Developing a Measure of Sarcasm Understanding in South African Children

Atida M. Moyo and Chloé E. Carr

MYXATI001   CRRCHL001

ACSENT Laboratory

Department of Psychology

University of Cape Town

15 January, 2021

Supervisor: A/Prof Susan Malcolm-Smith

Word count:

Abstract: 249

Main text: 7918

PLAGIARISM DECLARATION

1.   I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend that it is one's own.

2.   I have used the *American Psychological Association (APA)* convention for citation and referencing. Each significant contribution to, and quotation in, this essay / report / project / from the work, or works, of other people has been attributed, and has been cited and referenced.

3.   This essay /report /project / is my own work.

4.   I have not allowed, and will not allow, anyone to copy my work with the intention of passing if off as his or her own work.

5.   I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

SIGNATURE: Atida Moyo

                        Chloe´ Carr

Abstract

Sarcasm understanding is typically measured using lie-joke story tasks, which assess children's ability to discern when a speaker means the opposite of what is said. The literature suggests that most lie-joke measures demonstrate methodological inconsistencies and are limited to a Western context. In response to an ineffective lie-joke task from the UCT Theory of Mind Battery, the present study aimed to develop a contextually appropriate lie-joke measure to assess sarcasm understanding in English-speaking South African children. A cross-sectional, repeated measures design was used to examine the reliability, validity, and effectiveness of the revised lie-joke measure. It was hypothesised that our revised measure would be both reliable and valid; would be more effective at discerning age-appropriate sarcasm understanding than the original lie-joke task; and would indicate a significant difference in sarcasm understanding between younger and older children. South African school-going participants (N = 63) of a high-socioeconomic status (SES) were selected via snowball sampling. We initially planned to recruit children of a low-SES; however, this was not possible due to the lockdown restrictions that resulted from the COVID-19 pandemic of 2020. Our measure was found to be reliable, demonstrated reasonable convergent validity, was more effective in discerning age-appropriate sarcasm understanding than the original lie-joke task, and indicated a significant difference in sarcasm understanding between younger and older participants. Overall, we found our revised lie-joke measure to be useful in assessing sarcasm understanding in high-SES, English-speaking South African children.

Keywords: Sarcasm understanding, lie-joke measure, South African children, psychometrics, between-group differences

Developing a Measure of Sarcasm Understanding in South African Children

The acquisition of socio-cognitive skills is a critical aspect of human development. Skills such as emotion recognition, social learning and theory of mind (ToM) play an essential role in one's ability to interpret social interactions and agents' behaviour (Happé et al., 2017). ToM is the understanding of one's own and others' mental states, including thoughts, emotions, beliefs and intentions (Bailey & Im-Bolter, 2020; Happé et al., 2017; Lagattuta et al., 2015). This understanding increases throughout childhood (Lagattuta et al., 2015), with belief-desire ToM (i.e., the understanding that beliefs and desires determine the actions of others) typically emerging at the age of three; perspective-taking (i.e., the awareness that people can form beliefs about the beliefs of others), between the ages of six and seven; and more abstract forms of thinking, from eleven onwards (Gopnik & Wellman, 1992).

ToM research has focused primarily on early-childhood, specifically assessing the age at which children pass the benchmark of ToM acquisition, false-belief reasoning, indicating an understanding that people's beliefs do not always match reality (Lagattuta et al., 2015; Wellman et al., 2001). In contrast, research on ToM development in middle childhood has been described as both limited and unsystematic (Im-Bolter et al., 2016; Lagattuta et al., 2015). Important social, emotional and cognitive development takes place between the ages of 6 and 12 years old, where children begin to develop the necessary skills to navigate the complexities of social relationships (Lagattuta et al., 2015). Improved ToM abilities in middle childhood provide children with the necessary skills to better understand others' mental states and grasp more subtle forms of communication, including deception, persuasion, faux pas, and sarcasm (Happé, 1994; Lagattuta et al., 2015).

Understanding Sarcasm

Well-developed ToM abilities are fundamental to comprehending sarcasm, as the meaning of a sarcastic utterance is opposite to the verbalisation (Filik et al., 2019). Grice (1975) suggests that sarcasm violates the three conversational maxims of literal communication: truth, belief and literalness. A sarcastic comment is false, thus violating the truth maxim; the speaker does not believe what they have uttered, thus violating the belief maxim; and the utterance contradicts what is meant, thus violating the literalness maxim. Prior studies show that children find the violation of the literalness maxim the most challenging to comprehend because it contradicts reality (Angeleri & Airenti, 2014). Consequently, children interpret sarcastic utterances based on the violations of only the truth and belief maxims and misperceive sarcasm as deception (Winner et al., 1987).

Cues Necessary for Sarcasm Comprehension

Listeners need to recognise verbal and non-verbal cues to detect the literalness maxim violation and perceive the humour or criticism in a sarcastic message (Winner et al., 1987). Prosody (i.e., rhythm, tempo, pitch, and volume) is arguably one of the more salient cues used to detect sarcasm, demonstrated through the use of monotone or exaggerated intonation in several studies (Capelli et al., 1990; Glenwright et al., 2014). Interestingly, some studies have found cross-cultural differences in the use of sarcastic intonation. For example, Cheang and Pell (2009) found differences in the mean fundamental frequency, amplitude rate, mean amplitude, and harmonics-to-noise ratio between Cantonese and English speakers. These findings not only highlight the importance of prosody as a cue for conveying sarcasm but also show that specific acoustic conventions for communicating sarcasm may differ across cultures.

There is consensus in the literature that verbal cues are critical in detecting the discrepancy between the verbal and contextual meanings in a sarcastic scenario (Capelli et al.,

1990). However, some scholars argue that non-verbal cues, such as facial expression, are equally important; suggesting that eye rolls, squinting, smirking, raised eyebrows and a blank face are all useful cues in sarcasm comprehension (Attardo et al., 2003). Facial expressions aid communication of the emotions behind a sarcastic comment, including anger, embarrassment or contempt (Rockwell, 2001). However, the meaning behind non-verbal cues is contextually dependent. For example, smiling may accompany sarcasm when one is embarrassed, indicating the relevance of context when interpreting facial expressions (Russell, 1994).

Development of Sarcasm Comprehension in a Western Context

The age that children begin to utilise these cues and develop an understanding of sarcasm has long been debated (Capelli et al., 1990; Demorest et al., 1984; Dews et al., 1996). Research in Western contexts suggests that children as young as 4 years old demonstrate the basic understanding that sarcasm implies a meaning opposite to reality (Ackerman, 1983). However, the ability to comprehend the humorous intent behind a sarcastic utterance requires more developed ToM skills and only emerges in middle childhood (Cutting & Dunn, 1999; Pexman, 2018). Younger children's inability to identify sarcastic intent is associated with their failure to infer the complex mental states of others. An emergent appreciation of the humour in sarcasm develops in children from 8 or 9 years old, due to their increased ToM abilities (Angeleri & Airenti, 2014). By 10 years old, children demonstrate the ToM necessary to infer the mental states of others and comprehend the teasing or mocking intent behind a sarcastic comment (Recchia et al., 2010).

