# Children's Social Understanding Scale-Short Form: Adaptation to Turkish Sample

Müge Ekerim Akbulut, Evren Etel, Deniz Tahiroğlu & Ayşe Bilge Selçuk

Routledge
Taylor & Francis Group

Check for updates

# Children's Social Understanding Scale-Short Form: Adaptation to Turkish Sample

Müge Ekerim Akbulut [a], Evren Etel[b], Deniz Tahiroğlu[c], and Ayşe Bilge Selçuk[d]

aDepartment of Psychology, Istanbul 29 Mayis University; bSchool of Applied Psychology, Griffith University; cDepartment of Psychology, Boğaziçi University; dDepartment of Psychology, Koç University

**ABSTRACT**

*Research Findings:* The Children's Social Understanding Scale (CSUS) assesses young children's theory of mind (ToM). The current study adapted the parent version of the CSUS-short form to Turkish and examined its utility, for the first time, as an assessment tool for teachers. The associations of parent and teacher reports of the CSUS-short form with classical behavioral assessments of ToM were investigated along with such known correlates of ToM as receptive language, executive functions (EF), and social competence. Two hundred and two children ($M_{age}$ = 53.98 months), their parents and 169 teachers participated in the study at Time 1, and one year later at Time 2, 132 parents and 109 teachers completed the CSUS-short form again. Similar to the original version of the scale, both parent and teacher forms yielded one-factor structure, demonstrated high internal consistencies at both measurement points, and were positively correlated with each other across time. Parent and teacher forms were also concurrently and positively associated with behavioral ToM tasks, receptive language, EF, and social competence. *Practice or Policy:* These results pointed to the reliability and validity of the parent and teacher versions of the CSUS-short form as a multi-informant measure for Turkish preschoolers and lay the ground for cross-cultural comparison of children's ToM scores with the other cultural adaptions of the CSUS.

A child's insight into the subjective mental states of others, termed theory of mind (ToM), positively relates to various social outcomes in early childhood, such as prosocial behaviors and peer popularity (Wellman, 2014). The power of ToM in predicting important social outcomes in children's lives has led researchers to concentrate on the age at which children acquire ToM, as well as the factors that predict this acquisition. One of the major concerns at the center of this research endeavor has been the accurate measurement of ToM. Various behavioral tasks have been developed to assess children's ToM but important shortcomings of these tasks have become apparent with their repeated use across studies. These include the impracticality in assessment of diverse ToM concepts, performance demands and problems in criterion validity (also see Hutchins et al., 2012). The Children's Social Understanding Scale (CSUS; Tahiroglu et al., 2014) has been developed as a multi-informant report which could compensate for these disadvantages in behavioral ToM tasks. It also provides a way to cross-examine children's scores through reports from different informants.

The present study aims to adapt the parent version of the CSUS-short form into Turkish, to investigate the consistency in its scores over a year, and to examine its construct validity by exploring links of the CSUS with widely known social and cognitive correlates of ToM in early childhood; namely, executive functions (EF), receptive language and social competence. Furthermore, the current

**CONTACT** Müge Ekerim Akbulut ✉ makbulut@29mayis.edu.tr ▣ Department of Psychology, Elmalıkent Mah, Istanbul 29 Mayis University, Elmalıkent Cad, Umraniye, Istanbul 34764, Turkey

study aims to explore the advantages of the CSUS as a multi-informant measure and examines the psychometric properties of the teacher version of the CSUS-short form by testing its construct validity and reliability over a year.

## Measures of Theory of Mind

ToM encompasses children's ability to infer various subjective mental states of others including desires, emotions, beliefs and intentions (Wellman, 2014), and has been used in the literature (and also in the present paper) as a term to denote children's social understanding (Hughes, 2011). The overwhelming majority of the studies in literature have used behavioral assessments to evaluate ToM performance in children and exclusively focused on false belief understanding as the litmus test of mental state inference. These behavioral assessments of ToM yield reliable and valid results which usually predict critical social behaviors (e.g., prosocial behaviors and peer popularity; Imuta et al., 2016; Slaughter et al., 2015), and are widely used across cultures (Shahaeian et al., 2011; Yagmurlu et al., 2005). However, researchers have also noted the limitations of the behavioral ToM tasks (Hutchins et al., 2012). First, the pass/fail nature of behavioral assessments of ToM prevent researchers from measuring the variance in children's ToM scores (Peterson & Slaughter, 2003). Secondly, the tasks inadvertently limit the content of assessment to a few dimensions of ToM (e.g., false belief), as sometimes it is not time-efficient to measure the many different facets of ToM (e.g., desires, intentions) in one or two sessions with young children (Wellman & Liu, 2004). Thirdly, these tasks pose high performance demands on preschoolers' still-developing executive functions and language skills (Wellman et al., 2001). Finally, these behavioral assessments are sometimes limited in capturing a child's subtle insights into the minds of others, for example, those which are revealed in a child's everyday communications and interactions with others in natural contexts (Astington, 2003).

The abovementioned limitations of behavioral ToM assessments led researchers to develop alternative scale-based measures that could be applied to diverse informants, such as parents and teachers. These measures offer potential to reduce the measurement error associated with single-informant ratings, yield results with high ecological validity, and provide feasible and easy testing for children of different ages. To our knowledge, different scales are commonly used when evaluating mental state understanding (and empathy) in older age groups, such as adolescents and adults (see Interpersonal Reactivity Index, Davis, 1996; Theory of Mind Assessment Scale, Bosco et al., 2016), and younger age groups (only recently have ToM scales been applied; see the Theory of Mind Inventory, Hutchins et al., 2012; and Empathy and Theory of Mind Scale; Wang & Wang, 2015).

Tahiroglu et al. (2014) developed the Children's Social Understanding Scale (CSUS) as an alternative multi-informant measure to assess the individual variation in preschool children's ToM. The CSUS emphasized six mental state concepts and included seven items to measure each of these. This resulted in 42 items in total, tapping into such multiple facets of ToM as diverse desires (understanding differences in peoples' desires), diverse beliefs (being aware that beliefs can change over time), knowledge access (understanding that people differ in their knowledge states), intentions (understanding that people's actions arise from their intentions), emotions (understanding that same situation can cause multiple emotions in different people), and perception (knowing that the perceptual attention of others can be changed and that perception and reality might clash). The items in the scale included questions that directly asked informants if the child is aware of subjective mental states of others (e.g., "understands that telling lies misleads other people"), involved statements asking whether the child uses diverse mental state terms (e.g., "talks about the difference between intentions and outcomes – 'he tried to open the door, but it was locked'") or included situations which pondered real-life behaviors that reflects a child's grasp of diverse minds (e.g., "is good at playing hide and seek"). Parents were asked to rate their children on a 4-point Likert scale ranging from "*definitely untrue of my child*" to "*definitely true of my child*." In addition to the 42-item scale, a short form of the scale was also created from a subset of items. Three items were selected from each mental state concept based on the item-total correlations and correlations with behavioral measures, which resulted in 18-

items in total representing the same six mental states. The main reason for development of the CSUS-short form was to provide researchers with a quick assessment tool, which is of critical importance when ToM skills of children are measured as part of research into other developmental constructs (Tahiroglu et al., 2014).

When applied to typically developing North-American preschoolers and their parents, the CSUS revealed good internal consistency in both its long ($\alpha$ = .94) and short forms ($\alpha$ = .89), and demonstrated significant associations with behavioral ToM assessments (e.g., contents false belief and appearance-reality distinction tasks) as well as children's cognitive skills (e.g., working memory and planning) (Tahiroglu et al., 2014). These results led to the adaptation of CSUS's long-form to different cultural samples such as Canadian-French (Brosseau-Liard & Poulin-Dubois, 2018), Polish (Białecka-Pikul & Stępień-Nycz, 2019; Smogorzewska et al., 2019) and Korean (Jang & Shin, 2018) preschoolers and their parents. Using the long-form of the CSUS, these cultural adaptations showed reliable scores and significant associations with important correlates of ToM in preschoolers, such as attention and inhibition (Brosseau-Liard & Poulin-Dubois, 2018), receptive language (Białecka-Pikul & Stępień-Nycz, 2019), EF (Jang & Shin, 2018) and ability to cope with difficult social situations (e.g., entry to peer group). Of note, the reliability and validity of the adaptations of the CSUS were obtained for both typical and atypical school-aged children (Smogorzewska et al., 2019).

