JSLHR

Research Article

Idiom, Syntax, and Advanced Theory of Mind Abilities in Children With Autism Spectrum Disorders

Elisabeth M. Whyte,[a] Keith E. Nelson,[a] and K. Suzanne Scherf[a]

Purpose: When researchers investigate figurative language abilities (including idioms) in children with autism spectrum disorder (ASD), syntax abilities may be more important than once considered. In addition, there are limitations to the overreliance on false-belief tasks to measure theory of mind (TOM) abilities. In the current study, the authors investigated idiom, syntax, and advanced TOM abilities in children with ASD compared to children with typical development (TD).
Method: Twenty-six children with ASD, ages 5 to 12 years, were compared to individuals in each of 2 control groups of children with TD: 1 matched on chronological age and nonverbal IQ, and 1 matched on syntax age-equivalence and raw scores. Idiom comprehension, syntax, vocabulary, and 2 measures of advanced TOM abilities were examined.
Results: Although children with ASD performed worse on idiom comprehension compared to the age-matched group with TD, they exhibited comparable idiom performance to the syntax-matched group with TD. Advanced TOM abilities were related to idiom comprehension for children with ASD, but not for children with TD, above the contributions of basic language abilities.
Conclusion: Syntax abilities should be used as a matching variable when examining figurative or other late-developing language skills.

Key Words: autism, language, developmental disorders, figurative, social communication

Children with autism spectrum disorders (ASD) are characterized as having deficits in figurative language, including difficulties in understanding metaphors, idioms, and humor (Dennis, Lazenby, & Lockyer, 2001; Happé, 1993; Kerbel & Grunwell, 1998a, 1998b; Lyons & Fitzgerald, 2004). This is particularly concerning in that misinterpretations of figurative language may impact children's classroom learning because teachers frequently use figurative phrases to illustrate concepts during their lessons (Kerbel & Grunwell, 1997; Lazar, Warr-Leeper, Nicholson, & Johnson, 1989). Idioms provide an interesting opportunity to examine the nature of figurative language deficits in children with ASD because the relatively fixed nature of many expressions (e.g., *raining cats and dogs,* not related patterns such as *raining cars and trucks*) mean that they are potentially easy to directly teach in interventions (Abrahamsen & Smith, 2000; Ezell & Goldstein, 1992; Lundblom & Woods, 2012; Whyte, Nelson, & Khan, 2013).

There are several theories positing why children with ASD fall behind in their figurative language abilities, but two have particular relevance to the current study. First, the *relevance theory* suggests that figurative language requires the listener to have some understanding of the speaker's intentions (Happé, 1993). Thus, perspective-taking abilities, and theory of mind (TOM) in particular should be strongly tied to figurative language abilities (Happé, 1993). In contrast, a second theory emphasizes the role of basic language abilities in understanding the linguistic context and suggests that vocabulary and syntax are the most important predictors of success in understanding idioms (Norbury, 2004). For example, Gernsbacher and Pripas-Kapit (2012) criticized the majority of research on both figurative language and TOM abilities in individuals with ASD for failing to measure or control for potential limitations in understanding syntax. The goal of the current study was to evaluate the potential contribution of TOM abilities and more basic language abilities (i.e., syntax) to differences for idiom comprehension in children with ASD and two comparison groups of children with typical development (TD): one matched on age and nonverbal IQ and the other matched on syntax abilities.

[a]The Pennsylvania State University, University Park
Correspondence to Elisabeth M. Whyte: emv131@psu.edu

Editor: Janna Oetting
Associate Editor: Elizabeth Crais
Received September 25, 2012
Revision received January 11, 2013
Accepted April 25, 2013
DOI: 10.1044/1092-4388(2013/12-0308)

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Linguistic Theories of Figurative Language Development

For individuals with TD, idiom comprehension develops slowly across childhood and into adolescence (Ackerman, 1982; Qualls, O'Brien, Blood, & Hammer, 2003; Spector, 1996). The theories of idiom comprehension in individuals with TD have largely focused on the linguistic aspects of these figurative phrases, including how some idiom items differ from others (such as familiarity and decompositionality). These linguistic theories of idiom comprehension share the idea that "acquisition of idioms occurs as part of the general process of language and word knowledge development" (Levorato & Cacciari, 1995, p. 261). For example, some research emphasizes the role of familiarity and repeated exposure to the phrases for increasing comprehension (Ackerman, 1982; Nippold & Taylor, 2002). The implication of the role of familiarity is that low-familiarity idioms are less likely to be fully lexicalized (or memorized) compared to high-familiarity idioms.

In addition, these theories of idiom comprehension emphasize the potential importance of the individual words for aiding the interpretation of an idiomatic phrase (Hamblin & Gibbs, 1999; Nippold & Duthie, 2003). The *decompositionality* of an idiomatic phrase refers to how closely the figurative meaning is related to the individual words in the phrase (Hamblin & Gibbs, 1999; Nippold & Duthie, 2003). Research on decompositionality predicts that highly decompositional idioms such as *saved my skin* (meaning "saved someone from getting hurt") will be more easily learned than idioms with low decompositionality, such as *hit the sack* (which means "go to bed"). It is interesting to note that the literature is inconsistent concerning whether children (with either ASD or TD) use the individual words from idiomatic phrases to help them understand the figurative meaning (Abrahamsen & Burke-Williams, 2004; Nippold & Duthie, 2003; Nippold & Rudzinski, 1993; Norbury, 2004; Whyte et al., 2013).

Consistent with the linguistic theories of figurative language development, delays in syntax and vocabulary have been implicated as a possible source of the delays in figurative abilities for children with ASD (Norbury, 2004, 2005). Gernsbacher and Pripas-Kapit (2012) stated that if individuals with ASD "don't have difficulty comprehending language in general, they don't have difficulty comprehending metaphoric language in particular" (p. 94). However, few studies examining figurative language comprehension have controlled for vocabulary, and even fewer have controlled for syntax when matching groups. Gernsbacher and Pripas-Kapit (2012) further argued that vocabulary, the most commonly used language variable for matching groups, may be a poor matching variable to control for structural language abilities. It is possible that children with ASD could perform well on measures of vocabulary but still have impairments in syntax, and this could lead to making biased conclusions about figurative impairments in ASD (Gernsbacher & Pripas-Kapit, 2012).

