| Is there a need for a clue for the child in the questions scored?                     | 0     |   | 100  | 100 | 100  | 100  |      |      |      | 100  |
|   | 1     | 33.4  |      |     |      |      |      | 33.4 | 33.4 | 33.4 |
|   | 2     | 66.6  |      |     |      |      |      | 66.6 | 66.6 | 66.6 |
| Is there consistency between the type of task and the ability it intends to evaluate? | 0     |   |      |     |      |      |      |      |      |      |
|   | 1     |   |      |     |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100 | 100  | 100  | 100  | 100  | 100  | 100  |
| Is the ability to be evaluated well represented in the item?                          | 0     |   |      |     |      |      |      |      |      |      |
|   | 1     |   |      |     |      |      |      |      |      |      |
|   | 2     | 100   | 100  | 100 | 100  | 100  | 100  | 100  | 100  | 100  |
| On a scale of 0 to 10, how difficult is the item?                                     | M     | 3.34  | 5.34 | 3   | 3.34 | 3.34 | 5    | 5.34 | 5.34 | 6    |

0 - no; 1 - yes, with reservations; 2 - yes.

In each item of subtests 1 to 3, to facilitate the understanding of the child and the applicator, the same sequence of instruction presentation was maintained. Thus, in the application protocol, the materials necessary for the item are described first, then the application for the examiner is described, as well as the question that the child will be asked and, finally, a table with the scoring criteria. The types of answers considered correct

and considered errors or when a clue is needed according to the response of the child are also explained in this. The clues were improved.

In the registration protocol spaces were inserted to describe the response of the child, instead of just assigning a score. In the register of subtest 1, reminders were inserted in items 5 and 6, since there is no correct response a priori; the correct response

Table 4

*Agreement of the judges in relation to each evaluation criterion for the TMEC items - Subtest 4*

| Main criteria   | Score | Agreement of the judges in relation to each evaluation criterion for the TMEC items (%) * |      |      |      |      |      |      |
|---|-------|---|------|------|------|------|------|------|
|   |       | Subtest 4   |      |      |      |      |      |      |
|   |       | 1   | 2    | 3    | 4    | 5    | 6    | 7    |
| Relevant for children aged from 4 to 6 years?   | 0     |   |      |      |      |      |      |      |
|   | 1     | 33.4  |      |      |      | 33.4 |      |      |
|   | 2     | 66.6  | 100  | 100  | 100  | 66.6 | 100  | 100  |
| Is the application description clear?   | 0     |   |      |      |      |      |      |      |
|   | 1     | 33.4  | 33.4 | 33.4 | 33.4 | 33.4 | 33.4 | 33.4 |
|   | 2     | 66.6  | 66.6 | 66.6 | 66.6 | 66.6 | 66.6 | 66.6 |
| Is the description for the allocation of points clear?                                | 0     | 33.4  | 33.4 | 33.4 | 33.4 | 33.4 | 33.4 | 33.4 |
|   | 1     |   |      |      |      |      | 33.4 | 33.4 |
|   | 2     | 66.6  | 66.6 | 66.6 | 66.6 | 66.6 | 33.4 | 33.4 |
| Is the story/vignette clear (succinct and easy to understand) for a child?            | 0     |   |      |      |      |      |      |      |
|   | 1     | 33.4  | 33.4 |      |      | 33.4 |      |      |
|   | 2     | 66.6  | 66.6 | 100  | 100  | 66.6 | 100  | 100  |
| Are the questions clear (succinct and easy to understand) for a child?                | 0     |   |      |      |      |      |      |      |
|   | 1     | 33.4  | 33.4 | 33.4 | 33.4 | 33.4 | 66.6 | 33.4 |
|   | 2     | 66.6  | 66.6 | 66.6 | 66.6 | 66.6 | 33.4 | 66.6 |
| Is there consistency between the type of task and the ability it intends to evaluate? | 0     |   |      |      |      |      |      |      |
|   | 1     |   |      |      |      | 33.4 |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 66.6 | 100  | 100  |
| Is the ability to be evaluated well represented in the item?                          | 0     |   |      |      |      |      |      |      |
|   | 1     |   |      |      |      | 33.4 |      |      |
|   | 2     | 100   | 100  | 100  | 100  | 66.6 | 100  | 100  |
| Is the implied content of the item clear in the vignette?                             | 0     |   |      |      |      |      |      |      |
|   | 1     |   | 33.4 | 33.4 | 33.4 | 33.4 |      |      |
|   | 2     | 100   | 66.6 | 66.6 | 66.6 | 66.6 | 100  | 100  |

0 - no; 1 - yes, with reservations; 2 - yes.

will depend on the choice of the stimulus previously made by the child. Reminders were also inserted in the register of items 7 and 8, as when the first part of the item (part A) is responded to correctly, the child does not answer the second part (B) and the examiner must go straight to the last part (C). Part B is performed only when the child responds incorrectly to part A.