Cultural Influences on Sarcasm Comprehension

Several non-Western studies oppose the idea that there is a particular age that children begin to comprehend sarcasm, arguing that the development of sarcasm understanding is

influenced by sociocultural factors (Hughes et al., 2014). Cultural differences in communication patterns within families are thought to influence children's understanding and use of sarcasm (Recchia et al., 2010), with an Iranian study finding that ToM development is a function of cultural norms and parental involvement in sibling play and conflict (Shahaeian, 2015). Furthermore, a Polish study concluded that children as young as 4 years old demonstrated a well-developed understanding of sarcasm due to cultural norms and the regular use of figurative language in Polish communication (Banasik-Jemielniak & Bokus, 2019). Culture underlies the social norms regarding power dynamics and respect that influence individuals' use of sarcasm in communication (Shahaeian et al., 2014). A child will only understand the implied sarcasm in a statement if they are familiar with society's norms surrounding sarcasm use (Creusere, 1999). For example, children may be "more likely to expect sarcasm from their older brother than from their teacher" (Capelli et al., 1990, p. 1837). Therefore, it is imperative to consider the cultural complexities of communication and embed sarcasm measures within familiar scenarios that promote understanding (Angeleri & Airenti, 2014).

Measuring Sarcasm Understanding

Sarcasm understanding is generally measured using lie-joke story tasks that assess children's ability to discern whether a speaker means the opposite of what is said (Angeleri & Airenti, 2014; Banasik-Jemielniak & Bokus, 2019; Demorest et al., 1984; Dews et al., 1996; Miller, 2006; Perner et al., 2002; Shamay-Tsoory et al., 2005). Most lie-joke measures are embedded within a Western, English-speaking context due to a lack of research into cultural differences in sarcasm comprehension (Banasik-Jemielniak & Bokus, 2019). Research indicates that available lie-joke measures vary in complexity, narrative length, vocabulary, syntax, number of characters, and the relationship between interlocutors. The abovementioned irregularities

place inconsistent demands on children's memory and comprehension skills, leading to contradictory conclusions regarding children's understanding of sarcasm in middle childhood (Banasik-Jemielniak & Bokus, 2019; Creusere, 1999).

Rationale

Overall, the literature demonstrates that research on sarcasm understanding in middle childhood is methodologically inconsistent (Rockwell & Theriot, 2001) and limited to a Western context (Creusere 1999; Shahaeian et al., 2014). A systematic study into the development of sarcasm understanding within a variety of social and cultural settings is required, which indicates the need for the construction of a reliable and valid lie-joke measure that accurately assesses sarcasm understanding in middle childhood in the South African context.

Research Aims and Questions

Following a previously ineffective lie-joke measure in the UCT Theory of Mind Battery (Hoogenhout & Malcolm-Smith, 2014; see Appendix A), this study aimed to develop a contextually appropriate lie-joke measure (see Appendix B) to assess sarcasm understanding in English-speaking South African children.

Research Question 1

Is the New Measure Reliable? We calculated the interrater reliability of the revised lie-joke measure, using Cohen's Kappa. The split-half reliability of this measure was estimated using the Spearman-Brown prophecy formula.

Hypothesis 1. The revised lie-joke measure will demonstrate good interrater reliability and internal consistency.

Research Question 2

Is the New Measure Valid? We investigated the convergent validity of the revised lie-joke measure by correlating it with four measures with which it should be associated: the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II) vocabulary subtest (Wechsler, 2011; see Appendix C), the parent-report Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers et al., 2011; see Appendix D), and the strange stories and faux pas tasks from the UCT Theory of Mind Battery (Hoogenhout & Malcolm-Smith, 2014; see Appendices E & F, respectively).

Hypothesis 2. The revised lie-joke ToM scores will moderately correlate with the four convergent measures: the WASI-II vocabulary subtest, parent-report QCAE, strange stories, and faux pas scores.

Research Question 3

Will Participants Aged 10 Years and Older Demonstrate Better Sarcasm Comprehension on the Revised Lie-Joke Measure when Compared to the Original Lie-Joke Measure? Research suggests that children aged 10 years and older should demonstrate a good understanding of sarcasm.

Hypothesis 3. The participants who are 10 years and older will demonstrate a better understanding of sarcasm on the revised lie-joke measure when compared to the original lie-joke measure.

Research Question 4

Is there a Significant Difference in Sarcasm Understanding between Children Younger than 10 Years Old and Children Aged 10 Years and Older? Research suggests that sarcasm understanding develops across middle childhood, with older children demonstrating better sarcasm comprehension than younger children.

Hypothesis 4. There will be a significant age difference in sarcasm understanding, with children 10 years and older demonstrating better sarcasm understanding than children younger than 10 years old.

Method

Design and Setting

This research formed part of an ongoing study on the development of moral reasoning and empathy in children. A cross-sectional, repeated measures design was implemented to examine the reliability, validity, and effectiveness of the revised lie-joke measure (Appendix B) and quantitative measures were used to assess the participants' empathy, vocabulary, and ToM. The study required children to complete six tasks over two 30-40 min sessions and parents to complete two questionnaires in their own time. Data were collected online, and participants were recruited via snowball sampling.

Participants

This study aimed to recruit 70 participants, who were school-going children from a high-socioeconomic background. Ideally, we should have included children of low-socioeconomic status (SES), as the majority of South African households fall within this bracket (Western Cape Government, 2017). However, due to the COVID-19 pandemic of 2020, all non-essential contact was limited throughout the year. Therefore, the study design was adapted to measure sarcasm understanding in children with unlimited internet access and who possessed the necessary technology for online communication. All participants were from South Africa, and data collection took place between July and September 2020.

Estimated Required Sample Size

Assessing the convergent validity and effectiveness of our measure were both critical aspects of this study. G*Power analysis (Faul et al., 2009) indicated that with parameters set to analysis = Pearson product-moment correlation test, effect size (r) = .45 (moderate), α = .05, power (1- β) = .80, an N of 36 was sufficient to demonstrate the validity of our revised lie-joke measure. Both within and between group differences were key to assessing the effectiveness of our measure. It was determined that an N of 43 was sufficient, with parameters set to analysis = paired samples t-test, effect size (d) = .45 (medium), α = .05, power (1- β) = .80, to assess if participants' sarcasm comprehension scores improved from the original to the revised lie-joke measure. Furthermore, it was determined that an N of 60 was sufficient, with parameters set to analysis = independent samples t-test, effect size (d) = .65 (medium), α = .05, power (1- β) = .80, to assess if there was a significant difference in sarcasm understanding between younger and older participants (i.e., participants younger than 10 years old vs participants aged 10 years and older). The recruited sample size was ample in terms of power requirements.

Snowball Sampling

Recruitment and Age. Participants were recruited via word-of-mouth. Researchers contacted individuals known personally to them and requested that they forward a WhatsApp message (see Appendix G) that briefly outlined the proposed study to any of their contacts with children, aged 6 to 14 years. This message requested that any interested parents email the researchers directly if they would like to receive information regarding how to participate. Two interested participants did not fit within the age parameters set; however, they were allowed to participate because their ages fell within one month of the age limits (i.e., participant 53 was aged 5y11m and participant 6 was aged 15y1m).