Evidence supporting the importance of syntax abilities in comprehending idioms comes from two reports that children with autistic social symptoms who have high language comprehension abilities may not have impairments in understanding idioms or metaphors (Norbury, 2004, 2005). Norbury (2004) found that children and adolescents with ASD, ages 8 to 15, showed impairments only in idiom comprehension compared to controls with TD if they had structural language impairments, but not when their structural language abilities (vocabulary and syntax) fell in the typical range. Overall, the results of regression analyses from Norbury (2004) suggest that syntax and vocabulary abilities may be better predictors of figurative language comprehension than the presence of autistic social impairments or performance on false-belief measures of TOM. However, research finds that false-belief understanding and basic language abilities (vocabulary and syntax) are strongly tied together in early development; thus, TOM abilities more generally may still be related to the development of figurative language.

The Relationship Between TOM and Language Abilities

The literature on the relationship between TOM and figurative language has had mixed results. Support for the relevance theory came from findings that children with TD as well as those with ASD are more likely to understand figurative language (idioms and metaphors) if they perform well on TOM tasks, mainly the classic tests of reasoning about first- and second-order false beliefs (Caillies & Le Sourn-Bissaoui, 2008, 2012; Happé, 1993; Martin & McDonald, 2004). However, other studies have failed to support this relationship (Norbury, 2004, 2005). Gernsbacher and Pripas-Kapit (2012) argue that a fundamental flaw in the majority of existing studies of figurative language in ASD stems largely from an overreliance on these false belief tasks to measure TOM abilities. Studies of the relationship between false belief and the structural aspects of language suggest that this commonly used task to measure TOM abilities is strongly dependent on very basic aspects of language development that may be delayed or impaired in individuals with ASD (Astington & Jenkins, 1999; Hale & Tager-Flusberg, 2003; Milligan, Astington, & Dack, 2007; Tager-Flusberg, 2007; Tager-Flusberg & Joseph, 2005). In addition, some preschool children are able to pass both first- and second-order false-belief tasks, so these measures may not be developmentally sensitive to the continued progression of TOM development across older childhood and into adolescence (Steele, Joseph, & Tager-Flusberg, 2003; Sullivan, Zaitchik, & Tager-Flusberg, 1994; Wellman, Cross, & Watson, 2001).

There are several tasks that assess advanced TOM abilities, which may be more developmentally appropriate for older children, including the "strange stories" task (Happé, 1994; O'Hare, Bremner, Nash, Happé, & Pettigrew, 2009) and the children's version of the "reading the mind in the eyes" (RMTE) task (Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001). The strange stories task requires children to interpret the intentions and mental states (lies, persuasion, forgetting, etc.) of characters in short stories composed of multiple sentences. The task now includes a revised coding scheme that rewards credit for providing partial descriptions of mental state responses (O'Hare et al., 2009). The children's version of

the RMTE requires participants to match words (*jealous*) or short phrases (*thinking about something*) describing complex mental states with facial expressions from pictures of the eye regions of faces (Baron-Cohen et al., 2001). The RMTE task may have lower demands on complex language processing than the strange stories task because it presents only words or phrases, rather than requiring subjects to interpret stories comprised of multiple sentences. Both the strange stories and the RMTE tasks show increases with age for children with TD, and they are developmentally appropriate for use with children with both ASD and TD who are ages 5 to 12 years (Baron-Cohen et al., 2001; O'Hare et al., 2009).

Previous studies have found deficits on the strange stories and RMTE tasks for individuals with ASD relative to individuals with TD (Baron-Cohen et al., 2001; Jollife & Baron-Cohen, 1999; Kaland, Callesen, Moller-Nielsen, Mortensen, & Smith, 2008; White, Hill, Happé, & Frith, 2009). However, several previous studies of these advanced TOM measures for individuals with ASD have had the same weakness related to group matching (failing to properly match groups on language and, especially, syntax abilities). This leaves open the question of whether diagnosis or language abilities best account for the group differences on these advanced TOM tasks.

Current Study

The current research examines idiom comprehension abilities in a group of high-functioning children with ASD compared to two matched control groups of children with TD, one matched on chronological age and nonverbal IQ, the other matched on syntax age-equivalence and raw scores. In other words, each child with ASD was yoked to two individual children with TD, one who was of similar chronological age and nonverbal IQ and another who had similar syntactic knowledge (based on syntax age-equivalence and raw scores), but was likely a bit younger in chronological age.

We hypothesized that the children with ASD would perform poorly on the idiom comprehension task when compared to the age and nonverbal IQ-matched children with TD. However, we also predicted that children with ASD would not show deficient figurative language abilities when compared to the syntax-matched group, consistent with linguistic theory. The current study also examined idiom performance in relation to advanced TOM abilities, as measured by both the RMTE and strange stories tasks. The linguistic theory would predict that there should not be a significant relationship between idiom comprehension and advanced TOM abilities after controlling for basic language (syntax and vocabulary) abilities. However, the relevance theory would predict that TOM abilities would still correlate with idiom comprehension after controlling for basic language abilities.

Method

Design and Participants

The current study included highly verbal children with ASD and children with TD. All of the children were native English speakers. The sample was largely composed of middle-class families from central Pennsylvania. Parent report indicated that 64 of the 72 children were non-Hispanic and Caucasian. The other parents classified their child's ethnicity as Hispanic ($n = 6$) or their race as American Indian ($n = 2$), African American ($n = 1$), or biracial ($n = 5$).

Individual children with TD were yoked to children in the ASD group based either on chronological age and nonverbal IQ or on syntax age-equivalence and raw scores from a larger sample of children with TD. Participants were excluded from the typical control groups if they had a diagnosis of any developmental disorder or language impairment. Participants were recruited using flyers distributed through several participant recruitment databases. Additional children with ASD were recruited using flyers distributed at schools and social skills intervention programs. Table 1 provides the full demographic information for the three groups.

Group with ASD. The group with ASD consisted of 26 children (21 boys, 5 girls), ages 5 to 12 years ($M = 9.08$, $SD = 1.87$). Children were included in the ASD group based on parental report of a previous diagnosis of an autism spectrum disorder. Parents also completed the Social Responsiveness Scale (SRS; Constantino, 2002), which measures autistic-like social symptoms in a child during the prior 6 months. A score of 60 or greater on this screening measure is highly associated with an actual diagnosis of a disorder on the autism spectrum (Constantino, 2002). In this sample, scores on the SRS ranged from 60 to 153 for the children in the ASD group (see Table 1).