The short-form of the CSUS has not been adapted for use across cultures. However, the opportunity to have quick measurement of ToM with a shorter scale could be advantageous for researchers, as well as for parents and teachers whose motivation to participate in research might increase if the assessment process is brief. Indeed, the utilization of concise surveys is highly recommended, as they lower the risk of distracting participants' attention and provide more accurate results compared to similar yet longer questionnaires (Sheatsley, 1983). This point might be especially crucial when data is obtained from parents with limited time and varying levels of attention.

Adopting a multi-informant approach is also a critical consideration for the accuracy of the assessment. Evaluation of the same construct by multiple informants can generate enriched insight for capturing the different facets of the same construct and decrease source and setting errors found in single-informant approaches (De Los Reyes et al., 2015). Despite being acknowledged as a potential issue by Tahiroglu et al. (2014), previous CSUS assessment tool studies (including their own) have failed to obtain reports from people other than parents (e.g., Brosseau-Liard & Poulin-Dubois, 2018; Smogorzewska et al., 2019). However, teachers in addition to parents, may also provide valuable information about children's socio-cognitive skills, since they have the chance to observe children's social behaviors when they are interacting with peers in a structured environment outside the home (Fuhs & Day, 2011). Consequently, information gathered from teachers can further improve the accuracy of ToM measurement, especially when it converges with reports from the parents.

### The Present Study

The main aim of the current study was to adapt the parent version of the CSUS-short form into Turkish and explore its psychometric properties. Thus, our main research question was whether the CSUS-short form displayed acceptable scores for validity and reliability in this novel context. To achieve this, we examined the convergent validity of the CSUS-short form and investigated its links with children's behavioral ToM performance, as well as with the known correlates of mental state understanding such as EF, receptive language and social competence. We hypothesized that the CSUS-short form would emerge as a valid instrument for assessing the ToM performance of young Turkish children and yield positive associations with children's behavioral ToM scores, EF, receptive language and social competence. We also investigated inter-rater reliability and over-time stability of the CSUS ratings. For this, we examined the link between parent and teacher ratings of the CSUS and calculated the consistency of the CSUS scores across a one-year period. Given that social behaviors of children

reflect their socio-cognitive skills across different settings (Astington, 2003), we hypothesized that parents and teachers would converge in their ratings and the CSUS scores would show consistency over-time.

Contrary to previous studies, which solely used the CSUS as a parent report, the present study adopted a multi-informant approach. Thus, another research question of the current study was to test if the CSUS could be utilized as an instrument for teachers. Given teachers are accepted to be as reliable informants about children's social behaviors as parents (Arseneault et al., 2003), items of the CSUS could capture teachers' rich insights into the mental state understanding skills of children. Therefore, as with the parent version of the CSUS, we investigated the psychometric properties of the teacher version of the CSUS-short form by examining its convergent validity, inter-rater reliability and over-time stability across one year. We hypothesized that similar to the parent form, the teacher form of the CSUS would display significant associations with behavioral ToM, EF, language, and social competence, and yield over-time stability across one year. Finally, we hypothesized that, as with the original CSUS form that was designed to globally assess children's inference of multiple mental state concepts (Tahiroglu et al., 2014), a one-factor structure would emerge from both parent and teacher versions of the CSUS-short form.

In addition to the concurrent links of the CSUS, we examined whether the CSUS scores at T2 demonstrate similar convergent validity to the CSUS scores at T1. Based on the extant literature showing that children's EF and language skills were linked with their both concurrent and future ToM performance (Ruffman, 2014), we hypothesized that T1 EF and T1 receptive language would predict T2 CSUS scores. For social competence, however, the literature supports bidirectional associations, whereby advanced ToM relates to harmonious concurrent and future social relationships through which children encounter diverse minds and further strengthen their mental state understanding skills (Weimer et al., 2021). Hence, we expected that T1 social competence would positively predict T2 CSUS as social interactions expose children to different mental states and facilitate their ToM. In summary, we hypothesized that similar to the T1 CSUS scores, T2 CSUS scores would show convergent validity and would be predicted by earlier EF, receptive language, social competence and behavioral ToM scores.

Parents and teachers filled out the CSUS-short form twice, one year apart at T1 and T2, and also reported on children's social competence scores at T1. Individual assessments were conducted to measure children's EF, receptive language and behavioral ToM performance at T1. We chose knowledge access, contents false belief and unexpected change in location as our target behavioral ToM tasks because they have been commonly used in previous studies with Turkish samples and revealed variability in Turkish children's ToM performance (Etel & Yagmurlu, 2015; Yagmurlu, 2014). These tasks were also utilized in the original study of the CSUS (Tahiroglu et al., 2014) because they tap into diverse aspects of ToM and maximize the chances of capturing individual differences. Consistent with previous adaptations of the CSUS-long form (Brosseau-Liard & Poulin-Dubois, 2018; Smogorzewska et al., 2019), we examined the Turkish version of the CSUS-short form in a wide age group, starting from preschool through to the elementary school period, which is considered a milestone in the development and assessment of ToM (Wellman et al., 2001).

## Methods

### Participants

A total of 202 children (101 girls; age range = 34–80 months, $M_{age}$ = 53.98 months, $SD$ = 10.40), their parents, and 169 teachers took part in the study at T1 (see Table 1 for descriptives).

Parent's mean age was 34.93 years ($SD$ = 5.14). Although literate, 5% of the participant parents had no school experience; 13% had primary school education, 7% had a middle school degree, 30% were high school graduates, and 45% had a university education. At the time of data collection, the monthly household income of 30% of the families was less than 2,000 Turkish liras (TL; ~695 USD), 41% of

**Table 1.** Descriptive statistics.

| Variables | 3-year-olds (n = 59) | | | | 4-year-olds (n = 77) | | | | 5-year-olds (n = 56) | | | | 6-year-olds (n = 10) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M | SD | Min | Max | M | SD | Min | Max | M | SD | Min | Max | M | SD | Min | Max |
| Demographic Variables | | | | | | | | | | | | | | | | |
| Parent education (1–10) | 8.19 | 1.87 | 2.00 | 10.00 | 6.71 | 2.67 | 1.00 | 10.00 | 5.28 | 3.12 | 0.00 | 10.00 | 7.20 | 3.55 | 0.00 | 10.00 |
| Monthly household income (1–7) | 5.31 | 1.20 | 3.00 | 7.00 | 4.68 | 1.7 | 1.00 | 7.00 | 3.48 | 1.48 | 1.00 | 7.00 | 4.10 | 1.52 | 2.00 | 7.00 |
| Study Variables | | | | | | | | | | | | | | | | |
| T1 CSUS parent-report (1–4) | 2.76 | 0.41 | 1.72 | 3.61 | 2.84 | 0.42 | 1.71 | 3.78 | 3.02 | 0.37 | 1.98 | 3.78 | 2.94 | 0.48 | 2.08 | 3.56 |
| T1 CSUS teacher-report (1–4) | 2.83 | 0.76 | 1.17 | 3.82 | 2.96 | 0.52 | 1.67 | 3.83 | 3.23 | 0.45 | 2.39 | 3.94 | 3.16 | 0.55 | 2.22 | 3.83 |
| T2 CSUS parent-report (1–4) | 3.11 | 0.42 | 2.22 | 3.89 | 3.21 | 0.33 | 2.44 | 3.72 | 3.32 | 0.32 | 2.56 | 3.82 | 3.50 | 0.28 | 3.06 | 3.83 |
| T2 CSUS teacher-report (1–4) | 3.10 | 0.54 | 2.00 | 3.94 | 3.17 | 0.46 | 2.17 | 3.83 | 3.37 | 0.47 | 2.33 | 4.00 | 3.08 | 0.30 | 2.83 | 3.41 |
| T1 Behavioral measures of ToM (0–3) | 0.98 | 0.93 | 0.00 | 3.00 | 1.65 | 0.80 | 0.00 | 3.00 | 2.13 | 0.80 | 0.00 | 3.00 | 2.30 | 0.95 | 0.00 | 3.00 |
| T1 EF (z score) | -0.70 | 0.77 | -2.00 | 0.84 | 0.13 | 0.77 | -2.00 | 1.03 | 0.57 | 0.52 | -1.80 | 1.03 | 0.51 | 0.96 | -1.80 | 1.03 |
| T1 Receptive language (z score) | 0.32 | 0.33 | -0.30 | 1.25 | 0.35 | 0.38 | -0.40 | 1.59 | 0.44 | 0.41 | -0.40 | 1.14 | 0.30 | 0.40 | -0.50 | 0.71 |
| T1 Social competence mother-report (1–4) | 2.75 | 0.53 | 1.38 | 3.93 | 2.74 | 0.47 | 1.57 | 3.79 | 2.97 | 0.42 | 1.58 | 3.71 | 3.02 | 0.38 | 2.50 | 3.79 |
| T1 Social competence teacher-report (1–4) | 2.89 | 0.56 | 1.86 | 4.00 | 2.99 | 0.51 | 1.71 | 4.00 | 3.19 | 0.52 | 2.21 | 4.00 | 3.09 | 0.54 | 2.36 | 4.00 |
| T1 Social competence (composite) (1–4) | 2.81 | 0.45 | 1.69 | 3.65 | 2.85 | 0.43 | 1.86 | 3.75 | 3.08 | 0.38 | 1.90 | 3.79 | 3.06 | 0.44 | 2.46 | 3.89 |