Gender. Due to the sampling strategy employed, the gender of potential participants was unknown, and we could not plan to match for gender.

Inclusion and Exclusion Criteria

All participants were South African, aged between 5y11m and 15y1m, and spoke English as their home language. Verbal and social skills are strongly associated with ToM ability (Happé, 1995). Therefore, children were excluded from the study if they demonstrated marked learning delays, obtained a vocabulary subtest scaled score below 6, or had ever been diagnosed with a severe language impairment; autism spectrum disorder; or a behavioural disorder, including conduct disorder or oppositional defiant disorder.

Procedure

After receiving ethical approval, the recruitment WhatsApp message (Appendix G) was sent to individuals known personally to the researchers. Parents of children who were interested in participating emailed the researchers directly and were sent an information letter (see Appendix H) detailing a brief overview of the study, as well as Google Form links to the consent form (see Appendix I) and the child and teenage assent forms (see Appendix J & K, respectively). Parents who returned the above items were emailed Google form links to a demographic questionnaire (see Appendix L) and the QCAE (Reniers et al., 2011; Appendix D), as well as a link to schedule their child's participation in the study via the online scheduling platform: Setmore Appointments. The vocabulary and ToM tasks were conducted online and took place over two Zoom video calls to minimise fatigue. The first Zoom session commenced with a brief 5 min introduction between the researcher and participant. Following this, the researcher activated the screen share option on Zoom to conduct the WASI-II vocabulary test over approximately 10 min. The participant was then guided through a practise story item (see

Appendix M) before commencing the location change false-belief and strange story tasks (Appendix N & E, respectively), which took approximately 20 min. During the second session, the researcher administered the original lie-joke task (Appendix A), the revised lie-joke task (Appendix B), and the faux pas task (Appendix F) via the screen share option on Zoom. The tasks mentioned above took approximately 15 min, 12 min, and 10 min, respectively and the items from the original and revised lie-joke tasks were counterbalanced to control for order effects.

Measures

Demographic Questionnaire

The items on the general demographic questionnaire (Appendix L) related to demographic information (i.e., age, gender and home language) and clinical or medical information needed to determine if participants met the inclusion criteria.

Parent-Report Measure

Questionnaire of Cognitive and Affective Empathy (QCAE). This 31-item parent-report measure (Reiners et al., 2011; Appendix D) was implemented as one of four convergent validity measures, as research suggests that empathy and ToM are related constructs. ToM plays a vital role in the empathetic process by enabling an individual to infer another person's emotional state through the correct processing of contextual cues (Bird & Viding, 2014). The QCAE uses a 4-point Likert scale for responses, ranging from 'strongly disagree' (1) to 'strongly agree' (4), and consists of both cognitive and affective empathy subscales, which respectively assess the ability to understand and experience another's internal emotional state (Queirós et al., 2018). The QCAE demonstrates good factor structure, reliability (Queirós et al., 2018), validity and internal consistency (Reiners et al., 2011). Furthermore, a study (n = 92) found it to be a

reliable measure of overall empathy ($\propto = 0.95$) in a South African context (Louw, 2014). Research on the validity of the QCAE in South Africa is limited, but there is a current study by Pileggi et al. (2020) that supports its application in this context.

Child Task Measures

Vocabulary Measure. The Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Wechsler, 2011) measures cognitive abilities of individuals ranging from 6 to 89 years old. The WASI-II Vocabulary subtest (Appendix C) measures verbal comprehension and was implemented as one of the four measures of convergent validity, as the literature suggests that there is an association between verbal intelligence (VIQ) and ToM abilities (Happé, 1995; Osterhaus et al., 2020). The vocabulary subtest was also used to exclude participants with extremely poor verbal skills relative to age-appropriate expectations (i.e., a scaled score below 6). This test is administered individually and consists of 31 items that measure word knowledge, degree of language development, and formation of verbal concepts. Participants were asked to provide definitions of words read aloud to them from a list of increasing difficulty. The WASI-II demonstrates strong reliability and validity in a Western context (Hasson, 2015). Furthermore, a study by Ferret (2011) found that all four WASI-II subtests were suitable for English-speaking South African participants, with an advantaged quality of education, when original age-adjusted standardisation norms were used.

Theory of Mind Measures. The UCT Theory of Mind Battery (Hoogenhout & Malcolm-Smith, 2014) measures a wide range of ToM abilities in children and consists of four modules in ascending order of difficulty: early, basic, intermediate and advanced. Our study utilised four tasks from this battery. The location change false-belief task, from the basic module (see Appendix N), was used as a baseline to ensure all participants demonstrated basic ToM

understanding that people can have false beliefs. The following tasks were implemented to assess the convergent validity and effectiveness of the revised lie-joke measure: (1) the strange stories task from the intermediate module (Appendix E), which assessed participants' ability to interpret non-literal language correctly; (2) the original lie-joke task from the advanced module (Appendix A), which measured participants' understanding of sarcastic meaning and intent; and (3) the faux pas task from the advanced module (Appendix F), which evaluated participants' ability to identify social errors and misunderstandings (Kelly, 2013).

Researchers from the University of Cape Town (UCT) found that South African schoolchildren, aged 10 years and older, from the Western Cape, had difficulty comprehending the sarcasm in the lie-joke measure from the advanced module (Hoogenhout & Malcolm-Smith, 2014; Appendix A). Interestingly, these children generally perceived the sarcastic character to be lying as opposed to joking. This original lie-joke measure was scored out of 16 and consisted of four stories, each with two main characters and a third background character. The dialogue between the two main characters consisted of a child directing a sarcastic utterance towards an adult (i.e., their parent or teacher). We identified various factors in the measure that might have influenced the children's lack of sarcasm understanding. Firstly, informal questioning by the original research team suggested that the children answered items in the measure incorrectly because they were culturally inappropriate. A child uttering a sarcastic statement towards an adult was seen by them as disrespectful in the cultural context of the Western Cape (S. Malcolm-Smith, personal communication, February 19, 2020). Sarcastic statements are better understood when uttered by an adult, as adults verbalise non-literal language more often than children, and children are more aware of the communicative competence of adults (Banasik-Jemielniak & Bokus, 2019). Therefore, it is contextually understandable that participants in the previous study

perceived a child's sarcasm towards an adult as deception. Secondly, some vocabulary used was not relevant to a South African context, and we suspect the addition of a third character added unnecessary complexity and length to the measure. Difficulty comprehending an item can create the false impression of a lack of sarcasm understanding. Therefore, to ensure that a measure is only assessing sarcasm comprehension, items must contain contextually appropriate vocabulary and be controlled for length and complexity (Banasik-Jemielniak & Bokus, 2019).