The children with ASD had to be highly verbal (able to speak in sentences) to complete the tasks. As a result, the majority of the participants with ASD in the current study were high functioning. Parent report indicated that six of the children were previously diagnosed with autism, 13 with Asperger's syndrome, five with pervasive deficit disorder, not otherwise specified (PDD-NOS), and two with high-functioning autism. However, many of the children with ASD still had delays in their language abilities, with standardized syntax scores ranging from 51 to 104 ($M = 82.9$) based on the Syntax Construction subtest of the Comprehensive Assessment of Spoken Language (CASL, Carrow-Woolfolk, 1999). Parent report indicated that 17 of the children were currently enrolled in speech therapy or were enrolled in speech therapy in the last school year. Four additional children had a reported history of speech therapy ending prior to the last school year. The most common comorbid/secondary diagnoses included attention-deficit hyperactivity disorder (ADHD; $n = 7$), oppositional defiant disorder (ODD; $n = 3$), and anxiety ($n = 3$).

Chronological age-matched control group (CAM). The CAM group included 26 children with TD (18 boys, eight girls), ages 5 to 12 years ($M = 9.03$, $SD = 1.95$). Parent report indicated that the children with TD were free of a diagnosis of ASD or other developmental disorder, and all children scored below 60 on the parent-reported SRS (range = 2–56). The children in the CAM group were individually matched to children with ASD based on chronological age (in years) and nonverbal IQ age-equivalence and raw scores from the Kaufman Brief Intelligence Test, 2nd edition (KBIT–2,

Table 1. Mean age and scores on background measures for the individuals with autism spectrum disorder (ASD) and the two typically developing (age-matched and language-matched) control groups.

	ASD		Age-matched		Language-matched	
Participant characteristic	***M***	***SD***	***M***	***(SD)***	***M***	***(SD)***
Age (in months)	115.5	22.7	113.9	22.5	100.0*	24.4
NVIQ age-equivalent	120.0	52.1	128.9	50.6	116.6	48.2
NVIQ raw	26.9	7.9	28.6	7.8	26.3	8.2
Vocab age-equivalent	111.1	29.3	134.4*	38.4	118.3	33.0
Vocab raw	51.2	14.2	60.9*	15.5	54.2	14.3
Syntax age-equivalent	93.0	27.8	116.3*	32.2	94.9	29.5
Syntax raw	27.2	8.6	34.0*	8.7	28.1	9.6
SRS total raw	107.1	24.5	23.3*	13.2	25.0*	15.1

Note. $n = 26$ for each of the three groups. NVIQ = nonverbal IQ subtest from the Kaufman Brief Intelligence Test—2 (KBIT–2); Vocab = Verbal IQ subtest of the KBIT–2 measuring vocabulary; Syntax = Syntax Construction subtest of the Comprehensive Assessment of Spoken Language; SRS = Social Responsiveness Scale.

*$p < .05$ (indicates group is significantly different from ASD group).

Kaufman & Kaufman, 2004). Two-tailed t tests revealed that the children in the ASD and CAM groups did not significantly differ on chronological age, $t(50) = 0.257$, $p > .7$, $d = 0.07$; nonverbal IQ age-equivalence scores, $t(49) = 0.640$, $p > .5$, $d = 0.18$; or nonverbal IQ raw scores $t(49) = 0.770$, $p > .45$, $d = 0.22$. As expected, the children in the CAM group performed significantly higher than the ASD group on syntax (CASL) age-equivalence scores, $t(50) = 2.79$, $p < .05$, $d = 0.78$, and syntax raw scores, $t(50) = 2.847$, $p < .01$, $d = 0.8$. The children in the CAM group performed significantly higher than the ASD group on vocabulary (KBIT–2 verbal IQ) age-equivalence scores, $t(49) = 2.43$, $p < .05$, $d = 0.69$, and vocabulary raw scores, $t(49) = 2.336$, $p < .05$, $d = 0.67$.

Language age-matched control group (LAM). The LAM group consisted of a separate group of 26 children with TD (14 boys, 12 girls), ages 5 to 12 years ($M = 7.88$, $SD = 2.04$). Parent report indicated that these children were free of a diagnosis of ASD or other developmental disorder, and all children scored below 60 on the parent-reported SRS (range = 4–57). It is important to note that the children in the LAM group were also free of a diagnosis of any language disorder. As a result, we used age-equivalence scores and raw scores, rather than standardized scores, to match the children in our LAM group to the children with ASD. Specifically, children in the LAM control group were individually yoked to children in the ASD group based on age-equivalence and raw scores from the Syntax Construction subtest of the CASL.

The children in the LAM and ASD groups did not significantly differ on syntax age-equivalence scores, $t(50) = 0.242$, $p > .7$, $d = 0.06$, or syntax raw scores, $t(50) = 0.35$, $p > .7$, $d = 0.09$. In addition, the children in the LAM and ASD groups did not significantly differ on vocabulary age-equivalence scores, $t(50) = 0.884$, $p > .4$, $d = 0.24$, or vocabulary raw scores, $t(50) = 0.758$, $p > .45$, $d = 0.21$. The children in the LAM and ASD groups also did not significantly differ on nonverbal IQ age-equivalence scores, $t(49) = 0.218$, $p > .8$, $d = 0.06$, or nonverbal IQ raw scores, $t(49) = 0.254$, $p > .8$ $d = 0.07$. As expected, the children in the ASD group were significantly older than the children in the LAM group, $t(50) = –2.19$, $p < .05$, $d = 0.61$. There were more girls in the LAM group compared to the group with ASD. However, a comparison of male versus female participants in the LAM group found that idiom performance did not differ by gender ($p > .8$).

Measures

Idiom comprehension. A total of 20 idioms were presented, each in the context of a short paragraph that supports the figurative meaning of the phrase. The 20 idioms varied on their familiarity and decompositionality. Half of the idioms were high on decompositionality (e.g., above 60%) and half were low on decompositionality (e.g., below 40%) based on the ratings from Titone and Connine (1994). The familiarity of the idioms was determined by adult ratings of familiarity in a pilot study with healthy undergraduate students. Idioms were only included when the adult ratings of familiarity from pilot testing were consistent with the Titone and Connine (1994) scores. Children were provided with one practice item and feedback before the start of the test items. An experimenter read the vignette containing the idiom to the child and asked the child to verbally define the meaning of the idiom (see Appendix).