Parent education: 1 = having no formal education but being literate; 10 = having graduate school degree. Monthly household income: 1 = less than 630 TL (~219 USD); 7 = higher than 12,000 TL (~4180) USD. T1 = Time 1. T2 = Time 2. CSUS = Children's Social Understanding Scale. ToM = Theory of Mind. EF = Executive Functions.

families were earning between 2,000 and 7,000 TL (~2,430 USD), and 29% of families had a monthly income of more than 7,000 TL (see Table 1). Overall, the demographic information demonstrated that families were from diverse socioeconomic backgrounds that included low, middle and high incomes, and different levels of education.

Of the 202 parents who participated in T1 data collection, 132 parents completed the CSUS-short form at T2 one year later. Children had a mean age of 65.09 months at T2 ($SD$ = 10.56, age range = 46–88 months; 63 girls and 69 boys). One hundred and nine teachers completed the CSUS-short form at T2. Due to the changes in classroom and school allocations and/or children or teachers who changed their schools, most teachers who completed the CSUS at T1 were different than the ones who completed the CSUS at T2. The rest of the parents and teachers declined participation due to lack of time or moving from their previous location. No difference was found between the T1 teacher-reported CSUS scores of those who remained in the study ($M$ = 3.07, $SD$ = 0.66) and those who dropped out ($M$ = 3.04, $SD$ = 0.57), $F$ (1, 162) = 0.09, $p$ = .76, $\eta2$ = 0.001. However, T1 parent-reported CSUS scores were higher for those who remained in the program ($M$ = 2.93, $SD$ = 0.40) than for those who dropped out ($M$ = 2.80, $SD$ = 0.45), $F$ (1, 200) = 4.43, $p$ = .04, $\eta2$ = 0.022.

## Measures

### Children's Social Understanding Scale

The 18-item short form of the Children's Social Understanding Scale (CSUS; Tahiroglu et al., 2014) was used. Bilingual researchers and authors of the current study translated the scale into Turkish, which was then followed by back-translation to English and finally re-translation to Turkish. Some of the names and objects in the exemplar items of the CSUS were adapted into Turkish culture based on their commonality in use (e.g., coffee in the original items was replaced with tea). One major change has been made regarding the wording of the Likert scale. Previous research experience revealed that Turkish parents have an easier time grasping and answering scale questions when they are rated in terms of frequency of the behavior (e.g., ranging from *never* to *always*) rather than in terms of descriptiveness of behavior (e.g., ranging from *definitely untrue* to *definitely true of my child*) (Yagmurlu & Sanson, 2009). Thus, wording of the scale ratings was changed from descriptions of the behavior in the original CSUS to frequency of the behavior in the Turkish version. Parents and teachers were asked to answer each of the 18 questions on a 4-point Likert scale (from 1 = *never* to 4 = *always*), also given a don't know option, which was coded as missing values in the analyses.

Consistent with Tahiroglu et al. (2014)'s original study, cases missing more than 20% of the data (e.g., more than 3 items) were excluded from the analyses.[1] Data for cases having 3 or fewer missing items were imputed through the multiple imputation method. This method creates five sets of plausible values for missing data by incorporating random variation of the existing items and then pools these plausible values to come up with the best estimation . Multiple imputation allows researchers to obtain unbiased estimates of all parameters and their standard errors. It is recommended for data points that are missing at random (Little & Rubin,9). After the imputation, composite scores for the CSUS were obtained by averaging responses across the 18 items.

### Behavioral Measures of Theory of Mind

Knowledge access, contents false belief and unexpected change tasks were used to assess children's behavioral ToM performance.

*Knowledge Access Task.* We used the knowledge access task in the Wellman and Liu's (2004) ToM Scale. Children were first presented with a closed box and asked to guess its content ("What do you think is in this box?"). After giving their initial answer, they were shown that the box actually contained a small dog ("Let's see what is in there. There is a small dog inside!"). As a memory control question, children were asked what was really inside the box. Then, a doll figure was introduced, and

children were told that this doll had never seen the content of the box. As the target question, children were asked if this doll who had never seen the content of the box knew what was inside ("Does this doll know what is inside the box?"). In order to pass this task, children had to correctly answer both the memory control and the target questions. Those who passed both questions received 1 point and those who correctly answered the target question but failed the memory control question, received 0 (see Wellman & Liu, 2004 for similar scoring).

*Contents False Belief Task.* In the contents false belief task (Perner et al., 1987), children were presented with a crayon box that in fact included a few band-aids. At first, children were asked to guess the content of this crayon box ("Here is a box, what do you think is in the box?"). After receiving the children's answer, the experimenter opened the box and showed the actual content, the band-aids, and closed the box. As a memory control question, children were asked to indicate what was really inside the box ("So, what is in the crayon box?"). Following this memory control question, the experimenter introduced a doll to the children and told them that this doll had never seen the content of the box. As the target question, the experimenter asked the child what the doll would think the box contains ("What does this doll think is in the crayon box? Band-aids or pencils?"). Children received a score of 1 if they passed both the memory control and target question and 0 if they failed either one of the questions or both questions (as recommended by Wellman & Liu, 2004).

*Unexpected Change Task.* In this task developed by Wimmer and Perner (1983), children were introduced to two doll figures (one male and one female), two colored boxes (blue box and yellow box) and a ball. The experimenter told children that after playing with the ball, the male doll put it in the blue box and left. Next, the female doll arrived, took the ball out of the blue box, played with it, and put it in the yellow box. As a memory control question, children were asked where the male doll put the ball, and as the target question they were asked where the male doll would look for the ball. Children were given a score of 1 if they answered both memory and control questions correctly, and 0 if they failed either one of these questions or both questions.

Knowledge access, contents false belief and unexpected change tasks were all significantly correlated with each other ($r$'s > . 20, $p$'s < .001). Therefore, they were summed to create a total behavioral ToM score, with 0 indicating the lowest score and 3 indicating the highest score.

### Receptive Language

Children's receptive language was assessed with the Turkish Expressive and Receptive Language Test (TIFALDI; Berument & Güven, 2010), a valid and reliable measurement tool for assessing receptive language skills of preschoolers. In TIFALDI, children were shown sets of four pictures and asked to point to the one that matches what the experimenter says. Receptive language scores were calculated with the three-parameter item response theory. This gives latent language ability scores by accounting for the relative difficulty of each question, the probability of giving a correct response just by guessing, and the discrimination of each item in determining the rate of success on the test (see Ekerim & Selcuk, 2018 for similar computations).

### Executive Functions

Executive functions were assessed based on the complex response inhibition performance of children (Garon et al., 2008) in day-night and peg tapping tasks. These tasks require children to (1) remember a new rule, (2) detect the conflict between dominant and subdominant responses, (3) inhibit the dominant response and (4) perform the subdominant response (see Etel & Yagmurlu, 2015 for same procedure with Turkish children).