Revised Lie-Joke Measure. We designed a revised lie-joke measure that takes the above-mentioned factors into account (Appendix B). We limited each story's length to 65 words (vs the previous stories of 165 words), reduced the number of characters from three to two, and aimed to use appropriate vocabulary and story settings for a South African context. All stories were accompanied by explicit facial expressions and sarcastic intonation (i.e., lengthened phrases and lowered pitch) to emphasise the non-literal meaning of sarcastic statements (Banasik-Jemielniak & Bokus, 2019). Items were counterbalanced to control for order effects. To control for participant structure effects on sarcasm comprehension, half of the stories included an adult speaking to a child, and the other half involved two children conversing. As a result of controlling for participant structure effects on sarcasm comprehension, the revised lie-joke measure consisted of eight story items and was thus scored out of 32. To ensure that scores from the original lie-joke measure (initially scored out of 16) were directly comparable to the revised measure, scores from the former were proportionally adjusted (i.e., doubled) to be scored out of 32.

Ethical Considerations

Ethical approval was obtained for The Moral Development Study from UCT's ethical review board on 5 March 2013 (see Appendix O). Ethical approval for this component of the

approved Moral Development Study was obtained from UCT's ethical review board on 23 July 2020 (see Appendix P).

Consent, Voluntary Participation, and Confidentiality

All individuals recruited via WhatsApp, who emailed the researchers to show their interest in participating in our study, were provided with an outline of the procedure, requirements, and the researcher's contact information for any questions regarding the study. Willing parents were required to sign a consent form stating their permission for their child to participate in the study. This consent form provided contact information for the Psychology Department Ethics Board and assured the parent and child that all data gathered throughout the study would remain confidential and protected. Furthermore, this form assured participants that their data would not be linked to any identifying information in reports pertaining to this study. If permission was granted for the child to participate in the study, the child received an assent form, simply stating what would be expected of them. If the child wished to participate, they were required to sign the assent form, stating that their parents had read the introductory letter and study outline to them and they understood that they were allowed to withdraw from the study at any time, without penalty.

Participation in this study carried no risk for the parents and carried a minimal risk of fatigue for the children. Both parties were informed that the children were allowed to take breaks or continue the assessment on a different day if they became fatigued.

Upon completion of the study, parents will receive a brief research report that provides general feedback regarding the research findings.

Statistical Analyses

RStudio (version 1.2.5033) was used for all statistical analyses, and statistical significance was determined with α set at .05. Effect size in correlational analyses was considered low if less than .40, moderate if between .40 and .70, and high if above .70 (Lachenicht, 2013). Effect size in comparison of group differences was considered small if less than .20, medium if .50, and large if above .80 (Lakens, 2013).

Preliminary Analyses

A complete set of descriptive statistics was compiled to ensure that the parametric statistical test assumptions were met. These preliminary analyses provided an initial summary of the participants' demographic characteristics, as well as their empathy, vocabulary, and ToM scores. When parametric statistical test assumptions were not met, appropriate non-parametric tests were implemented.

Psychometric Analysis

We determined the inter-rater reliability of the revised lie-joke measure, using Cohen's kappa, and calculated the split-half reliability of this measure using the Spearman-Brown prophecy formula. The convergent validity of the revised lie-joke measure was assessed by conducting bivariate correlational analyses (using Pearson's product-moment correlation coefficient) to establish the magnitude of the association between performance on the revised lie-joke measure and that on each of the four convergent measures.

Within and Between Group Differences

The effectiveness of the revised lie-joke measure was investigated by conducting a test of group differences that compared the sarcasm comprehension scores of children aged 10 years and older on the original and revised lie-joke measures. Furthermore, we investigated if there was an age difference in understanding of sarcasm in the revised lie-joke measure, by conducting

a test of group differences that compared the sarcasm comprehension scores of younger participants (aged 5 to 9 years old) with the scores of older participants (aged 10 years and older). Paired samples and independent samples t-tests were to be conducted for the above-mentioned comparison of means if data were normally distributed. Alternatively, if the data demonstrated a skewed distribution, non-parametric tests would allow for the comparison of medians.

Results

Sample Demographics

The final sample included N = 63 participants, with n = 33 younger participants in the age range between 5y11m and 9y11m, and n = 30 older participants in the age range between 10y0m and 15y1m. Purposive sampling resulted in a high-SES sample, with more male participants (n = 43) than female participants (n = 20). We obtained all required data for 62 out of the 63 participants, as we did not receive a response for the parent-report QCAE for one participant.

Preliminary Analyses

Participants completed the WASI-II Vocabulary subtest (Wechsler, 2011), location change false-belief, strange stories, lie-joke, and faux pas tasks from the UCT ToM Battery (Hoogenhout & Malcolm-Smith, 2014), as well as our revised lie-joke measure. All participants adequately understood the measures, indicated by the negatively skewed control data on all of the following tasks: location change false-belief (score range = 12, median = 12), strange stories (score range = 24, median = 24), original lie-joke (score range = 32, median = 32), revised lie-joke (score range = 32, median = 32) and faux pas (score range = 40, median = 38) tasks. The participants' ToM scores were relatively normally distributed in most tasks (see Table 1), except

for the slightly positively skewed distribution of the original lie-joke ToM scores (Figure 1; score range = 32; mean = 24.46; median = 24.0; SD = 2.08), and the negatively skewed distribution of the revised lie-joke ToM scores (Figure 2; score range = 32: mean = 28.53; median 30, SD = 3.61). Non-parametric tests were thus used in statistical analyses that involved the lie-joke data, as the median is a more appropriate indicator of central tendency than the mean when data are skewed. The QCAE parent-report data (score range = 100; mean = 84.5; SD = 13.5) and faux pas ToM scores (score range = 40; mean = 29.98; SD = 6.24) demonstrated wide variation.

Table 1

Descriptive Statistics of All Participants' Scores on Measures (Aged 6 to 15 years old)

Measures	Control		ToM	
	M (SD)	Median	M (SD)	Median
WASI-II Vocabulary T-score (max = 100)	61.84 (8.01)	62		
Parent-report QCAE * (max = 100)	84.5 (13.15)	84		
Location Change False-Belief (max = 12)	12.0 (0.00)	12.0	12.0 (0.00)	12.0
Strange Stories (max = 24)	23.67 (0.91)	24.0	22.49 (1.68)	22.5
Original Lie-Joke (max = 32)	31.94 (0.25)	32.0	24.46 (2.08)	24.0
Revised Lie-Joke (max = 32)	31.71 (1.36)	32.0	28.53 (3.61)	30.0
Faux Pas (max = 40)	37.5 (2.5)	38.0	29.98 (6.24)	30.5

Note. Means with (SD) are reported. * Only completed by 62n

Figure 1

Distribution of Original Lie-Joke ToM Scores
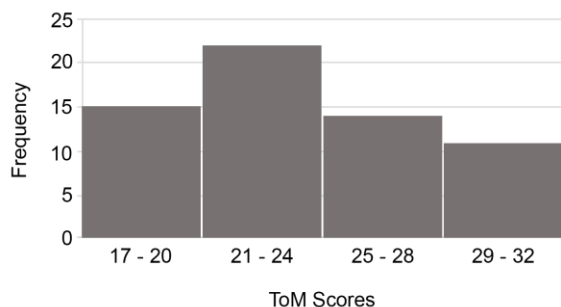
Figure 2

Distribution of Revised Lie-Joke ToM Scores
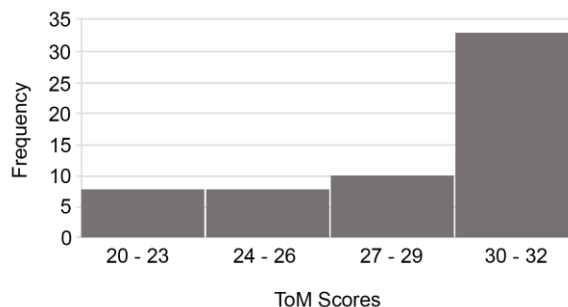
Psychometric Analysis

Reliability

Internal Consistency. One of the primary characteristics of a reliable measure is that it yields consistent measurement results (Guilford & Fruchter, 1978). We used the Spearman-Brown prophecy formula, an estimate of internal consistency, to calculate the split-half reliability of the revised lie-joke measure. We limited this computation to participants between the ages of 10 and 15 years old, as research suggests that children aged 10 years and older should be able to understand sarcasm (Recchia et al., 2010). Based on the Pearson correlation coefficient ($r = .94$, $p = .032$) we found that the revised lie-joke measure demonstrates good internal consistency.