Idiom responses were classified as correct figurative meaning, related figurative meaning, literal, restated, not related (to the idiom meaning), and no response ("I don't know"). Correct responses were scored as 2 points, related responses were worth 1 point, and all other answers were worth 0 points. Two undergraduate research assistants, blind to diagnosis, independently scored the idiom responses. Percent agreement between the two coders was high (87%), K = .818. When there were discrepancies in the coding of an item, the coders met and resolved the difference. Scores could range from 0 to 40.

Strange stories task. The strange stories task was used to measure TOM performance based on a subset of six of the mentalizing stories from O'Hare et al. (2009). Because the

full task included some figurative language, any stories that included figurative language (including jokes, figures of speech, and sarcasm) were excluded. Thus, a subset of six stories was chosen that illustrated several different kinds of mentalizing (lie, white lie, misunderstanding, contrary emotions, appearance/reality, and forget). Children were read a short paragraph (i.e., story), and then asked questions such as "Is it true what Peter said?" and "Why does he say that?" Two clip art pictures depicting people or objects relevant to each story were presented as memory aids. It is important to note that the O'Hare et al. (2009) measure scores the justification questions on a scale of zero to two points, where partial mental state responses earn one point, and fully correct mental state responses earn two points. In addition, exclusion of the ambiguous figurative items for the current study meant that the question of "Is it true…?" had a clear yes or no answer appropriate for scoring. Thus, each story is worth up to three points (one point for "Is it true…?" and two points for justification), for a total possible range of scores of 18. Two undergraduate research assistants, blind to diagnosis, independently coded the strange stories measure. Percent agreement between the two coders in the current study was high (92%), K = .87. When there were discrepancies, the two coders met to resolve differences in their coding.

RMTE task. The children's version of the RMTE task was completed (Baron-Cohen et al., 2001). During this task, children were presented with pictures of the eye region of a face displaying an expression and had to select (among four options) the appropriate words or phrases that depicted the mental state of the face (i.e., shy, worried, not believing, or kind). The experimenter read the answer choices aloud for each picture and asked which word matched the thoughts/emotions of the person in the picture ("What is this person feeling?"). Children could either point to the correct answer printed on the page or say the answer out loud. Correct responses were scored as 1, incorrect responses as 0. There are a total of 28 items, with a possible range of scores from 0 to 28.

Syntax. Syntax abilities were assessed using the Syntax Construction subtest of the CASL. The test involves presenting pictures to a child and requires that the child verbally complete sentences based on a model provided. For example, the experimenter could say, "Finish what I say: Here is one book (pointing to a picture of a single book). Here (pointing to a picture of a set of three books) ____," and a correct response would include "are three books." Two practice items were administered before the start of the task and children were provided with feedback. Each item targeted a specific syntactic form, focusing on the use of syntax rules. The syntactic forms tested include structures such as plurals, verb tenses, dependent clauses, and adverbs. The reported internal reliability for this subtest is high, .73 to .88 (Carrow-Woolfolk, 1999). Raw scores (out of a possible 60 items) were converted into age-equivalence scores and standard scores.

Vocabulary. The two verbal IQ subtests of the KBIT–2 were used to measure vocabulary abilities. The Verbal Knowledge subtest measured receptive vocabulary. The Riddles subtest measured verbal comprehension, reasoning, and vocabulary knowledge. The internal-consistency reliability reported for ages 4 to 18 on this verbal IQ measure is high, .90 (Kaufman & Kaufman, 2004). The raw scores could range from 1 to 60 for verbal knowledge and from 1 to 48 for riddles. Total raw scores were computed by summing the two subtests and then converting into standard and age-equivalence scores.

Nonverbal IQ. The nonverbal IQ score consisted of the Matrices subtest of the KBIT–2 (Kaufman & Kaufman, 2004). The participants were asked to choose which picture best completed the matrices of pictures (2 × 2 or 3 × 3). Internal-consistency reliability for ages 4 to 18 on this subtest is high, .86 (Kaufman & Kaufman, 2004). Raw scores could range from 1 to 46 and were converted to age-equivalence scores.

Procedure

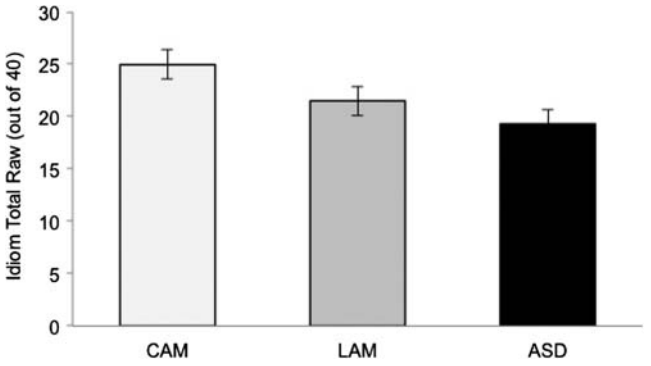
A parent or legal guardian provided informed consent prior to each child's participation in the study. An experimenter read an assent form out loud to each child, and all children provided verbal assent prior to participation. The experimental procedures were approved by and complied with the standards of the university's internal review board. Testing took place either in offices in the research lab or in a quiet room in the child's home (for 12 of the children with ASD who were unable to travel to the lab). The study was completed over two sessions, lasting approximately 1 hr each (for a total of 2 hr of data collection). For most lab visits, the sessions were completed approximately 1 week apart. For home visit appointments or for lab visits when the family was unable to schedule two different appointment days (due to traveling distance or scheduling conflicts), both sessions were completed on the same day. Breaks were given between tasks, and testing procedures were kept as similar as possible across testing locations. All tasks were completed in the same order for all of the participants. All measures were read aloud to the children by an experimenter, and children were asked to respond verbally (or by pointing when appropriate).

Results

Overall Idiom Performance

Figure 1 illustrates the performance of each of the three groups (ASD, CAM, and LAM) on idiom comprehension scores. Idiom performance was examined for the children with ASD compared to each of the two TD control groups. A one-way analysis of variance (ANOVA) was conducted comparing total idiom scores (out of 40 possible points) for the three groups (ASD, CAM, and LAM). There was an effect of group on idiom comprehension total scores, $F(2, 75) = 3.506$, $p < .05$, $\eta^2 = .09$. The children in the CAM group ($M = 24.96$, $SD = 7.02$, range = 5–34) performed significantly higher than the group with ASD ($M = 19.30$, $SD = 8.06$, range = 0–31) on idiom comprehension, $t(50) = 2.69$, $p < .01$, $d = 0.76$. However, there was no significant difference between the children in the LAM group ($M = 21.5$,

Figure 1. Mean (with standard error of the mean) idiom score (out of 40 points) for the three groups: chronological age-matched controls (CAM), language age-matched controls (LAM), and children with autism spectrum disorder (ASD).