*Day-Night Task.* In the modified version of the day-night task (Gerstadt et al., 1994), children were shown one day-card depicting sunshine and daylight, and one night-card that displays moon and the stars. Children were expected to point to the night card when the experimenter said "day" and point to

the day card when the experimenter said "night." After three practice trials, ten consecutive trials were administered. Children received 1 point for each of their correct pointing responses, leading to a minimum score of 0 and a maximum score of 10 in total.

*Peg Tapping Task.* In the first part of the peg tapping task (Diamond & Taylor, 1996), children were asked to tap a wooden peg twice when the experimenter tapped once, and to tap once when the experimenter tapped twice. Then, a new rule was added in the second part and children were asked not to tap at all when the experimenter tapped three times. In each part of the task, children were given practice trials that were followed by 12 consecutive test trials and each of their correct responses in these trials was coded as 1. Overall peg tapping scores ranged from a minimum score of 0 and maximum score of 24.

Due to the high correlation between day-night and peg tapping tasks ($r = .62$, $p < .001$), we created an overall EF score by taking the mean of the standardized ($z$) scores of day-night and peg tapping tasks (see Etel & Yagmurlu, 2015 for the same procedure and scoring with Turkish children).

### Social Competence

Children's social competence was assessed with the Social Competence subscale of the Social Competence and Behavioral Evaluation Scale (SCBE; LaFreniere & Dumas, 1996) and the Play Interaction subscale of the Penn Interactive Play Scale (PIPPS; Fantuzzo et al., 1995). Both measures were reliable and valid assessment tools for Turkish children (see Etel & Yagmurlu, 2015; Korucu et al., 2016). The Social Competence subscale of the SCBE had 10 items (e.g., *works well in groups*) rated on a 4-point Likert scale (1 = *"never"* and 4 = *"always"*) by both parents and teachers. High internal consistency scores were observed in both parent and teacher reports (for parent ratings Cronbach's $\alpha = .74$; for teacher ratings Cronbach's $\alpha = .81$). The Play Interaction subscale of the Penn Interactive Peer Play Scale had 8 items (e.g., *helps settle peer conflicts*) rated on a 4-point Likert scale (1 = *"never"* and 4 = *"always"*) by parents and teachers. High internal consistency scores were observed in both parent and teacher reports (for parent ratings Cronbach's $\alpha = .77$; for teacher ratings Cronbach's $\alpha = .83$).

A composite social competence score was created by calculating the mean of the two subscales (Social Competence and Peer Interaction) separately for parent and teacher reports. However, given the significant correlation between parent and teacher reports of social competence ($r = .40$, $p < .001$), we calculated an aggregate average social competence score for teacher and parent reports.

### Procedure

The study was approved by the XXX University Institutional Review board. Written consent from the parents and the directors of kindergartens/preschools, and verbal assent from the children were obtained. Children were recruited via convenience sampling from kindergartens and preschools in five socio-economically diverse cities of Turkey: Istanbul, Bursa, Balikesir, Tekirdag and Mugla. Individual assessments of receptive language, ToM and EF took place in a separate and quiet room at the children's preschools or homes and lasted for about 45–50 minutes for each child. Children were given stickers as an expression of appreciation. Questionnaires (CSUS, SCBE and PIPPS) were mailed to parents' home addresses and collected by teachers, while teachers completed these questionnaires at preschools (over two weeks). Teachers filled out the questionnaires for multiple children if children were attending the same class. Participants were free to withdraw at any stage. The data of the current study is available from the corresponding author upon reasonable request.

### Data Analysis Plan

We first investigated the psychometric properties of the CSUS-short form by examining its factor structure through confirmatory factor analysis (CFA) which allows for comparing the data against a theoretical model (Hurley et al., 1997). Since the CSUS has been found to have a one-factor structure which assesses the inferences relating to multiple mental states (Tahiroglu et al., 2014), we tested if the CSUS-short form ratings of parents and teachers fit the original one-factor model. We examined the reliability scores of the CSUS by calculating inter-item reliability through Cronbach's alpha and inter-rater reliability, and through correlations between parent and teacher scores. With regard to the convergent validity, we examined the associations of the CSUS-short form with behavioral ToM scores, EF, receptive language and social competence. We examined both zero-order as well as partial correlations controlling for demographics (e.g., age and gender). Finally, given the continuous improvement in ToM with age, we explored the stability of the CSUS scores over a one-year period by calculating correlations between the first and the second year CSUS ratings (T1 and T2). In addition, to confirm convergent validity of the subsequent CSUS scores (T2), we ran hierarchical regression analyses and examined if concurrent correlates of ToM could be linked with T2 CSUS ratings, as found with the T1 CSUS ratings.

## Results

### Confirmatory Factor Analysis

When the factor structure of the T1 CSUS was examined, CFA results revealed acceptable fit indices for the one-factor model in both parent ($RMSEA$ = 0.07, $CFI$ = 0.90, $SRMR$ = 0.06) and teacher forms ($RMSEA$ = .09, $CFI$ = 0.91, $SRMR$ = 0.07). Three items in the parent form and one item in the teacher form had marginal factor loadings ($\lambda < 0.30$, see Table 2). However, given the acceptable fit indices, these items were retained in line with the original structure and previous adaptations of the CSUS (see Smogorzewska et al., 2019 for a similar procedure).

### Reliability and Correlations of the CSUS-Short Form Score with Demographics

The parent (Cronbach's $\alpha$ = .84) and teacher (Cronbach's $\alpha$ = .90) T1 CSUS-Short forms revealed high inter-item reliability; and the ratings of mothers and teachers were significantly and positively correlated with each other, $r$ = .25, $p < .05$. Intra-class correlation coefficient (ICC) was also calculated and found to be .70,

**Table 2.** Standardized factor loadings in CFA for parent- and teacher-reported CSUS-Short form (SE in brackets).

| Item | CSUS Parent report | CSUS Teacher report |
|------|--------------------|---------------------|
| 1 | .51 (.06) | .74 (.04) |
| 2 | .37 (.06) | .65 (.05) |
| 3 | .30 (.07) | .66 (.05) |
| 4 | .04 (.08) | .26 (.08) |
| 5 | .23 (.07) | .63 (.05) |
| 6 | .59 (.05) | .80 (.03) |
| 7 | .56 (.06) | .78 (.03) |
| 8 | .55 (.06) | .66 (.05) |
| 9 | .33 (.07) | .68 (.04) |
| 10 | .01 (.08) | .53 (.06) |
| 11 | .53 (.06) | .78 (.03) |
| 12 | .45 (.06) | .57 (.06) |
| 13 | .69 (.05) | .66 (.05) |
| 14 | .68 (.05) | .74 (.04) |
| 15 | .50 (.06) | .55 (.06) |
| 16 | .66 (.05) | .74 (.04) |
| 17 | .54 (.06) | .66 (.05) |
| 18 | .70 (.04) | .85 (.03) |
| RMSEA | 0.07 | 0.09 |
| CFI | 0.90 | 0.91 |
| SRMR | 0.06 | 0.07 |

indicating acceptable level of agreement between the ratings of parents and teachers. Child age was significantly correlated with the parent T1 CSUS-Short form scores, $r = .25$, $p < .001$. Parents scored girls ($M = 2.94$, $SD = 0.39$) higher on the T1 CSUS than boys ($M = 2.80$, $SD = 0.45$), $F (1, 196) = 4.89$, $p = .03$, $\eta2 = .02$, which fit the trend observed with gender difference in behavioral ToM performance, with girls ($M = 1.76$, $SD = 0.93$) outperforming boys ($M = 1.48$, $SD = 0.97$), $F (1, 200) = 4.64$, $p = .03$, $\eta2 = .02$. Similar to the parent report, teacher report of the T1 CSUS was positively correlated with child age, $r = .30$, $p < .001$. Teachers also rated girls ($M = 3.18$, $SD = 0.44$) higher on the T1 CSUS than boys ($M = 2.89$, $SD = 0.66$), $F (1, 162) = 11.64$, $p = .001$, $\eta2 = .07$. Given these correlations, we calculated the convergent validity analyses by controlling for age and gender. On the other hand, neither parents' education level (for T1 CSUS parent reports $r = .13$, $p = .06$; for T1 CSUS teacher reports $r = .13$, $p = .09$) nor the monthly family income (for T1 CSUS parent reports $r = .14$, $p = .06$; for T1 CSUS teacher reports $r = .03$, $p = .69$) was significantly linked with the CSUS ratings of parents or teachers, thus, these variables were not accounted for in the analyses.