Interrater Reliability. Two raters were involved in the scoring of the revised lie-joke measure. The Cohen's kappa findings suggest that overall, there was satisfactory agreement between the raters (Kappa $= .76$).

Validity

Convergent Validity. We investigated the convergent validity of the revised lie-joke measure by correlating it with the four convergent measures discussed previously, namely, the WASI-II Vocabulary subtest ($r = .45$, $p = .043$), strange stories task ($r = .51$, $p = .038$), faux pas task ($r = .52$, $p = .041$) and parent-report QCAE. We found significant moderate correlations

between the revised measure and the first three convergent measures mentioned above. We found only small, non-significant correlations between the revised lie-joke measure and the cognitive subscale ($r = .23$, $p = .057$), affective subscale ($r = .20$, $p = .055$) and overall QCAE scores ($r = .18$, $p = .053$).

Within and Between Group Differences

Comparison of Original vs Revised Lie-Joke ToM Scores in Older Children

A non-parametric test of group differences was used to investigate the hypothesis that children, aged 10 years and older, would demonstrate better sarcasm understanding in the new lie-joke measure when compared to the prior lie-joke measure. The above hypothesis was supported, indicated by a significant difference between the median ToM scores of the original and revised lie-joke measures (scaled original lie-joke median = 24; revised lie-joke median = 30; $V = 222$, $p < .001$). A large effect size was calculated (Cohen's $d = 1.04$), indicating a meaningful difference between groups.

Comparison of ToM Scores Across Age Groups on the Revised Lie-Joke Measure

A non-parametric test of group differences was used to investigate the hypothesis that there would be a significant difference in ToM scores across age bands for the revised lie-joke measure. The median score of younger participants, aged 5 to 9 years old (median = 27.5), and older participants, aged 10 to 15 years old (median = 32), were compared. The hypothesis was supported, indicated by a significant difference in ToM scores between the two age groups for the revised lie-joke measure ($W = 853$, $p < .001$). The meaningfulness of this finding was further supported by a large effect size (Cohen's $d = 1.37$).

Discussion

Our results indicate that the revised lie-joke measure evidenced good reliability, as well as convergent validity with three of the four indicators. Sarcasm understanding scores in children 10 years and older were significantly improved when compared to the initial task, and the expected age difference between older and younger children on sarcasm understanding was apparent.

Reliability of the Revised Lie-Joke Measure

As hypothesised, our revised lie-joke measure demonstrated strong internal consistency. This was evidenced by the high Pearson correlation coefficient that we found when calculating the split-half reliability of our measure. It is worth noting that one of the difficulties of using split-half reliability to estimate the internal consistency of a measure is that the calculation involves splitting the measure in half and thus undermines the overall internal consistency of the measure (Sijtsma, 2009). Hence, to ensure that the estimate we found and reported on was, in fact, accurate, we used the Spearman-Brown prophecy formula (yielding the Pearson correlation coefficient), which accounts for the splitting of the halves (Peters, 2014). We also found good inter-rater reliability on our revised lie-joke measure based on the kappa statistic of 0.76 that we obtained. A reliable measure yields consistent results across items and raters, and our revised measure met these requirements (Guilford & Fruchter, 1978).

Validity of the Revised Lie-Joke Measure

Regarding the validity of our revised lie-joke measure, we found reasonable indications of convergent validity. As was hypothesised, we found a moderate, positive correlation between performance on the WASI-II Vocabulary subtest and the revised lie-joke measure. The correspondence between higher scores on the WASI-II Vocabulary subtest and better performance on the revised lie-joke measure aligns with the literature, which documents a

positive correlation between ToM abilities and vocabulary skills (Angeleri & Airenti, 2014). In addition to this, we found a moderate, positive correlation between our revised lie-joke measure and the strange stories and the faux pas tasks, as was hypothesised. The importance of these correlations lies in the fact that, like the lie-joke measure, these measures assess more advanced ToM abilities seen in older children (Hoogenhout & Malcolm-Smith, 2014).

We found only a weak, positive correlation between participants' ToM scores on the revised lie-joke measure and their empathy scores for both the cognitive and affective empathy subscales of the QCAE. This was unexpected because empathy is reported to be positively correlated with ToM abilities (Bird & Viding, 2014). Accordingly, we expected to find a moderate, positive correlation between our revised lie-joke measure and the parent-report QCAE scores. There are various possible reasons for the weak correlation between the revised lie-joke measure and the parent-report QCAE scores. For one, perceptual bias may have influenced the parent's observations of their children. In other words, some parents may perceive their children to be more or less empathetic than they are. Thus, their reports, based on their subjective observations, may yield biased scores. Additionally, due to social desirability bias, parents may have given responses that they deemed to be more socially acceptable, to avoid embarrassment or to make their child "look good" (Fisher & Katz, 2000. pp.109). As a result, the parent report QCAE scores we obtained may have been a measure of the parents' socially desirable responses and not their actual perceptions of their children's empathic abilities. Both perceptual and social desirability biases affect the validity of scores yielded from a measure (Van de Mortel, 2008). Another consideration is that the QCAE was an indirect measure of the children's general empathic tendencies, whereas the other three converging measures, and the lie-joke measure itself, were direct assessments of their ToM abilities. Thus, the disparity between the directness

and indirectness in the assessment, as well as the specificity of ToM abilities versus the generality of dispositional empathic tendencies, may have contributed to the weak correlation we obtained (Longobardi et al., 2019). Finally, it may be that the reports made by parents were limited to their understanding of their child's empathy within the family context. One of the challenges of using parent reports is that parents primarily base their reports on the family setting, which leaves a measure of the child's empathy beyond this context unaccounted for (Rieffe et al., 2010). Hence, the ratings would be contextually limited (i.e., they would not provide a full picture of the children's empathic abilities). That being said, parent reports were the most practical measure to use, as children below the age of eight lack the verbal and cognitive abilities to report on their internal states and older children are highly influenced by demand characteristics (Dadds et al., 2008). Despite the weak correlation found between the revised lie-joke measure and the QCAE; our revised lie-joke measure demonstrated satisfactory indications of validity with the other three converging measures.