SD = 8.14, range = 5–34) and the group with ASD on idiom performance ($p > .3$).

Idiom Item-Level Factors

To examine the effect of familiarity on idiom comprehension for each group, a repeated-measures ANOVA was conducted with familiarity (low vs. high) as a within-subjects factor and diagnostic group as a between-subjects factor. There was a main effect of familiarity, where performance was more accurate for high-familiarity idioms ($M = 12.17$, $SD = 4.19$) than for low-familiarity idioms ($M = 9.75$, $SD = 4.41$), $F(1, 75) = 44.50$, $p < .01$, $\eta_p^2 = .37$. There was not a significant interaction between familiarity and diagnostic group ($p > .9$).

To examine the effect of decompositionality on idiom comprehension for each group, a mixed-factors ANOVA was conducted with decompositionality (low vs. high) as a within-subjects factor and diagnostic group as a between-subjects factor. There was a main effect of decompositionality, although an examination of the means shows that this is in an unexpected direction, where performance was better for low decompositional (more opaque) idioms ($M = 11.97$, $SD = 4.34$) than for high decompositional (more transparent) idioms ($M = 9.95$, $SD = 4.21$), $F(1, 75) = 36.37$, $p < .001$, $\eta_p^2 = .33$. There was no interaction between decompositionality and diagnosis ($p > .3$). Although the current study found the opposite of what could be expected from the theories of how individuals use decompositionality to understand idioms (Hamblin & Gibbs, 1999; Nippold & Duthie, 2003), the same trend as the current study was found by Norbury (2004) using a similar task and coding scheme for children with TD and ASD.

Overall TOM Performance

One-way ANOVAs were conducted comparing strange stories and RMTE accuracy scores separately for each of the three groups (ASD, CAM, and LAM). There was a significant difference between the groups on the strange stories task, $F(2, 75) = 8.59$, $p < .001$, $\eta^2 = .19$. The children with ASD ($M = 9.88$, $SD = 4.94$, range = 1–18) were less accurate in interpreting the mental states described in the strange stories than both the CAM group ($M = 14.26$, $SD = 14.26$, range = 8–18), $t(50) = 4.36$, $p < .001$, and the LAM group ($M = 13.04$, $SD = 4.94$, range = 7–18), $t(50) = 2.646$, $p < .05$. In contrast, the children in the ASD ($M = 15.38$, $SD = 3.60$, range = 9–22), CAM ($M = 16.46$, $SD = 3.66$, range = 10–23), and LAM ($M = 17.42$; range = 7–26) groups did not perform significantly differently on the RMTE ($p > .1$).

The Relationship Between TOM and Idiom Abilities

To examine the relationships between idiom comprehension, basic language abilities, and TOM, two-tailed Pearson correlations were conducted between the measures. The two control groups (CAM, LAM) were combined into a single group of children with TD for the correlation analysis, because they came from the same general population of children with TD. The correlations between measures were examined separately in the groups of children with ASD and TD. Total idiom comprehension, vocabulary raw scores, syntax raw scores, RMTE, and strange stories were entered as variables for each of the two groups. Table 2 illustrates the full set of correlational analyses among the variables for the children with ASD. For the children with ASD, idiom comprehension significantly correlated with vocabulary raw scores, syntax raw scores, RMTE, and strange stories ($r = .841$, .734, .646, and .699, respectively). Table 3 illustrates the full set of correlational analyses among the variables for the children with TD. Similarly, among children with TD, idiom comprehension scores significantly correlated with vocabulary raw scores, syntax raw scores, RMTE, and strange stories scores ($r = .840$, .822, .547, and .477, respectively).

To examine whether or not TOM abilities predict idiom comprehension above and beyond the contributions of language, partial correlations were conducted for each of the groups. One-tailed partial correlations, controlling for both vocabulary and syntax raw scores, were conducted separately for children with ASD and TD, to reexamine relationships between total scores on the idioms, RMTE, and strange stories. For the children with ASD, idiom comprehension was related to performance on the strange stories task, $r(22) = .399$, $p < .05$, as well as the on the RMTE, $r(22) = .466$, $p < .05$, even after controlling for basic language abilities. In contrast,

Table 2. Bivariate correlations between scores on vocabulary, syntax, the "reading the mind in the eyes" (RMTE) and the strange stories tasks, and idiom comprehension for the group of children with ASD ($n = 26$).

Variable	1	2	3	4	5
1. Vocabulary	—	.853**	.511**	.632**	.841**
2. Syntax		—	.436*	.651**	.734**
3. RMTE			—	.642**	.646**
4. Strange stories				—	.699**
5. Idioms					—

*$p < .05$. **$p < .01$.

Table 3. Bivariate correlations between scores on vocabulary, syntax, the RMTE and strange stories tasks, and idiom comprehension for the group of children with TD (n = 52).

Variable	1	2	3	4	5
1. Vocabulary	—	.894**	.588**	.479**	.840**
2. Syntax		—	.559**	.496**	.822**
3. RMTE			—	.245	.547**
4. Strange stories				—	.477**
5. Idioms					—

*p < .05. **p < .01.

for the children with TD, there was no relationship between idiom comprehension and either the RMTE or strange stories measures of TOM when controlling for basic language abilities (p > .1).

Discussion

The current study evaluated the relationship between idiom, syntax, and TOM abilities in high-functioning children with ASD. We found that both syntax and TOM abilities predict idiom comprehension success in children with ASD. We uncovered revealing interactions in the way that syntactic knowledge and TOM abilities contribute to figurative language comprehension in ASD. Specifically, children with ASD can determine the meaning of idioms presented in a supportive linguistic context in the same way that their language-matched (but younger) peers do. However, advanced TOM abilities appear to also be important for idiom comprehension in children with ASD, because TOM abilities predicted relative deficits in idiom comprehension even after controlling for the contributions of basic language abilities. This suggests that both linguistic skills and advanced TOM abilities contribute to idiom comprehension for children with ASD.