In subtest 4, there was a reformulation of vignettes in agreement with the indications of the judges, adjusting them to the age group and their implicit content, as well as making them clearer for the understanding of the children. Also the questions after each vignette were revised and a clearer and more detailed description of the allocation of points was made, tailoring the items to the comments of the judges.

In relation to the step to certify the adequacy of the application and registration protocols (evaluation of one child), there were no disagreements among the evaluators related to the allocation of the scores, so that 100% scored 1 for the correct response, zero for the errors and 0.5 points for the questions where the child presented the correct answer after the clue. However, changes were specifically made in subtests 1 and 2.

The stimulus in item 3 of subtest 1, the picture of which was horizontal, was replaced with a vertical picture. At the time of application it was found that different responses could be correct depending on the angle the child looked at the figure, even if the examiner presented the stimulus in a standardized way. Therefore, a substitution was chosen, so that only one answer could be scored as correct and the other would be considered an error. In item 1 of subtest 2, initially with four questions, it was decided to eliminate the last question of the item. During the application all the evaluators agreed that it was a question in which the answer was very similar to the previous question, and therefore there were two questions with the same answer, one of which had to be excluded.

Finally, in the application of the protocol to the 6 children, with the observations made, there was no need for changes in relation to the items or the stimuli, since the contents did not seem strange to the children. However, changes were made in the standardization of the description of the instructions of the items, since the evaluators made some errors of application, due to lack of clarity in the instructions (some aspects were emphasized, including with visual highlight). The errors committed most during the application included incomplete presentation of the instructions or addition of statements not planned in the application protocol that could hinder the understanding of the child and absence of clues in some items when necessary.

The errors committed may have occurred because of the amount of detail present in the application, as most of the items are composed of stories that are interrupted by questions and clues from the evaluator. Therefore, the statements of the examiner were highlighted in relation to the description of the application, as well as the questions asked of the child. The items that require clues and the types of response the child can present were also highlighted.

These procedures led to the final version of the TMEC, with the same number of items as the initial version, same division into subtests, however, with description and presentation of items and the application and scoring procedure reformulated. Examples of items of each subtest are shown below. Figure 1 shows items from subtests 1 to 3. Table 5 shows an item from subtest 4.

The TMEC was developed considering different levels of ToM. Items considered easier, such as comprehension of perspective and attribution of thought, up to attribution of emotions and comprehension of mental states in complex situations, according to previous literature (Baron-Cohen et al., 1999; Corcoran et al., 1995; Happé, 1994; Howlin et al., 1999; Wellman & Liu, 2004), were included.

The first version of the instrument was submitted to the judges, who aimed to perform the evaluation of the content and relevance of the items. This is an important stage in the development of an instrument, as the judges are experts in the subject and can provide data regarding the representativeness of the items based on an investigation of the comprehensiveness of the domain to be evaluated (Pasquali, 2004; Primi et al., 2009). The items that presented inter-rater agreement of 80% or more were not modified. The others were carefully reviewed.

There were some differences regarding the degree of difficulty of the items of the subtests assigned by

Table 5

*Example of an item from Subtest 4. In the item, the implicit content is 'gaffe'*

---

Subtest 4 - Theory of Mind from complex situations and emotions.

Juliana got a teddy bear from her cousin at Christmas. A few days later, they were playing and Juliana's cousin inadvertently lost the bear. Juliana told her cousin: "Its alright, he was very ugly and I did not even like him. Someone gave it to me for Christmas".

Comprehension of reality question: What did Juliana's cousin give her at Christmas?

Comprehension of mental state question: In the story, was it better for Juliana not to have said that the teddy bear was ugly and that she did not like it? Why?