## Convergent Validity Analyses

### Associations of the CSUS with the Behavioral Measures of ToM

The parent-reported T1 CSUS was significantly associated with children's ToM performance on knowledge access, unexpected change in location and contents false belief tasks (see Table 3). When gender and age were controlled, only knowledge access remained significantly linked with the T1 CSUS. The correlation of the parent-reported T1 CSUS with the composite behavioral ToM score was significant accounting for child age and gender. The teacher-reported T1 CSUS was associated with knowledge access and unexpected change in location scores of children, but not with contents false belief. Only knowledge access remained significant in its association with the teacher-reported T1 CSUS after controlling for age and gender. Similar to parent-reports, the teacher-reported T1 CSUS scores were significantly linked with the composite behavioral ToM scores of children. However, this association was no longer significant after controlling for child age and gender (see Table 3).

### Associations of the CSUS with Correlates of ToM

Both the parent-reported T1 CSUS and the teacher-reported T1 CSUS were significantly associated with children's EF, receptive language and social competence[2] even after accounting for child age and gender (see Table 3).

### Over-Time Stability and Predictors of the CSUS at T2

At T2, the parent and teacher reported CSUS scores displayed high internal consistencies (Cronbach's $\alpha = .84$ and $\alpha = .93$, respectively). Given the developmental continuity in ToM, we examined the over-time stability of the CSUS ratings and found positive associations between

Table 3. Zero-order and partial correlations of T1 CSUS with the study variables controlling for child age and gender.

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.T1 CSUS parent report | - | .40** | .24** | .21** | .16* | .28** | .37** | .27** | .51** | .52** | .25* |
| 2.T1 CSUS teacher report | .33** | - | .24** | .16* | .09 | .22** | .47** | .36** | .60** | .19 | .24* |
| 3. Knowledge access | .15* | .15* | - | .34** | .22** | .68** | .45** | .15* | .19** | .18* | .36** |
| 4. Unexpected change | .13 | .02 | .21* | - | .27** | .79** | .43** | .18* | .25* | .19* | .14 |
| 5. Contents false belief | .11 | −.04 | .08 | .15* | - | .69** | .30** | .12 | .27** | .04 | .12 |
| 6. Behavioral ToM | .19* | .05 | .55** | .74** | .66** | - | .54** | .21** | .28** | .19* | .26* |
| 7. EF | .27** | .35** | .17* | .26** | .10 | .27** | - | .30** | .41** | .29** | .30** |
| 8. Receptive language | .28** | .35** | .10 | .15 | .07 | .21** | .31** | - | .23** | .13 | .25* |
| 9. Social competence | .48** | .56** | .07 | .05 | .21* | .17* | .25** | .24** | - | .34** | .38** |
| 10.T2 CSUS parent report | .45** | .23 | .14 | −.22 | −.25 | −.19 | .10 | .09 | .27* | - | .26* |
| 11.T2 CSUS teacher report | .19 | .23 | .23 | .03 | .07 | .14 | .16 | .18 | .20 | .19 | - |

Zero-order correlations were presented above the diagonal and partial correlations were below the diagonal. T1 = Time 1. T2 = Time 2. CSUS = Children's Social Understanding Scale. ToM = Theory of Mind. EF = Executive Functions.
* $p < .05$. ** $p < .001$.

the T1 CSUS and the T2 CSUS scores in both parent ($r = .52$, $p < .001$) and teacher forms ($r = .24$, $p < .05$). Similar to ratings at T1 ($r = .25$, $p < .05$), parents and teachers both showed consistency in their evaluation of children's mental state understanding at T2, as indicated by significant positive correlations between parent and teacher forms of CSUS at T2 ($r = .26$, $p < .05$). ICC was .63, showing a moderate agreement.

We further explored early predictors of the T2 CSUS scores and tested a series of hierarchical linear regression models (see Table 4). We included behavioral ToM, EF, receptive language and social competence as criterion variables in the first step and found that, only early social competence significantly predicted the subsequent parent-reported CSUS scores at T2. However, when the T1 parent-reported CSUS was entered into the model, only parent-reported CSUS at T1 significantly predicted parent-reported CSUS scores at T2.[3] VIF value as an indicator of multicollinearity between the variables ranged from 1.14 to 1.56 at both steps of the regression, remaining below the recommended cutoff score of 5.00 (Hair et al., 1995). Social competence only marginally predicted the teacher-reported CSUS scores at T2 in the first step of the regression and when the teacher-reported CSUS at T1 was entered in the second step, none of the variables contributed to the model predicting the teacher reported CSUS at T2. VIF values were between 1.16 to 2.23.

Finally, given significant concurrent associations between the CSUS and behavioral ToM performance of children, we explored whether both scores can be equally used as predictors of the CSUS ratings at T2 (see Table 5), and tested whether T1 behavioral ToM would predict the T2 CSUS independent of the T1 CSUS. Behavioral ToM scores of children significantly predicted the T2 parent-reported CSUS scores in the first step of the regression. When the T1 parent-reported CSUS scores were added in the second step, only parents' ratings but not behavioral ToM emerged as a significant predictor of the CSUS at T2. For the T2 teacher-reported CSUS, behavioral ToM was a significant predictor in the first step of the regression, but after the addition of the T1 teacher-reported CSUS in the second step, none of the variables predicted the CSUS scores at T2.

**Table 4.** Hierarchical regression analyses predicting T2 CSUS scores from behavioral ToM, EF, receptive language, social competence and T1 CSUS.

|  | B | SE | β | $R^2$ | F |
|---|---|---|---|---|---|
| DV: T2 CSUS parent-report |  |  |  | .15 | 4.35** |
| Behavioral ToM | 0.03 | 0.05 | .06 |  |  |
| EF | 0.07 | 0.05 | .15 |  |  |
| Receptive language | −0.02 | 0.11 | −.02 |  |  |
| Social competence | 0.25 | 0.09 | .28* |  |  |
| DV: T2 CSUS parent-report |  |  |  | .30 | 20.42** |
| Behavioral ToM | 0.00 | 0.04 | .01 |  |  |
| EF | 0.05 | 0.05 | .11 |  |  |
| Receptive Language | −0.01 | 0.10 | −.01 |  |  |
| Social competence | 0.10 | 0.09 | .11 |  |  |
| T1 CSUS parent-report | 0.40 | 0.09 | .43** |  |  |
| DV: T2 CSUS teacher-report |  |  |  | .15 | 2.65* |
| Behavioral ToM | 0.10 | 0.07 | .19 |  |  |
| EF | 0.03 | 0.09 | .04 |  |  |
| Receptive language | 0.19 | 0.18 | .13 |  |  |
| Social competence | 0.26 | 0.16 | .22† |  |  |
| DV: T2 CSUS teacher-report |  |  |  | .15 | 0.01 |
| Behavioral ToM | 0.10 | 0.07 | .19 |  |  |
| EF | 0.03 | 0.10 | .04 |  |  |
| Receptive language | 0.19 | 0.19 | .13 |  |  |
| Social competence | 0.27 | 0.18 | .22 |  |  |
| T1 CSUS teacher-report | −0.01 | 0.13 | −.01 |  |  |

T1 = Time 1. T2 = Time 2. CSUS = Children's Social Understanding Scale. ToM = Theory of Mind. EF = Executive Functions.
† $p < .10$ * $p < .05$. ** $p < .001$.

**Table 5.** Hierarchical regression analyses predicting T2 CSUS scores from behavioral ToM and T1 CSUS.

|  | B | SE | β | $R^2$ | F |
|---|---|---|---|---|---|
| DV: T2 CSUS parent-report |  |  |  | .04 | 4.21* |
| Behavioral ToM | 0.08 | 0.04 | .20* |  |  |
| DV: T2 CSUS parent-report |  |  |  | .27 | 33.66** |
| Behavioral ToM | 0.03 | 0.03 | .07 |  |  |
| T1 CSUS parent-report | 0.46 | 0.08 | .50** |  |  |
| DV: T2 CSUS teacher-report |  |  |  | .06 | 4.17* |
| Behavioral ToM | 0.12 | 0.06 | .25* |  |  |
| DV: T2 CSUS teacher-report |  |  |  | .10 | 2.70 |
| Behavioral ToM | 0.10 | 0.06 | .20 |  |  |
| T1 CSUS teacher-report | 0.15 | 0.09 | .20 |  |  |

T1 = Time 1. T2 = Time 2. CSUS = Children's Social Understanding Scale. ToM = Theory of Mind.
* $p < .05$. ** $p < .001$.