Linguistic Theories of Idiom Comprehension

Overall, these results largely support the hypothesis of Gernsbacher and Pripas-Kapit (2012), who suggested that children with ASD would not show delays in their figurative language abilities when compared to children with TD matched on syntax abilities. The choice of matching variables (either matching on nonverbal IQ and chronological age, or matching on syntax abilities) was particularly revealing about whether the children with ASD showed a significant delay compared to typically developing controls. Although the children with ASD scored lower than the nonverbal IQ and age-matched children with TD (CAM group) on idiom comprehension, the children with ASD did not have significantly lower idiom scores than the younger language-matched children with TD (LAM group). In addition, the significant correlations between idiom comprehension and both vocabulary and syntax raw scores provide further evidence of the importance of basic language abilities for idiom comprehension. This suggests that development in idiom comprehension is predicted by development of basic language abilities, consistent with the linguistic theories (however, the current study found that TOM abilities also contribute to idiom comprehension for children with ASD).

The effect of previous exposure to the figurative phrases (familiarity), prior to participating in the study, was examined. The children (both with ASD and TD) scored higher on high-familiarity idioms (where they likely had more exposure to the phrases) than on low-familiarity idioms. In addition, there was not a significant interaction between diagnosis and familiarity, suggesting that familiarity had a similar effect across the groups. This is consistent with the findings of previous studies of idiom comprehension in children and the linguistic theories related to the role of familiarity in idiom comprehension (Ackerman, 1982; Nippold & Taylor, 2002).

When the individual words in phrases are used to help figure out the idiom's meaning, the linguistic theories suggest that such high decompositional idioms should be easier to understand than low decompositional idioms (Nippold & Duthie, 2003; Nippold & Rudzinski, 1993). The current study did not find support for this role of decompositionality, and instead found the opposite trend. Low decompositional idioms had higher scores in all the groups compared to high decompositional idioms. Several other previous studies examining idiom comprehension using similar tasks and coding schemes have also failed to find the expected effect for decompositionality (Abrahamsen & Burke-Williams, 2004; Norbury, 2004; Whyte et al., 2013). In addition, there was not a significant interaction between diagnosis and decompositionality, suggesting that all three groups in the current study showed a similar pattern of results.

Rather than relying on the decompositionality of the phrases, children may sometimes instead rely upon the more reliable information available (such as the other linguistic or social contextual cues, or previous experience with the phrase) for interpreting the correct meaning of the idioms. One limitation in the current study is that the ratings of decompositionality and familiarity came from adults (and not the children themselves). Future research examining the linguistic theories of idiom comprehension should try to use familiarity and decompositionality ratings from the children participating in the study to further clarify the role of these item-level variables. In addition, future research could vary the task demands and amount of context to see how this impacts how much children rely on item decompositionality as a factor in their comprehension strategy. Future research examining interactions between decompositionality and the supportive linguistic context should also include novel idiomatic expressions, with no possible prior exposure, to prevent familiarity from serving as a potential confound in examining the strategies that individuals with ASD use to understand figurative phrases (Cain, Oakhill, & Lemmon, 2005).

The Relationship Between TOM and Language

Previous studies have provided inconsistent results about a potential relationship between TOM abilities (using

first- and/or second-order false-belief performance) and figurative language for children, with some finding a significant relationship (Caillies & Le Sourn-Bissaoui, 2008, 2012; Happé, 1993; Martin & McDonald, 2004) and others failing to find a significant relationship (Norbury, 2004, 2005). False-belief measures of TOM are very highly tied to vocabulary and syntax development, and they have the additional problem of not being developmentally appropriate for older typically developing children. Thus, the current study examined performance on two TOM tasks that varied in their linguistic demands and were more developmentally appropriate for our age group: the RMTE and strange stories tasks.

The individuals with ASD in the current study showed a wide range of performance on the RMTE and strange stories tasks (with some individuals with ASD actually performing quite well on these measures). Overall, the group of children with ASD showed lower performance than both groups of children with TD on the strange stories task, but not the RMTE task. Although Gernsbacher and Pripas-Kapit (2012) suggest that linguistic differences can account for the differences in TOM performance, the lower performance for the ASD group relative to the language-matched group of children with TD on the strange stories task (and a lack of any group difference on the RMTE) suggests that syntax abilities alone cannot account for the performance of children with ASD on either advanced TOM measure.

We found that for children with ASD and TD, both measures of advanced TOM abilities correlated with basic language abilities (vocabulary and syntax raw scores), as well as idiom comprehension. On the one hand, this finding seems to support the relevance theory (Happé, 1993). However, when we controlled for basic language abilities (vocabulary and syntax raw scores), performance in the TOM tasks only predicted idiom comprehension for children with ASD, and not for the children with TD. In addition, this set of findings was consistent across both the task requiring interpretations of mental states based on listening to story paragraphs (the strange stories task), and a task that required matching words or phrases, describing mental states, to pictures of faces (the RMTE task).

An alternative to the relevance theory for explaining the relationship between TOM and figurative language abilities may be that language skills and TOM skills more broadly develop together across time, and that figurative language abilities do not have a privileged relationship with TOM. Theories of the broader social communication development (explaining the relationship between TOM and basic language abilities) can potentially account for the relationship between TOM skills and figurative language, and are potentially consistent with the current linguistic theories of idiom development (Garfield, Peterson, & Perry, 2001; Miller, 2006). Miller (2006) suggested that "language skills grow and support a developing theory of mind, while at the same time, the increasingly sophisticated theory of mind makes it possible to engage in meaningful communication" (p. 147). Nelson and colleagues (Nelson, 2000; Nelson & Arkenberg, 2008; Nelson, Welsh, Vance Trup, & Greenberg, 2011; Whyte et al., 2013) have similarly stressed from a dynamic systems perspective that multiple domains of language develop in a closely interrelated fashion and that similar patterns of cognitive skills contribute to all areas of language development. However, longitudinal studies may be necessary to understand how these skills may develop together across childhood.