---

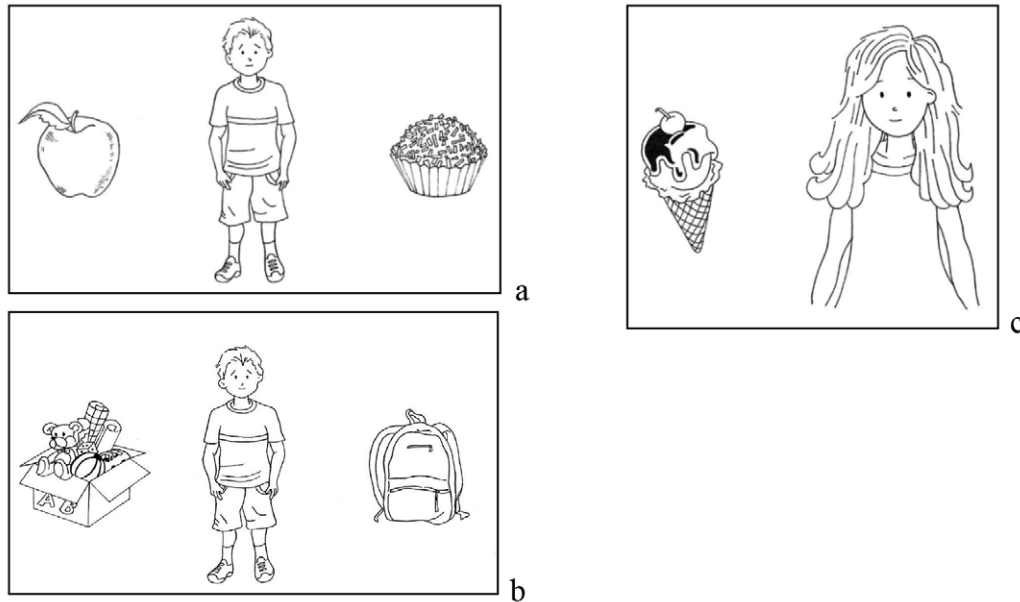


Figure 1. Example of TMEC items: a) Subtest 1 - Comprehension of Perspective (“This is Pedro and its snack time. In the lunch box there is an apple and a chocolate to eat. Which one do you like the most? After the child’s response the examiner should say “good choice, but Pedro likes \_\_\_ (say the name of the noun opposite to what the child likes). Now it’s snack time and Pedro can only choose one to eat. Which one will he choose, the chocolate or the apple?”) b) Subtest 2 - Attribution of Thought (“This is Gabriel. He is looking for his toy car. We know the toy car is in the backpack. But Gabriel thinks the car is in the toy box”. Where do you think Gabriel will look for the car?); c) Subtest 3 - Attribution of basic emotions (“This is Ana (point to the girl). This picture (pointing to the picture of the ice cream) shows what Ana wants. Ana wants ice cream. Ana’s mother bought her an ice cream. What does Ana want? This answer is not scored. If the child makes a mistake, say: “Look, this figure shows what Ana wants.” Provide the correction if the child makes a mistake. How will Ana feel when her mother gives her the ice cream?).

each judge. Some items were judged as easy by some judges and moderately difficult by others. Also, some items were judged as moderate by some of the judges and as difficult by others. These discrepancies may have occurred due to the amplitude of the scale (0 to 10), while in the other criteria the score was more restricted (0 to 2). In this sense, future studies with the TMEC that investigate the level of difficulty of the items through the Item Response Theory are suggested.

However, in general, the majority of the judges agreed that subtest 1 was easier, followed by subtest 2 and with subtest 3 being more complex. The progression according to the complexity of the items in the same subtest and between the subtests corroborates previous findings in the literature. These suggest a scaling of the ToM tasks, beginning with tasks in which the child must judge that two people have distinct desires or perspectives about the same object, passing through different beliefs, understanding the action of the other based on the knowledge of this other, comprehension of explicit and implicit false-belief, up to

the understanding of emotions of others based on their beliefs (Howlin et al., 1999; Wellman & Liu, 2004). Subtest 4 was developed based on a different assumption (Baron-Cohen et al., 1999; Corcoran et al., 1995; Happé, 1994), with a different format. As a result, its complexity was not evaluated based on the same metrics as the previous subtests.

After the review of the instrument from the analysis of the judges, the subsequent applications allowed difficulties to be identified and new adjustments to be made. In general, this first study resulted in a relatively broad and complete instrument, covering different levels of the construct, for the evaluation of children, a tool that, to date, is not available in the national context, despite the work of research groups with adaptations of tasks already used in other countries (Maluf, Penna-Gallo, & Santos, 2011; Pavarini & Souza, 2010). The data found in this study indicate evidence of validity for the content of the TMEC and demonstrate the importance of this type of evidence to produce a more reliable pilot instrument for the continuity of evidence

of validity, in this case, the empirical evidence, as highlighted by Pasquali (2010).