## Discussion

This study adapted the parent version of the CSUS-short form into Turkish, examined its psychometric properties over one year, and investigated the utility of the CSUS as an instrument for teachers. The findings showed that both the parent and teacher versions of the CSUS-short form had good internal consistency and acceptable stability in scores across a one-year period. Converging ratings of parents and teachers indicated that both groups were reliable informants for evaluating children's social understanding skills. Although parent- and teacher-reports have long been used in developmental research measuring children's social behaviors (e.g., aggression, prosocial behavior) (Fantuzzo et al., 1998), our findings, in addition to others (see Hutchins et al., 2012), supported that parents and teachers can also be consulted for brief assessment of children's socio-cognitive skills. Given that socio-cognitive skills have been measured predominantly through individual tasks that have required considerable time and effort from researchers and young children, the application of short scales such as the CSUS could be especially practical.

Moreover, the content of the CSUS-short form revealed a single factor structure in both parents' and teachers' ratings, which conforms to the original form of the CSUS (Tahiroglu et al., 2014) and its adaptations to other cultures (Smogorzewska et al., 2019). The CSUS was designed to incorporate different facets of ToM; however, on the grounds that social understanding requires simultaneous use of these different facets within the complexity of everyday interactions, Tahiroglu et al. (2014) recommended researchers to use the total score of the scale rather than the subscale scores. The single-factor structure of the CSUS-short form supports this, which comprehensively inspects children's reasoning about others' subjective perspectives.

In addition to internal consistency and reliability across the informants, zero-order correlations showed that parent- and teacher-reports of the CSUS were positively related with the overall scores of children in classical behavioral assessments of ToM, highlighting the convergent validity of the CSUS-short form in Turkish. In line with its content structure that taps into multiple mental states to assess a general underlying mental state understanding performance, the CSUS was related to the composite behavioral ToM scores rather than individual behavioral ToM tasks (except for knowledge access) which target a particular type of mental state (e.g., belief, perception, intention). However, when child age and gender were controlled, only parent- but not teacher-report of the CSUS significantly correlated with overall behavioral measures of ToM. Given that teachers observe behaviors of multiple children at varying ages, their ratings might be swayed by their expectations of the skills children would normatively display with age and gender (Auwarter & Aruguete, 2008). In contrast, parent ratings might reflect their attentiveness to individual behaviors of their children in their own entirety without much consideration of the child's age and gender. Nevertheless, considering that the diverse instruments used to assess ToM of children have minimal associations with each other, the significant zero-order correlations of parent- and teacher-reports of CSUS with behavioral measures of ToM appear even more critical. It should be noted that different but widely established ToM assessments

(e.g., appearance-reality distinction, faux-pas, false belief tasks) do not show coherence with each other in early and middle childhood (Warnell & Redcay, 2019), indicating that they are likely to evaluate diverse aspects of a multi-dimensional process called mental state inference (Apperly, 2012). Thus, within this multi-dimensionality of ToM as a construct, overlapping scores of the CSUS-short form with classical behavioral tasks in zero-order correlations demonstrate the validity of the CSUS as an appropriate measurement tool for researchers.

As for the strength of the correlations between the CSUS-short form and behavioral measures of ToM, the correlation coefficients Tahiroglu et al. (2014) reported in Study 1 ($r$ (81) = .37, $p < .01$) were comparable to the one in our study ($r$ (202) = .28, $p < .001$). Moreover, the correlation between the CSUS-short form and behavioral ToM scores was not significant controlling for child age ($r$ (81) = .17, ns) in Tahiroglu et al.'s (2014) Study 1 while it was significant in our study ($r$ (202) = .19, $p < .05$). In addition, the correlation coefficients reported in our study were similar to the correlations reported between behavioral ToM scores and other ToM-scales (see, Bosco et al., 2016; Wang & Wang, 2015). As for the ratings of teachers, a weak-to-modest association was found between teacher reports of the CSUS-short form and children's behavioral ToM scores. Teachers tried to gauge mental state understanding skills of children in the school context within the peer group and therefore, their inferences of children's ToM skill might slightly differ from inferences of the parents. This might be one of the reasons why the association of behavioral ToM with teacher ratings of the CSUS was somewhat weaker than that with parent ratings of the CSUS. Together these findings highlighted that for such a multi-faceted concept as ToM a low-to-moderate association is found between individual task performance and scale measures. Although caution should be taken while interpreting results of scale measures of ToM, such as the CSUS, our findings demonstrated that the CSUS scores can be used as approximate indicators of children's mental state understanding skills especially when behavioral ToM tasks are not possible options for assessment.

When the CSUS scores were examined in relation to other widely known correlates of ToM, both parent and teacher reports displayed significant and positive association with EF and language performance of children, independent of age and gender. The literature consistently shows that cognitive regulation (Devine & Hughes, 2014) and vocabulary comprehension (Milligan et al., 2007) are involved in children's awareness of diverse minds underlying others' observable behaviors. EF helps children inhibit their own perspective, focus on the subjective mental states of others and hence, enables accurate mental state inference. On the other hand, language provides the connection between observable behaviors and the unobservable minds behind these behaviors, through labeling diverse mental states. Similarly, our findings, in line with the other cultural adaptations of the CSUS (e.g., Białecka-Pikul & Stępień-Nycz, 2019), showed that higher EF and language performance of children were concurrently associated with both higher behavioral ToM scores and higher CSUS ratings of parents and teachers.

In addition to EF and language, the results showed that both behavioral ToM performance of children and CSUS ratings of parents and teachers, relate to higher social competence even when age and gender were controlled. In fact, among the correlates of ToM, the CSUS ratings of parents and teachers yielded the highest association with social competence. One possible explanation of this finding could be shared informant variance. Both the CSUS and social competence were measured using parent and teacher reports, whereas EF and receptive language were measured with individual tasks. The higher correlation between the CSUS and social competence compared to the associations between the CSUS and other ToM correlates might reflect parents' and teachers' parallel evaluation of children's mental state understanding, as well as their social behaviors. Future studies could address this shared informant variance by assessing social competence through independent observations. Another reason for this high correlation could be that the content of the CSUS has been designed to capture children's behaviors that reflect their mental state understanding in daily social interactions. Some items of the CSUS gauge children's mental state inference through their social exchanges, such as peer interactions and play behaviors, which also touches on children's social competence. Noting this however, the significant associations between social competence and both behavioral measures of

ToM and the CSUS support the construct validity of the CSUS-short form in the Turkish sample. This implies that the significance of the relationship between the CSUS and children's social competence could arise not primarily from the shared informant variance but rather from the true associations between the two constructs. It appears that children with higher ToM skills might better understand the subjective minds of their peers, which helps them solve emerging conflicts with peers and bolster their cooperative actions. Children's developing insight into diverse mental states of others makes their social interactions more sophisticated. Furthermore, higher mental state understanding skills might help young children display more harmonious and less disruptive behaviors within peer groups (Imuta et al., 2016).

Children's insight into other peoples' mental states displays a protracted development and therefore, investigating predictors of children's ToM development is of importance when planning appropriate interventions for enhancement of ToM (Wellman, 2014). While EF and language skills of children have long been debated as the early indicators of ToM (Apperly et al., 2009), our findings showed that more than EF or language, it was the children's earlier social competence that predicted their future parent-reported CSUS performance. Based on the correlation and regression results, children's social competence was concurrently and subsequently linked to parents' ratings of mental state understanding. As indicated by the recent review of Weimer et al. (2021), toward the end of the preschool years social relationships become critical in providing a context where children can improve their social understanding skills. Findings show that those who are deprived of positive peer interactions are at risk of showing lower mental state understanding in later years (Banerjee et al., 2011) as peer relationships help children figure out how to make sense of the behaviors of their interaction partners. Although most of the studies focus on the role of early ToM in predicting children's future social competence and reported significant longitudinal associations in preschool period, once mental state understanding starts to get more advanced, social relationships may further enhance the existing mentalization skills of children (Weimer et al., 2021). Regardless, several mediating and moderating factors, including emotional self-regulation and reflective processes possibly exist to impact the direction of the relationship between ToM and social behaviors. These factors should be considered in future research when examining the direction of the association between children's mental state understanding and their social competence (Weimer et al., 2021).