Effectiveness of the Revised Lie-Joke Measure

Comparison of Original vs Revised Lie-Joke ToM Scores in Older Children

The effectiveness of our revised measure was supported by a significant difference between the median ToM scores of the original and revised lie-joke measures, with a large effect size (Cohen's d = 1.04). The significantly improved ToM scores of children aged 10 years and older in the revised lie-joke measure aligns with the existing literature, which indicates that children should be able to discern the teasing or mocking intent behind a sarcastic comment by 10 years old (Recchia et al., 2010). We considered multiple factors from prior research findings in order to design an effective lie-joke measure for the South African context. First, we ensured that the story vignettes were embedded within culturally familiar scenarios (Angeleri & Airenti,

2014). Second, we considered the effect of the relationship between interlocutors on sarcasm understanding and ensured that the story dialogue did not involve children being sarcastic towards adults (Banasik-Jemielniak & Bokus, 2019). Finally, we ensured that narrative length, vocabulary, syntax, and the number of characters did not place additional demands on the participants' memory and comprehension skills (Banasik-Jemielniak & Bokus, 2019; Creusere, 1999). The results indicate that our measure was relevant to the sample assessed in a South African context and the adjustments that we made to the length, complexity and participant structure of the stories were effective. However, it is beyond the scope of this study to discern which of the many changes we made were most influential to the participants' increased sarcasm understanding. Thus, we cannot make assumptions about the separate influence of the adjustments mentioned above and can only conclude that our changes were successful overall.

Although it did not affect the significance of our results, the researchers observed that some non-verbal cues in the revised lie-joke measure (Appendix B) were ambiguous and confused a minority of participants during the story sessions. First, several participants seemed unsure if the character in story one had kicked the ball after the researcher narrated "Tim runs to kick the ball but misses it and slips in the mud" and then immediately asked, "Did Tim kick the ball?". Most participants answered correctly that Tim had not kicked the ball but some of these participants hesitated, which indicated that they were unsure of their answer. More importantly, seven participants incorrectly answered "Yes", stating that Tim had kicked the ball. The participants' confusion prompted the researchers to informally question some children regarding why they appeared uncertain about whether or not Tim had kicked the ball. This questioning indicated that the first illustration in story one was confusing because the story specifically stated that Tim had missed the kick; however, the picture allowed for the interpretation that he had

kicked the ball because the soccer ball was flying through the air (see figure 3). Second, 12 participants incorrectly interpreted a white lie as sarcasm in story four, where a character states that some burnt muffins "smell delicious" so as not to offend another character who had made them. After several participants had incorrectly interpreted this story, the researchers were prompted to informally question some participants about why they thought the girl was being sarcastic when she said, "Yeah sure… they smell delicious". This informal questioning led the researchers to understand that the female character's sideways glance was ambiguous and could be interpreted as either a sarcastic eye roll or a deceptive glance (see figure 4).

Figure 3

Figure 4

Such questioning arose spontaneously and could not be planned for because the researchers did not expect the participants to find these questions challenging. However, by allowing for such informal questioning, we were able to recognise where there was a contradiction between the verbal and non-verbal cues in our revised lie-joke task. The participants who were informally questioned described using both intonation and contextual cues to determine if the characters in the stories were lying or joking. This aligns with the literature, which indicates the importance of both verbal and non-verbal cues in identifying sarcasm (Capelli et al., 1990; Winner et al., 1987). These participants' nuanced interpretations of non-verbal cues indicate the importance of clear illustrations, as ambiguous pictures may override

clear verbal cues and cause children to misinterpret a character's statement. Overall, our results supported the hypothesis that the participants would demonstrate better sarcasm understanding of the revised lie-joke measure when compared to the prior lie-joke measure. Thus, we can conclude that the revised lie-joke measure is an effective task for assessing sarcasm understanding in high-SES children in a South African context.

Comparison of ToM Scores Across Age Groups on the Revised Lie-Joke Measure

A significant difference in ToM scores was found between the median scores of children aged 10 years and older and those of children younger than 10 years old, with a large effect size (Cohen's d = 1.37): The older children performed much better than the younger age group. The significant difference in sarcasm understanding across the age groups in the revised lie-joke measure aligns with current research, which suggests that sarcasm understanding develops across middle childhood, with older children demonstrating better sarcasm comprehension than younger children (Ackerman, 1983; Angeleri & Airenti, 2014; Cutting & Dunn, 1999; Pexman, 2018; Recchia et al., 2010). Many younger participants demonstrated a developing sense of sarcasm and first-order reasoning skills in their identification of a sarcastic utterance as teasing; however, they consistently incorrectly identified this 'teasing' as deception. For example, when participants were asked the open-ended question, "Why did Sarah's father say that her room was tidy?" in story 5 of the revised lie-joke measure (Appendix B), the younger participants often stated that the father was teasing (e.g., "Sarah would know that dad was teasing. She knew that the room was not looking nice"; participant 26; age 6). However, when asked the closed-ended question, "Was Sarah's dad lying, joking, or telling the truth?", the younger participants all chose "lying" because, as research suggests, what was said contradicted the reality of the shared situation and young children do not possess the second-order reasoning necessary to infer the

speaker's beliefs about the listener's beliefs (Angeleri & Airenti, 2014). As was hypothesised, the revised lie-joke measure demonstrated a significant difference in sarcasm understanding between age groups, which indicates that this measure successfully reflects developmentally appropriate norms in sarcasm comprehension (Ackerman, 1983; Angeleri & Airenti, 2014; Cutting & Dunn, 1999; Pexman, 2018; Recchia et al., 2010).

Successful social interactions depend on an individual's ability to understand non-literal language, which is based on one's ability to correctly interpret verbal and non-verbal cues (Attardo et al., 2003; Capelli et al., 1990; Happé et al., 2017). ToM abilities are critical in the interpretation of such cues, which allow individuals to infer others' mental states and grasp complex forms of communication, including discerning between the subtle nuances of deception and sarcasm (Happé, 1994; Lagattuta et al., 2015). The misinterpretation of others' intentions in non-literal speech can create tensions and negative impressions within relationships and social interactions (Happé et al., 2017). Thus, an effective lie-joke task that accurately measures sarcasm understanding is a necessary tool for identifying ToM deficits in middle childhood; after which language complexities make social relationships increasingly difficult to navigate.

Limitations and Future Directions

There are a number of limitations to our study that we must acknowledge. To start with, our sample consisted of only high-SES, English-speaking children, which excludes children of low-SES and many South African children who do not speak English. Thus, the most significant limitation of our study was that of selection bias, as our sample was not representative of all South African children, making it difficult to generalise our findings to the broader population. This selection bias was unavoidable due to the lockdown restrictions present in South Africa this year, which prevented us from being able to access a low-SES sample through local schools as

initially planned. Such limitations led us to use online platforms as our next best alternative, with an available sample of high-SES children who had unlimited internet access. As a result, findings from our study cannot be generalised to other South African contexts, and further research into the applicability of the revised lie-joke measure for children who are from low-SES backgrounds or are non-English speaking is necessary. We suggest that future studies investigate sarcasm understanding in low-SES, English-speaking children first, and following this, the measure can then be translated into other South African languages such as isiXhosa and Afrikaans. We also recommend the use of a larger sample size so that inferences can be made to the broader population.