### Final considerations

The present study aimed to verify evidence of validity for the content of the TMEC, an instrument that evaluates ToM in children from 4 to 6 years of age, since there are no instruments for this in the national context. It should be noted that studies carried out in the country use some tasks already consolidated in the literature, developed in other contexts. The development of the TMEC and the subsequent studies of its psychometric characteristics may be important for the area considering the absence of instruments for the evaluation of ToM in Brazil and the relevance of this type of measurement in some contexts, such as ASD, as well as the role of ToM in the understanding of pro-social motivation. Furthermore, considering the important development of this ability in the stage between 3 and 5 years of age, it is fundamental to have instruments that assess ToM in the pre-school population, allowing the early detection of problems in their development. Studies are underway to verify evidence of validity in relation to other variables and reliability of the TMEC. Future studies should cover samples of atypical development (e.g. ASD) and analysis of items aiming for the future availability of the instrument.

### References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Autor.
- Astington, J. W., & Barriault, T. (2001). Children's theory of mind: How young children come to understand that people have thoughts and feelings. *Infants and Young Children*, 13, 1-12. Retrieved from [http://journals.lww.com/iyjournal/abstract/2001/13030/children\\_s\\_theory\\_of\\_mind\\_\\_how\\_young\\_children\\_come.5.aspx](http://journals.lww.com/iyjournal/abstract/2001/13030/children_s_theory_of_mind__how_young_children_come.5.aspx)
- Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New ideas in Psychology*, 20(2), 131-144. doi:10.1016/S0732-118X(02)00005-3
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37-46. Retrieved from [http://www.autismtruths.org/pdf/3.%20Does%20the%20autistic%20child%20have%20a%20theory%20of%20mind\\_SBC.pdf](http://www.autismtruths.org/pdf/3.%20Does%20the%20autistic%20child%20have%20a%20theory%20of%20mind_SBC.pdf)
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press. doi:10.1016/0010-0277(85)90022-8
- Baron-Cohen, S., O'Riordan, M., Jones, R., Stone, V., & Plaisted, K. (1999). A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome. *Journal of Autism and Developmental Disorders*, 29, 407-418. Retrieved from [http://docs.autismresearchcentre.com/papers/1999\\_BCetal\\_Socialsensitivity.pdf](http://docs.autismresearchcentre.com/papers/1999_BCetal_Socialsensitivity.pdf)
- Brüne, M. (2005). "Theory of mind" in schizophrenia: a review of the literature. *Schizophrenia Bulletin*, 31(1), 21-42. doi: 10.1093/schbul/sbi002
- Capovilla, F. C., Negrão, V. B. & Damásio, M. (2011). *Teste de Vocabulário Auditivo e Teste de Vocabulário Receptivo*. São Paulo: Memnon.
- Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, 11, 73-92. doi: 10.1002/icd.298
- 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. Retrieved from [http://www.schres-journal.com/article/0920-9964\(95\)00024-G/pdf](http://www.schres-journal.com/article/0920-9964(95)00024-G/pdf)
- Dunn, J., Brown, J., Slomkowski, C., Tesla, C. and Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development*, 62, 1352-1366. doi: 10.1111/j.1467-8624.1991.tb01610.x
- Gamliel, I., & Yirmiya, N. (2009). Assessment of social behavior in autism spectrum disorders. In S. Goldstein., J. A. Naglieri., & S. Ozonoff (Orgs.), *Assessment of Autism Spectrum Disorders* (pp. 138-170). New York, NY: Guilford Press.
- Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts

- and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of autism and Developmental disorders*, 24, 129-154. doi: 10.1007/BF02172093.
- Harris, P. L., Donnelly, K., Guz, G. R., & Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Development*, 57, 895 – 909. Retrieved from [http://www.jstor.org/stable/1130366?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/1130366?seq=1#page_scan_tab_contents)
- Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, 3, 379 – 400. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/02699938908412713?journalCode=pcem20>
- Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind-read: A practical guide for teachers and parents*. J. Wiley & Sons.
- Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology*, 32(1), 70. doi: 10.1037/0012-1649.32.1.70
- Kerr, N., Dunbar, R. I., & Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. *Journal of affective disorders*, 73(3), 253-259. doi:10.1016/S0165-0327(02)00008-3
- Freitas, L. B. L., O'Brien, M., Nelson, J. A., & Marcovitch, S. (2012). A compreensão da gratidão e teoria da mente em crianças de 5 anos. *Psicologia: Reflexão e Crítica*, 25(2), 330-338. doi: 10.1590/S0102-79722012000200015
- Mecca, T. P. & Dias, N. M. (2015). *Teste de Teoria da Mente para crianças*. Teste em desenvolvimento.
- Maluf, M. R., Gallo-Penna, E., & Santos, M. J. (2011). Atribuição de estados mentais e compreensão conversacional: estudo com pré-escolares. *Paidéia*, 21(48), 41-50. Retrieved from
- 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. doi: 10.1111/j.1467-8624.2007.01018.x
- O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of autism and developmental disorders*, 39(6), 916-928. doi:10.1007/s10803-009-0699-2.
- Osório, A., Castiajo, P., Ferreira, R., Barbosa, F., & Martins, C. (2011). Metodologias de avaliação do desenvolvimento da cognição social da infância até à idade pré-escolar. *Análise Psicológica*, 29(2), 259-274. Retrieved from
- Pasquali, L. (2004). *Teoria dos testes na psicologia e na educação*. Petrópolis: Vozes.
- Pasquali, L. (2010). *Instrumentação psicológica: fundamentos e práticas*. Porto Alegre: Artmed.
- Pavarini, G., & Souza, D. D. H. (2010). Teoria da mente, empatia e motivação pró-social em crianças pré-escolares. *Psicologia em Estudo*, 15(3), 613-622. Retrieved from <http://www.scielo.br/pdf/pe/v15n3/v15n3a19.pdf>
- Pratt, C. & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel. *Child Development*, 61, 973-982. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.1990.tb02835.x/abstract;jsessionid=55A6EBD0AA5EB858B1A88C7185291990.f01t03>
- Primi, R., Muniz, M., & Nunes, C. H. S. S. (2009). Definições Contemporâneas de Validade de Testes Psicológicos. In: C. S. Hutz. (Org.), *Avanços e polêmicas em avaliação psicológica*. (p. 243-265). São Paulo: Casa do Psicólogo.
- Santana, S. M., & Roazzi, A. (2006). Cognição social em crianças: Descobrimo a influência de crenças falsas e emoções no comportamento humano. *Psicologia: Reflexão e Crítica*, 19(1), 1-8. doi: 10.1590/S0102-79722006000100002
- Urbina, S. (2007). *Fundamentos da Testagem Psicológica*. Porto Alegre: Artmed.
- Wellman, H., & Bartsch, K. (1989). Three-year-olds understand belief: A reply to Perner. *Cognition*, 33, 321 – 326. doi:10.1016/0010-0277(89)90033-4
- Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, 35, 245 – 275. Retrieved from <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/28537/0000335.pdf?sequence=1&isAllowed=y>

- Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought-bubbles and of thoughts. *Child Development*, 67, 768 – 788. doi: 10.1111/j.1467-8624.1996.tb01763.x
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 655-684. doi: 10.1111/1467-8624.00304
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, 75(2), 523-541. doi: 10.1111/j.1467-8624.2004.00691.x
- Wimmer, H., & Perner. J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128. doi:10.1016/0010-0277(83)90004-5

Recebido em: 03/06/2016  
Reformulado em: 31/10/2017  
Aprovado em: 10/11/2017

Sobre as autoras:

**Tatiana Pontrelli Mecca:** PhD in Developmental Disorders by the Universidade Presbiteriana Mackenzie. Professor of the Department of School Psychology at Centro Universitário Fieo – UniFieo  
E-mail: tati.mecca@gmail.com  
ORCID: 0000-0002-2009-6228

**Natália Martins Dias:** PhD in Developmental Disorders by the Universidade Presbiteriana Mackenzie. Professor of the Department of Psychology, Universidade Federal de Santa Catarina – UFSC. CNPq Productivity scholarship.  
E-mail: natalia.m.dias@ufsc.br  
ORCID: 0000-0003-1144-5657

**Patrícia Vieira de Oliveira:** Master of Science in Educational Psychology at Academic Centre FIEO, PhD Candidate on Education and Health at Federal University of São Paulo.  
E-mail: patricia@prvo.com.br  
ORCID: 0000-0001-6606-4734

**Monalisa Muniz:** PhD in Psychology from São Francisco University. Professor of Department of Psychology, Federal University of São Carlos.  
E-mail: monamuniz@gmail.com  
ORCID: 0000-0003-1628-6296

Contato com as autoras:

Profa. Dra. Tatiana Pontrelli Mecca  
Rua Franz Voegeli, n.300, Bloco Prata, 1º Subsolo. Secretaria de Pós-Graduação em Psicologia Educacional  
Osasco-SP, Brasil  
CEP: 06020-190