Although social competence is important for future ToM scores, our results revealed that when early parent-reported CSUS scores were considered, social competence no longer had significance in predicting children's subsequent CSUS scores. While shared variance might explain these results, as parents rated the CSUS two times one year apart, it is also possible that early accomplishments in ToM might pave the way to future advances in ToM, having greater impact on the development of ToM than social competence. A similar trend was observed when parent-reported CSUS scores at T2 were initially predicted by early behavioral ToM performance. These findings were later found to be non-significant when early CSUS ratings of parents were added to the regression model, showing that continuity in parents' CSUS ratings had an important and stronger role than behavioral ToM scores in estimating future mental state understanding performance of children. However, overall, and apart from early CSUS scores, the fact that children's previous performance in behavioral ToM tasks predicted future CSUS ratings implies convergence of the two assessments (e.g., behavioral and scale) and validity of the CSUS as an appropriate tool for ToM measurement.

### *Limitations*

Our study is not without limitations and some of our initial findings may be indicative of these. First, behavioral ToM performance of children initially predicted subsequent CSUS ratings of teachers but after early teacher-reported CSUS scores were entered into the regression, neither behavioral ToM nor early CSUS ratings predicted the CSUS scores at T2. Furthermore, EF,

language and social competence were not linked to subsequent teacher ratings of the CSUS before and after early teacher ratings were considered. The following methodological factors might explain these results. (1) Some teachers were asked to rate more than one child in a class for the CSUS, which violates the main assumption of independence of observations for the conducted analyses, and (2) most teachers who completed the CSUS at T1 were different than the teachers who completed the CSUS at T2. This could have influenced the over-time stability of the teacher ratings as well as the regression results. Despite these potential limitations, the teacher-reported CSUS showed similar overall psychometric properties (e.g., internal consistency, factor structure, over-time stability and concurrent associations with correlates of ToM) to parent-reported CSUS. To overcome the limitation regarding teacher ratings, the teacher reports could be administered at the beginning and end of the same school year, increasing the possibility of inclusion of the same teachers. Future research considering these limitations might provide statistically stronger inference.

As for the over-time stability of the parent and teacher-reported CSUS scores, drop-out rates at the second measurement point should be acknowledged. No difference was observed in the teacher-reported CSUS scores and behavioral ToM performance between children who were present at both time points and children who dropped out one year later. However, those who remained in the study had higher parent-reported CSUS scores at T1 than those who left the study at the second measurement. This attrition might have created a bias in the over-time stability of the parent-reported CSUS scores, such that children with initially higher mental state understanding continued to display higher performance over the year (contributing to consistency in the results of the CSUS). However, examination of parent-reported CSUS scores one year later revealed that parents' ratings demonstrated a wide range in all age groups, showing diverse scores from low to high. Therefore, variability in the CSUS scores was observed to continue despite the attrition over the year. This variability, along with the stability in teacher-ratings across time, lend support to the CSUS offering a reliable assessment of ToM over time.

Finally, although parent and teacher-reports of the CSUS were significantly associated with each other at both time points, the strength of the association appeared to be weak-to-modest. Three factors could be considered when interpreting these associations. First, the context in which the association was revealed is important when making sense of the degree of associations (Adachi & Willoughby, 2015). On its own, ToM is already a tricky concept to gauge, which is why it has been assessed so far only through behavioral tasks rather than scales (Tahiroglu et al., 2014). With the existing diversity in ToM tasks, researchers examined the convergence between them to determine if diverse behavioral measures are equally good indicators of ToM (Quesque & Rossetti, 2020). Except for false belief tasks, a weak-to-modest association has been found among scores of different behavioral ToM tasks (Warnell & Redcay, 2019). Thus, considering that even behavioral assessments of ToM could yield significant but weak correlations among each other, it is reasonable to expect that observers could only weakly-to-moderately agree among themselves on their ratings of children's ToM. Afterall, while filling out the CSUS, parents and teachers rely on their own observations of children's social behaviors and try to see if these behaviors carry the signs which display children's understanding of others' minds. Secondly, in addition to the difficulty of observing ToM in social behaviors of children (Astington, 2003), differences in time and contexts of observation (school vs. home) can influence the strength of the association between ratings of two informants (De Los Reyes & Kazdin, 2005). Divergence in ratings of informants is commonly seen even in the same contexts (see Gluck et al., 2021). Thirdly, it is important to highlight that the correlation coefficients we found are in line with those reported in other studies. Parent and teacher ratings usually overlap in modest strength especially for measurement of constructs that involve assessment of children's socio-cognitive and emotional skills (Gresham et al., 2018; Renk & Phares, 2004). Yet, the complications of measuring these constructs as well as the informants' diverse contexts of observation could make the statistically significant associations of informant ratings valuable despite the weak-to-modest correlation coefficients.

## Implications and Conclusions

In general, our study demonstrated that both parent and teacher versions of the CSUS- short form could be used as valid and reliable instruments to assess ToM skills of young Turkish children. The Turkish version of the CSUS could provide a way to compare data on ToM from different cultural groups, and hence, allows for studying development of mental state understanding in cross-cultural contexts. This point is crucial, since the social and cultural influences on ToM are widely studied from comparative perspectives (Wang et al., 2016), requiring culturally comparable assessment tools. While studies on ToM development of Turkish children predominantly rely on behavioral assessments translated from classical ToM tasks (Korucu et al., 2017; Selcuk et al., 2018), they could benefit from short and culturally comparable measures such as the CSUS.

The existing adaptations of the CSUS to different languages/cultures supports that the CSUS can be applied cross-culturally in intervention programs on ToM development. Children from diverse cultural groups can be assessed regarding their need for ToM intervention (e.g., especially those who experience significant problems in peer interactions due to deficient ToM), or alternatively, their improvement in mental state understanding over the course of an intervention can be evaluated through brief reports of parents and teachers. Moreover, in educational settings, CSUS can be used as a practical tool to assess variation in social understanding skills of children. Teachers may benefit from the brief structure of the CSUS and more easily apply it to identify the relative standing of the children in class in terms of their insight into unobservable mental states of others. This might help teachers tailor their educational practices according to the needs of the children. Teachers can also utilize the parent-form of the CSUS to cross-validate their assessment, and therefore have a grounded idea about socio-cognitive skills of children in their classroom. As such, due to its brief and practical application, CSUS can be used for both cross-cultural research settings as well as in early childhood educational contexts.

In short, this study showed for the first time that CSUS-short form can be utilized reliably as a parent and teacher report to assess general ToM performance of young Turkish children. Importantly, it also demonstrates that the CSUS-short form can be used as a practical tool to evaluate children's ToM development comprehensively, enabling different informants who observe children in different contexts (e.g., home and school) to report their observations and to identify individual differences in children's ToM scores for research purposes.

## Notes

1. In our sample, the number of missing values was low with 25, 16 and 11 of the 202 parents, missing one, two or three items (respectively), while 42, 38 and 14 of the 169 teachers had one, two or three items missing (respectively).
2. When examined separately, social competence ratings of parents were significantly linked with CSUS ratings of parents ($r = .63$, $p < .001$) and teachers ($r = .36$, $p < .001$). Likewise, social competence ratings of teachers were significantly correlated with CSUS ratings of parents ($r = .24$, $p < .001$) and teachers ($r = .62$, $p < .001$).
3. The statistical significance levels of the standardized beta coefficients were the same when age and gender were accounted for in the regression analyses.