A limitation of the current study was in choosing two TOM measures that are still highly intercorrelated with vocabulary and syntax, although this seems to be a limitation of the majority of TOM measures commonly used in the literature. Future research should include an even larger number of TOM measures tapping into varying aspects of TOM and vary more significantly in their reliance on verbal abilities than those used in the current study. Additional TOM measures available include the Theory of Mind Inventory and Theory of Mind Task Battery (Hutchins, Bonazinga, Prelock, & Taylor, 2008; Hutchins & Prelock, 2008; Hutchins, Prelock, & Bonazinga, 2012; Hutchins, Prelock, & Chace, 2008; Lerner, Hutchins, & Prelock, 2011). The Theory of Mind Inventory measure is particularly interesting because it involves parent report rather than direct assessment with the child, which may provide a way to reduce the continued reliance on the child's verbal responses for completing TOM tasks (Hutchins et al., 2012). Future research should include longitudinal studies to examine how a broad range of language abilities (including both syntax and figurative language) develops over time with a broad range of TOM measures.

Conclusion

Strengths of the current study examining idiom comprehension development include the use of two different typically developing control groups (including one matched on syntax development) to aid the interpretation of results for children with ASD, and the use of multiple TOM measures not previously examined in relation to idiom performance. The study provides a clear demonstration that children with ASD fall behind in their idiom comprehension compared to an age-matched and nonverbal IQ–matched typically developing control group, but do not show a deficit in idiom comprehension compared to a slightly younger control group matched on syntax raw and age-equivalence scores.

A limitation of the current research is that we were not able to use diagnostic tests with our sample, such as the ADOS (Lord, Rutter, DiLavore, & Risi, 1999). The SRS measure of current ASD symptoms is a screening tool rather than a diagnostic test. Future research examining the relationship between figurative language and advanced TOM abilities should include more thorough diagnostic testing. In addition, a limitation is that our sample is relatively high functioning, as our tasks required fairly sophisticated verbal responses. Future research should include measures of the production of idioms and other figurative abilities in the context of social interactions, as well as the use of other response formats for examining the use and comprehension of figurative abilities. Future research should include control participants with specific language impairments (without the ASD social impairments) to allow for matching on standardized language scores rather than age-equivalence or raw

scores. This may help sort out the relative contributions to figurative language skills of social skills, advanced TOM abilities, and language abilities.

Although the current study did not find a delay in idiom comprehension relative to younger language-matched peers, the children with ASD did show language deficits across multiple language domains relative to their age-matched peers. Thus, idioms and other domains of language, such as syntax abilities, should be considered as targets for language assessment for school-aged children with ASD. For individual children who show delays in these areas, interventions should target both the structural aspects of language and figurative language abilities. In particular, interventions designed to target idiom comprehension have been successful for children with ASD or other language impairments (Abrahamsen & Smith, 2000; Ezell & Goldstein, 1992; Lundblom & Woods, 2012; Whyte et al., 2013). A common feature of these few previous idiom interventions has been the embedding of idiomatic phrases in the context of supportive story paragraphs to highlight relevant contextual cues. Interventions designed to target language abilities in children with ASD or other language disorders should keep in mind the possible need for examining and intervening on multiple domains (syntax, idioms, and TOM abilities) together.

Acknowledgments

We thank several research assistants for their help in data collection for this study, including Caitlin Hall, Katherine Hansen, Jeffrey Krystek, and Tyler Naill. We also thank the study families who made this research possible. A Penn State Liberal Arts dissertation improvement award to the first author provided funding for this project.

References

Abrahamsen, E. P., & Burke-Williams, D. (2004). Comprehension of idioms by children with learning disabilities: Metaphoric transparency and syntactic frozenness. *Journal of Psycholinguistic Research, 33,* 203–215. doi:10.1023/B:JOPR.0000027962.42590.60

Abrahamsen, E. P., & Smith, R. (2000). Facilitating idiom acquisition in children with communication disorders: Computer vs. classroom. *Child Language Teaching and Therapy, 16,* 227–239.

Ackerman, B. P. (1982). On comprehending idioms: Do children get the picture? *Journal of Experimental Child Psychology, 33,* 439–454. doi:10.1016/0022-0965(82)90058-3

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35,* 1331–1320.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome. *Journal of Developmental and Learning Disorders, 5,* 47–78. doi:10.1111/j.1467-7687.2008.00720.x

Caillies, S., & Le Sourn-Bissaoui, S. (2008). Children's understanding of idioms and theory of mind development. *Developmental Science, 11,* 703–711.

Caillies, S., & Le Sourn-Bissaoui, S. (2012). Nondecomposable idiom understanding in children: Recursive theory of mind and working memory. *Canadian Journal of Experimental Psychology*. Advance online publication. doi:10.1037/a0028606

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology, 90,* 65–87. doi:10.1016/j.jecp.2004.09.003

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Circle Pines, MN: AGS.

Constantino, J. N. (2002). *Social Responsiveness Scale*. Los Angeles, CA: Western Psychological Services.

Dennis, M., Lazenby, A. L., & Lockyer, L. (2001). Inferential language in high-functioning children with autism. *Journal of Autism and Developmental Disorders, 31,* 147–154. doi:10.1023/A:1005661613288

Ezell, H. K., & Goldstein, H. (1992). Teaching idiom comprehension to children with mental retardation. *Journal of Applied Behavior Analysis, 25,* 181–191. doi:10.1901/jaba.1992.25-181

Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social cognition, language acquisition, and the development of theory of mind. *Mind & Language, 16,* 494–541. doi:10.1111/1468-017.00180

Gernsbacher, M. A., & Pripas-Kapit, S. R. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol, 27,* 93–105. doi:10.1080/10926488.2012.656255

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6,* 346–359. doi:10.1111/1467-7687.00289

Hamblin, J. L., & Gibbs, R. W. (1999). Why you can't kick the bucket as you slowly die: Verbs in idiom comprehension. *Journal of Psycholinguistic Research, 28,* 25–39. doi:10.1023/A:1023235403250

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of the relevance theory. *Cognition, 48,* 101–119. doi:10.1016/0010-0277(93)90026-R

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding story characters' thought and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24,* 129–154. doi:0162-3257/94/0400-0129507.00/0

Hutchins, T., Bonazinga, L., Prelock, P. A., & Taylor, R. S. (2008). Beyond false beliefs: The development and psychometric evaluation of the Perceptions of Children's Theory of Mind Measure—Experimental version (PCToMM–E). *Journal of Autism and Developmental Disorders, 38,* 143–155. doi:10.1007/s10803-007-0377-1

Hutchins, T., & Prelock, P. A. (2008). Supporting theory of mind development: Considerations and recommendations for professionals providing services to individuals with ASD. *Topics in Language Disorders, 28,* 340–364. doi:10.1097/01.TLD.0000341128.01158.f2

Hutchins, T., 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,* 327–341. doi:10.1007/s10803-011-1244-7

Hutchins, T., Prelock, P. A., & Chace, W. (2008). Test-retest reliability of Theory of Mind tasks representing a range of content and complexity and adapted to facilitate the performance of children with ASD. *Focus on Autism and Other Developmental Disabilities, 23,* 195–206. doi:10.1177/1088357608322998

Jolliffe, T., & Baron-Cohen, S. (1999). The strange stories test: A replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 29,* 395–404. doi:10.1023/A:1023082928366

Kaland, N., Callesen, K., Moller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced

theory of mind tasks. *Journal of Autism and Developmental Disorders, 38,* 1112–1123. doi:10.1007/s10803-007-0496-8

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Circle Pines, MN: AGS.