Another limitation was that our study was conducted online, and even though we communicated the importance of the participants completing the tasks independently, the researchers could not control for any silent cues from parents, beyond the view of the camera, or potential 'coaching' of children on sarcasm before the online sessions. Furthermore, snowball sampling resulted in siblings being recruited for the study. All attempts were made to interview siblings straight after one-another; however, this was not always possible, and hypothetically, siblings could have discussed the stories and answers with one another if they completed the story tasks on different days. Nevertheless, children's ability to remember their answers was likely undermined due to the great number of stories completed in a session and all parents were cooperative in keeping the children isolated during their story sessions. Ideally, this study should have been conducted in a quiet room at school, where there is minimal pressure from parents for their child to perform and no potential for discussion about the stories between siblings or peers. That being said, as already stated, online platforms were used out of necessity as non-essential contact was limited throughout this year. The above limitations are not in-built to our measure,

as all materials used in our study can be delivered and administered in person. Thus, our measure is accessible to children who do not have access to the online platforms and technology used in this study, allowing for a more diverse sample of children in future applications of the revised lie-joke measure.

Finally, our study required the participants to answer questions based on vignettes, which cannot capture the complexity of real-life interactions. The literature indicates that children demonstrate better sarcasm comprehension in real-life settings when compared to studies that use stories (Angeleri & Airenti, 2014). This factor undermines the ecological validity of our findings, as our participants' performance on the revised lie-joke measure may not necessarily reflect their understanding of sarcasm in real life. However, this is a general problem regarding ToM assessments. While they have contributed significantly to our understanding of mental state reasoning, they are limited in their artificiality (Gallotti & Frith, 2013). Thus, in recent years, many researchers have recommended that methods used to investigate socio-cognitive skills, including ToM abilities, should shift from mere reasoning about vignettes to more real settings in which participants engage with another person. Central to this argument is that second-person perceptual tasks are fundamentally different from real-life interactions with people (Quesque & Rossetti, 2020).

Conclusion

This study has demonstrated that first, our revised lie-joke measure is reliable, as it demonstrates good internal consistency and inter-rater reliability. Second, except for the QCAE, our revised lie-joke measure correlated moderately with the WASI-II Vocabulary subtest as well as the strange stories and faux pas tasks from the UCT Theory of Mind Battery, suggesting reasonable convergent validity. Third, participants aged 10 years and older demonstrated better

sarcasm understanding in the revised lie-joke measure when compared to the original lie-joke measure, which indicates its effectiveness as a measure of age-appropriate sarcasm comprehension. Finally, the revised lie-joke measure indicated a significant difference in sarcasm understanding between children younger than 10 years old and children aged 10 years and older, with the latter performing significantly better than the former. Informal follow-up with participants suggested that some necessary illustration edits should be implemented. Overall, we can conclude that the revised lie-joke measure is an effective task for assessing sarcasm understanding in middle childhood and appears to be appropriate for a South African context. Following the success of this measure in a high-SES, English-speaking sample, the revised lie-joke measure needs to be implemented with low-SES children to determine its reliability and validity in a sample that is more representative of the broader population's socioeconomic context.

Acknowledgements

References

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, 35(3), 487–508. https://doi.org/10.1016/0022-0965(83)90023-1

Angeleri, R., & Airenti, G. (2014). The development of joke and irony understanding: A study with 3- to 6-year-old children. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 68(2), 133–146. https://doi.org/10.1037/cep0000011

Attardo, S., Eisterhold, J., Hay, J. & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Journal of Humor Research,* 16(2), 243-260. https://doi.org/10.1515/humr.2003.012

Bailey, K. & Im-Bolter, N. (2020). Theory of mind and language in childhood epilepsy. *Mind, Brain, and Education*. 14(2), 146-154. https://doi.org/10.1111/mbe.12230

Banasik-Jemielniak, N., & Bokus, B. (2019). Children's comprehension of irony: Studies on Polish-speaking preschoolers. *Journal of Psycholinguistic Research*. 48, 1217-1240. https://doi.org/10.1007/s10936-019-09654-x

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*. 47, 520-532. https://doi.org/10.1016/j.neubiorev.2014.09.021.

Capelli, C. A., Nakagawa, N., & Madden, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development*, *61*(6), 1824–1841. https://doi.org/10.2307/1130840

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America.* 126(3), 1394. https://doi.org/10.1121/1.3177275

Creusere, M. A. (1999). Theories of adults' understanding and use of irony and sarcasm: Applications to and evidence from research with children. *Developmental Review*, *19*(2), 213–262. https://doi.org/10.1006/drev.1998.0474

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development*, *70*(4), 853–865. https://doi.org/10.1111/1467-8624.00061

Dadds, M. R., Hunter, K., Hawes, D. J., Frost, A. D. J., Vassallo, S., Bunn, P., Merz, S., & El Masry, Y. (2008). A measure of cognitive and affective empathy in children using parent ratings. *Child Psychiatry and Human Development*, *39*(2), 111-122. https://doi.org/10.1007/s10578-007-0075-4

Demorest, A., Meyer, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development*, *55*(4), 1527–1534. https://doi.org/10.2307/1130022

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., McGovern, A., Qualter, A., & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development.*, *67*(6), 3071–3085. https://doi.org/10.1111/j.1467-8624.1996.tb01903.x

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferret, H. L. (2011). *The adaptation and norming of selected psychometric tests for 12- to 15-year-old urbanized Western Cape adolescents* [Unpublished doctoral dissertation, University of Stellenbosch]. scholar.sun.ac.za › handle › ferrett_adaptation_2011

Filik, R., Țurcan, A., Ralph-Nearman, C., & Pitiot, A. (2019). What is the difference between irony and sarcasm? An fMRI study. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *115*, 112–122. https://doi.org/10.1016/j.cortex.2019.01.025

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing*, *17*(2), 105-120. https://doi.org/10.1002/(SICI)1520-6793(200002)17:2%3C105::AID-MAR3%3E3.0.CO;2-9

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160-165. https://doi.org/10.1016/j.tics.2013.02.002

Glenwright, M., Parackel, J. M, Cheung, K. R. J., & Nilsen, E. S. (2014). Intonation influences how children and adults interpret sarcasm. *Journal of Child Language*, *41*(2), 472–484. https://doi.org/10.1017/S0305000912000773

Gopnik, A. & Wellman, H.M. (1992). Why the child's theory of mind really is a theory. *Mind and Language, 7*(1-2), 145-171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

Grice, H. P. (1975). Logic and conversation. In *Speech acts*, 41-58. Brill. https://doi.org/10.1163/9789004368811_003

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). McGraw-Hill.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story

characters' thoughts and feelings by able autistic, mentally handicapped, and normal

children and adults. *Journal of Autism and Developmental Disorders*, 24(2), 129–154.

https://doi.org/10.1007/BF02172093

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task

performance of subjects with autism. *Child Development*, 66(3), 843-855.

https://doi.org/10.1111/j.1467-8624.1995.tb00909.x

Happé, F. G. E., Cook, J. L., & Bird, G. (2017). The structure of social cognition:

In(ter)dependence of socio-cognitive processes. *Annual Review of Psychology, 68*,

243-267. https://doi.org/10.1146/annurev-psych-010416-044046

Hoogenhout, M., & Malcolm-Smith, S. (2014). Theory of mind in autism spectrum

disorder: Does DSM classification predict development? *Research in Autism Spectrum

Disorders*, 8(6), 597-607. https://doi.org/10.1016/j.rasd.2014.02.005

Hasson, R. (2015). *The clinical utility of the WASI-II and its association with acculturation

among Arab-American adolescent males*. [Unpublished doctoral dissertation, Michigan

State University].

https://d.lib.msu.edu/islandora/object/etd:2843/datastream/OBJ/download/The_Clinical_

Utility_of_the_WASIII_and_its_association_with_acculturation_among_Arab_American

_Adolescent_Males.pdf.