## Acknowledgments

## Disclosure statement

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Funding

## ORCID

Müge Ekerim Akbulut ⓘD http://orcid.org/0000-0002-2074-8003

## References

Adachi, P., & Willoughby, T. (2015). Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science. *European Journal of Developmental Psychology*, *12*(1), 116–128. https://doi.org/10.1080/17405629.2014.963549

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. Quarterly Journal of Experimental Psychology, 65(5),825–839. https://doi.org/10.1080/17470218.2012.676055

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, *45*(1), 190–201. https://doi.org/10.1037/a0014098

Arseneault, L., Moffitt, T. E., Caspi, A., Taylor, A., Rijsdijk, F. V., Jaffee, S. R., . . . Measelle, J. R. (2003). Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *Journal of Child Psychology and Psychiatry*, *44*(6), 832–848. https://doi.org/10.1111/1469-7610.00168

Astington, J. W. (2003). Sometimes necessary, never sufficient: False belief understanding and social competence. In B. Repacholi& V. Slaughter (Eds.), *Individual differences in theory of mind* (pp. 13–38). Psychology Press.

Auwarter, A. E., & Aruguete, M. S. (2008). Effects of student gender and socioeconomic status on teacher perceptions. *The Journal of Educational Research*, *101*(4), 242–246. https://doi.org/10.3200/JOER.101.4.243-246

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

Berument, S. K., & Güven, A. G. (2010). *Turkish expressive and receptive language test: Receptive vocabulary sub-scale (TIFALDI-RT)*. Turkish Psychological Association.

Białecka-Pikul, M., & Stępień-Nycz, M. (2019). Parent-report measure of theory of mind: Adaptation to Polish of the Children's Social Understanding Scale. *European Journal of Developmental Psychology*, *16*(3), 318–326. https://doi.org/10.1080/17405629.2017.1406850

Bosco, F. M., Gabbatore, I., Tirassa, M., & Testa, S. (2016). Psychometric properties of the theory of mind assessment scale in a sample of adolescents and adults. *Frontiers in Psychology*, *7*(566), 1–12. https://doi.org/10.3389/fpsyg.2016.00566

Brosseau-Liard, P., & Poulin-Dubois, D. (2018). Fiabilité et validité de l'Échelle de compréhension sociale des enfants [Reliability and validity of the French adaptation of the Children's Social Understanding Scale]. *Psychologie Française* 64(4), 331–341. https://doi.org/10.1016/j.psfr.2018.01.003

Davis, M. H. (1996). *Empathy: A social-psychological approach*. Westview.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. Psychological Bulletin, 131(4), 483–509. doi:10.1037/0033-2909.131.4.483

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777–1794. https://doi.org/10.1111/cdev.12237

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." *Developmental Psychobiology*, *29*, 315–334. https://doi.org/10.1002/(SICI)10982302(199605)29:4<315::AIDDEV2>3.0.CO;2-T

Ekerim, M., & Selcuk, B. (2018). Longitudinal predictors of vocabulary knowledge in Turkish children: The role of maternal warmth, inductive reasoning, and children's inhibitory control. *Early Education and Development*, *29*(3), 324–341. https://doi.org/10.1080/10409289.2017.1407607

Etel, E., & Yagmurlu, B. (2015). Social competence, theory of mind, and executive function in institution-reared Turkish children. *International Journal of Behavioral Development*, *39*(6), 519–529. https://doi.org/10.1177/0165025414556095

Fantuzzo, J., Sutton-Smith, B., Coolahan, K. C., Manz, P. H., Canning, S., & Debnam, D. (1995). Assessment of preschool play interaction behaviors in young low-income children: Penn Interactive Peer Play Scale. *Early Childhood Research Quarterly*, *10*(1), 105–120. https://doi.org/10.1016/0885-2006(95)90028-4

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology*, *47*(2), 404–416. https://doi.org/10.1037/a0021065

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2–7 years old on a Stroop-like day–night test. *Cognition*, *53*(2), 129–153. https://doi.org/10.1016/0010-0277(94)90068-X

Gluck, S. C., Tahiroglu, D., & Moses, L. J. (2021). A comparison of mother and father reports of children's theory of mind: Further validation of the children's social understanding scale. *European Journal of Developmental Psychology*, *18*(3), 459–469. https://doi.org/10.1080/17405629.2020.1813100

Gresham, F. M., Elliott, S. N., Mlo, S., Byrd, S., Wilson, E., & Cassidy, K.et al (2018). Cross-informant agreement of children's social-emotional skills: An investigation of ratings by teachers, parents, and students from a nationally representative sample. Psychology in the Schools, 55(2), 208–223. https://doi.org/10.1002/pits.22101

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3rd ed.). Macmillan.

Hughes, C. (2011). *Social understanding and social lives from toddlerhood through to the transition to school*. Psychology Press.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, *18*(6), 667–683. https://doi.org/10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *42*(3), 327–341. https://doi.org/10.1007/s10803-011-1244-7

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology*, *52*(8), 1192–1205. https://doi.org/10.1037/dev0000140

Jang, M., & Shin, N. (2018). Effects of maternal parenting behavior and preschoolers' executive function on preschoolers' theory of mind. *Korean Journal of Child Studies*, *39*(1), 103–117. https://doi.org/10.5723/kjcs.2018.39.1.103

Korucu, I., Selcuk, B., & Harma, M. (2017). Self-regulation: Relations with theory of mind and social behaviour. Infant and Child Development, 26(3), e1988. https://doi.org/10.1002/icd.1988

Korucu, I., Selcuk, B., & Harma, M. (2017). Self-regulation: Relations with theory of mind and social behaviour. *Infant and Child Development*, *26*(3), e1988. https://doi.org/10.1002/icd.1988

LaFreniere, P. J. & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). Psychological Assessment, 8(4), 369–377.10.1037/1040-3590.8.4.369.

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. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.2307/1131316

Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, *18*(3), 399–429. https://doi.org/10.1016/S0885-2014(03)00041-8

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, *15*(2), 384–396. https://doi.org/10.1177/1745691619896607

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, *24*(2), 239–254. https://doi.org/10.1016/j.cpr.2004.01.004

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, *34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Selcuk, B., Brink, K. A., Ekerim, M., & Wellman, H. M. (2018). Sequence of theory-of-mind acquisition in Turkish children from diverse social backgrounds. *Infant and Child Development*, *27*(4), e2098. https://doi.org/10.1002/icd.2098

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239–1247. https://doi.org/10.1037/a0023899

Sheatsley, P. B. (1983). Questionnaire construction and item writing. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survery research* (pp. 195–230). Academic Press.

Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. Child Development, 86(4), 1159–1174. https://doi.org/10.1111/cdev.12372

Smogorzewska, J., Szumski, G., & Grygiel, P. (2019). The Children's Social Understanding Scale: An advanced analysis of a parent-report measure for assessing theory of mind in Polish children with and without disabilities. *Developmental Psychology*, *55*(4), 835–845. https://doi.org/10.1037/dev0000673

Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E., Olofson, E. L., & Sabbagh, M. A. (2014). The Children's Social Understanding Scale: Construction and validation of a parent-report measure for assessing individual differences in children's theories of mind. *Developmental Psychology*, *50*(11), 2485–2497. https://doi.org/10.1037/a0037914

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

Wang, Z., & Wang, L. (2015). The mind and heart of the social child: Developing the empathy and theory of mind scale. *Child Development Research*, *2015*, Article ID 171304. https://doi.org/10.1155/2015/171304

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*, 103997. https://doi.org/10.1016/j.cognition.2019.06.009

Weimer, A. A., Warnell, K. R., Ettekal, I., Cartwright, K. B., Guajardo, N. R., & Liew, J. (2021). Correlates and antecedents of theory of mind development during middle childhood and adolescence: An integrated model. *Developmental Review*, *59*, Article 100945. https://doi.org/10.1016/j.dr.2020.100945

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

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

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

Yagmurlu, B., Berument, S. K., & Celimli, S. (2005). The role of institution and home contexts in theory of mind development. Journal of Applied Developmental Psychology, 26(5), 521–537. https://doi.org/10.1016/j.appdev.2005.06.004

Yagmurlu, B., & Sanson, A. (2009). Parenting and temperament as predictors of prosocial behaviour in Australian and Turkish Australian children. Australian Journal of Psychology, 61(2), 77–88. https://doi.org/10.1080/00049530802001338

Yagmurlu, B. (2014). Relations among sociocognitive abilities and prosocial behavior. *Journal of Child and Family Studies*, *23*(3), 591–603. https://doi.org/10.1007/s10826-013-9726-1