Kerbel, D., & Grunwell, P. (1997). Idioms in the classroom: An investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching and Therapy, 13,* 113–123. doi:10.1177/026565909701300201

Kerbel, D., & Grunwell, P. (1998a). A study of idiom coprehension in children with semantic-pragmatic difficulties: Part 1. Task effects on the assessment of idiom comprehension in children. *International Journal of Language and Communication Disorders, 33,* 1–22. doi:10.1080/136828298247901

Kerbel, D., & Grunwell, P. (1998b). A study of idiom comprehension in children with semantic-pragmatic difficulties: Part 2. Between-groups results and discussion. *International Journal of Language and Communication Disorders, 33,* 23–44. doi:10.1080/136828298247910

Lazar, R. T., Warr-Leeper, G. A., Nicholson, C. B., & Johnson, S. (1989). Elementary school teachers' use of multiple meaning expressions. *Language, Speech, and Hearing Services in Schools, 20,* 4420–4430.

Lerner, M., Hutchins, T., & Prelock, P. A. (2011). Brief report: Preliminary evaluation of the Theory of Mind Inventory (ToMI) and its relationship to measures of social skills. *Journal of Autism and Developmental Disorders, 41,* 512–517. doi:10.1007/s10803-010-1066-z

Levorato, M. C., & Cacciari, C. (1995). The effects of different tasks on the comprehension and production of idioms in children. *Journal of Experimental Child Psychology, 60,* 261–283. doi:10.1006/jecp.1995.1041

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule–WPS (ADOS-WPS)*. Los Angeles, CA: Western Psychological Services.

Lundblom, E., & Woods, J. (2012). Working in the classroom: Improving idiom comprehension through classwide peer tutoring. *Communication Disorders Quarterly, 33,* 202–219. doi:10.1177/1525740111404927

Lyons, V., & Fitzgerald, M. (2004). Humor in autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 34,* 521–531. doi:10.1007/s10803-004-2547-8

Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of Autism and Developmental Disorders, 34,* 311–328. doi:10.1023/B:JADD.0000029553.52889.15

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech-Language Pathology, 15,* 142–154. doi:0.1044/1058-0360(2006/014)

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,* 622–646. doi:10.1111/j.1467-8624.2007.01018.x

Nelson, K. E. (2000). Methods for stimulating and measuring lexical and syntactic advances: Why Fiffins and lobsters can tag along with other recast friends. In L. Menn & N. B. Ratner (Eds.), *Methods for studying language production* (pp. 113–146). Hillsdale, NJ: Erlbaum.

Nelson, K. E., & Arkenberg, M. E. (2008). Language and reading development reflect dynamic mixes of learning conditions. In M. Mody & E. Silliman (Eds.), *Brain, behavior, and learning in language and reading disorders* (pp. 315–348). New York, NY: Guilford.

Nelson, K. E., Welsh, J. A., Vance Trup, E. M., & Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Language, 31,* 164–194. doi:10.1177/0142723710391887

Nippold, M. A., & Duthie, J. K. (2003). Mental imagery and idiom comprehension: A comparison of school-age children and adults. *Journal of Speech, Language, and Hearing Research, 46,* 788–799. doi:10.1044/1092-4388(2003/062)

Nippold, M. A., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech and Hearing Research, 36,* 728–731.

Nippold, M. A., & Taylor, C. L. (2002). Judgments of idiom familiarity and transparency: A comparison of children and adolescents. *Journal of Speech, Language, and Hearing Research, 45,* 384–391. doi:10.1044/1092-4388(2002/030)

Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research, 47,* 1179–1193. doi:1092-4388/04/4705-1179

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology, 23,* 383–399. doi:10.1348/026151005X26732

O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism and Developmental Disorders, 39,* 916–928. doi:10.1007/s10803-009-0699-2

Qualls, C. D., O'Brien, R. M., Blood, G. W., & Hammer, C. S. (2003). Contextual variation, familiarity, academic literacy, and rural adolescents' idiom knowledge. *Language, Speech, and Hearing Services in Schools, 34,* 69–79. doi:0161–1461/03/3401–0069

Spector, C. C. (1996). Children's comprehension of idioms in the context of humor. *Language, Speech, and Hearing Services in Schools, 27,* 303–313.

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Brief report: Developmental change in theory of mind abilities in children with autism. *Journal of Autism and Developmental Disorders, 33,* 461–767. doi:10.1023/A:1025075115100

Sullivan, K., Zaitchick, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30,* 395–402. doi:10.1037/0012-1649.30.3.395

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science, 16,* 311–315. doi:10.1111/j.1467-8721.2007.00527.x

Tager-Flusberg, H., & Joseph, J. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–318). New York, NY: Oxford University Press.

Titone, D. A., & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbolic Activity, 9,* 247–270. doi:10.1207/s15327868ms0904_1

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development, 80,* 1097–1117. doi:10.1111/j.1467-8624.2009.01319.x

Whyte, E. M., Nelson, K. E., & Khan, K. S. (2013). Learning of idiomatic language expressions in a group intervention for children with autism. *Autism, 17,* 449–464. doi:10.1177/1362361311422530

Appendix

Example Item and Scoring From the Idiom-in-Context Measure

In-context vignette: I was supposed to go to my friend's house after school on Friday. However, I didn't make it to my friend's house because it **slipped my mind**.

Question: What does "slipped my mind" mean?

Correct answer (2 points): You forgot about it.
Related figurative (1 point): You don't think about it (needs to say "forget" for 2 points).
Literal (0 points): Your mind crazy; Slip on soap or banana peel
Restated (0 points): Slipped your mind
Not related (0 points): I had to do chores; Skip