Hughes, C., Devine, R. T., Ensor, R., Koyasu, M., Mizokawa, A., & Lecce, S. (2014). Lost

in translation? Comparing British, Japanese, and Italian children's theory-of-mind

Performance. *Child Development Research*, *10*, 1-11.

http://dx.doi.org/10.1155/2014/893492

Im-Bolter, N., Agostino, A., & Owens-Jaffray, K. (2016). Theory of mind in middle childhood and early adolescence: Different from before? *Journal of Experimental Child Psychology*, 149, 98–115. https://doi.org/10.1016/j.jecp.2015.12.006

Kelly, T. (2013). *Gender difference in theory of mind development*. [Unpublished master's thesis, University of Cape Town]. http://www.psychology.uct.ac.za/sites/default/files/image_tool/images/117/Logos/thesis/ Taryn%20Kelly.pdf

Lachenicht, L. (2013). Correlation. In C. G. Tredoux & K. Durrheim (Eds.), *Numbers, Hypotheses & Conclusions* (pp. 181–200). Juta & Company Ltd.

Lagattuta, K. H., Kramer, H. J., Kennedy, K., Hjortsvang, K., Goldfarb, D., & Tashjian, S. (2015). Chapter six - Beyond Sally's missing marble: Further development in children's understanding of mind and emotion in middle childhood. *Advances in Child Development and Behavior*, 48, 185-217. https://doi.org/10.1016/bs.acdb.2014.11.005

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Longobardi, E., Spataro, P., & Rossi-Arnaud, C. (2019). Direct and indirect associations of empathy, theory of mind, and language with prosocial behavior: Gender differences in primary school children. *The Journal of Genetic Psychology*, *180*(6), 266-279. https://doi.org/10.1080/00221325.2019.1653817

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech-Language Pathology*, 15(2), 142-154. https://doi.org/10.1044/1058-0360(2006/014)

Osterhaus, C., Putnick, D. L., Kristen-Antonow, S., Kloo, D., Bornstein, M. H. and Sodian, B. (2020), Theory of mind and diverse intelligences in 4-year-olds: Modelling associations of false beliefs with children's numerate-spatial, verbal, and social intelligence. *British Journal of Developmental Psychology*, 38(4), 580-593. https://doi.org/10.1111/bjdp.12336

Perner, J., Kain, W., & Barchfeld, P. (2002). Executive control and higher-order theory of mind in children at risk of ADHD. *Infant and Child Development*, 11(2), 141-158. https://doi.org/10.1002/icd.302

Peters, G-J. Y. (2014). The alpha and the omega of reliability and validity: why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist, 16*(2), 56- 69. https://doi.org/10.31234/osf.io/h47fv

Pexman, P. M. (2018). How do we understand sarcasm? *Frontiers for Young Minds*, *16*, 1-8. https://doi.org/10.3389/frym.2018.00056

Pileggi, L., Lewis, R., & Malcolm-Smith, S (2020). *Measuring dispositional empathy in South African children: Psychometric properties of two well-known questionnaires in a low-to-middle income country.* [Manuscript submitted for publication]. Department of Psychology. University of Cape Town.

Queirós, A., Fernandes, E., Reiners., R., Sampaio, A. Coutinho, J., & Seara-Cardoso, A. (2018). Psychometric properties of the questionnaire of cognitive and affective empathy in a Portuguese sample. *PloS ONE,* 13(6), 1-14. https://doi.org/10.1371/journal.pone.0197755

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%2F1745691619896607

Recchia, H. E., Howe, N., Ross, H. S., & Alexander, S. (2010). Children's understanding

and production of verbal irony in family conversations. *British Journal of Developmental*

Psychology, *28*(2), 255–274. https://doi.org/10.1348/026151008X401903

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The

QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality*

Assessment, 93(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Rieffe, C., Ketelaar, L., & Wiefferink, C. H. (2010). Assessing empathy in young children:

Construction and validation of an empathy questionnaire (EmQue). *Personality and*

Individual Differences, *49*(5), 362-367. https://doi.org/10.1016/j.paid.2010.03.046

Rockwell, P. (2001). Facial expression and sarcasm. *Perceptual and Motor Skills*, 93(1), 47–

50. https://doi.org/10.2466/pms.2001.93.1.47

Rockwell, P., & Theriot, E. M. (2001). Culture, gender, and gender mix in encoders of

sarcasm: A self-assessment analysis. *Communication Research Reports, 18*(1), 44-

52. https://doi.org/10.1080/08824090109384781

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A

review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141.

https://doi.org/10.1037/0033-2909.115.1.102

Shahaeian, A. (2015). Sibling, family, and social influences on children's theory of

mind understanding: New evidence from diverse intracultural samples. *Journal of Cross-*

Cultural Psychology, 46(6), 1-16. https://doi.org/10.1177/0022022115583897

Shahaeian, A., Nielsen, M., Peterson, C. C., & Slaughter, V. (2014). Cultural and family

influences on children's theory of mind development: A comparison of Australian and

Iranian school-age children. *Journal of Cross-Cultural Psychology*, *45*(4), 555–568.

https://doi.org/10.1177/0022022113513921

Shamay-Tsoory, S. G., Tomer, R., & Aharon-Peretz, J. (2005). The neuroanatomical basis

of understanding sarcasm and its relationship to social cognition. *Neuropsychology*,

19(3), 288-300. https://doi.org/10.1037/0894-4105.19.3.288

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's

alpha. *Psychometrika*, 74, 107–120. https://doi.org/10.1007/s11336-008-9101-0

van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report

research. *The Australian Journal of Advanced Nursing,* 25(4), 40-48.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II)*

[Database record]. APA PsycTests. https://doi.org/10.1037/t15171-000

Wellman, H. M., Cross, D. and Watson, J. (2001), Meta-analysis of theory-of-mind

development: The truth about false belief. *Child Development*, 72: 655-684.

https://doi.org/10.1111/1467-8624.00304

Western Cape Government. (2017). *Socio-economic profile: City of Cape Town*.

Winner, E., Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., & Gardner, H. (1987).

Making sense of literal and nonliteral falsehood. *Metaphor & Symbolic Activity*,

2(1), 13–32. https://doi.org/10.1207/s15327868ms0201_2