

9-16-2019

# Teaching Children with Autism to Understand Metaphors

Gabrielle Lee  
glee329@uwo.ca

Sheng Xu  
*Chongqing Normal University*

Huiling Zou  
*Hainan Normal University*

Lina Gilic  
*State University of New York College at Old Westbury*

Michelle M. Lee  
*Michigan State University*

Follow this and additional works at: <https://ir.lib.uwo.ca/edupub>

 Part of the [Education Commons](#)

---

## Citation of this paper:

Lee, G. T., Xu, S., Zou, H., Gilic, L., & Lee, M. W. (2019). Teaching Children with Autism to Understand Metaphors. *The Psychological Record*, 1-14.

Teaching Children with Autism to Understand Metaphors

Suggested citation:

Lee, G. T., Xu, S., Zou, H., Gilic, L., & Lee, M. W. (2019). Teaching children with autism to understand metaphors. *Psychological Record*, Advance Online Publication.

### Abstract

The purpose of this study was to evaluate the effects of an instructional procedure on the acquisition and generalization of metaphorical understanding for children with autism spectrum disorder. Three students (two boys, one girl, 5-8 years old) participated but only two completed the study. A multiple-probe design across two behaviors and three participants was used. The metaphors were categorized by topography: metaphors involving physical features and metaphors involving abstract properties. The instruction consisted of intraverbal training using echoic prompts, picture prompts, and textual prompts. The results indicated that the instruction was effective in establishing metaphorical understanding of target metaphors. Generalized understanding to untaught metaphors occurred for the two students who completed the study, and all metaphors were maintained at a relatively high level for two months following the instruction.

*Keywords:* metaphors, metaphorical understanding, relational frame theory, intraverbal, autism spectrum disorder

### Teaching Children with Autism to Understand Metaphors

Metaphorical language use is pervasive in social conversations and written expressions in our daily lives, and is common in many cultures (Semino & Demjén, 2016). In elementary schools, teachers frequently use metaphorical language to introduce new topics, establish rules, and provide feedback and evaluations (Cameron, 2003). For example, a teacher may use this metaphor in a science class, “The solar system is a single family that has people, pets, and their stuff. The star cluster is an apartment building with many families in one building, each family being a solar system. The galaxy is a city, which is a big group of families and apartment buildings” (Professor Astronomy, 2010). In social conversation, a child may say, “Math is a piece of cake for me.” Any child who does not understand such language may feel confused, awkward, and excluded in school. Indeed, metaphors are present and inevitable in everyday communication, and thus, metaphorical comprehension plays an essential role in children’s knowledge acquisition and social interactions.

Previous research consistently suggests that children with autism spectrum disorder (ASD) are deficient in metaphorical understanding compared to typically developing peers (Melogno, D’Ardia, Pinto, & Levi, 2012; Rundblad & Annaz, 2010), or children with learning disabilities (Mashal & Kasirer, 2012; 2011). ASD is a neurodevelopmental disorder with core deficits in social communication, restricted interests and activities, or patterns of behavior (American Psychiatric Association, 2013). The underlying mechanisms responsible for metaphorical comprehension in individuals with ASD are unknown. Given the prevalence of figurative language (Semino & Demjén, 2016), difficulties of understanding metaphors in individuals with ASD have shown to aggravate their existing deficiency in social communication and further limit their social interactions. Such difficulties also hinder academic achievement, as adequate

understanding of metaphors is necessary when learning subject matters in schools (Cameron, 2003). Therefore, it is imperative to teach metaphorical comprehension so that individuals with ASD are able to engage in meaningful conversations with others and effectively acquire academic knowledge.

To devise instruction that effectively improves metaphorical comprehension, it is necessary to analyze the behavioral processes involved complex verbal behaviors. Skinner (1957) defines metaphor as one type of *extended tact*, in which properties of a stimulus evoke a response and share some, but not all, relevant properties of the stimulus that control the response. For example, a person may be called an owl because s/he tends to be energetic past midnight. Some properties of an owl, such as staying awake and active at night, are exhibited by the person. The description implying characteristics of an owl used to refer to a person constitutes a metaphorical extension. However, the analysis of extended tact does not fully explain the behavioral processes involved in the development of metaphorical verbal behavior. Some unanswered questions pertaining to the development of metaphorical verbal behavior include how relevant properties are selected, how irrelevant properties are suppressed, and how relevant properties of a stimulus are related to another stimulus (Stewart & Barnes-Holmes, 2001). Stewart and Barnes-Holmes (2001) used relational frame theory (RFT) to interpret metaphor as a behavioral phenomenon. RFT proposes that human language and cognition are behavioral phenomena developed via *arbitrarily applicable relational responding* (Hayes, Barnes-Holmes, & Roche, 2001). The cornerstone of complex human behavior lies in the ability to create bidirectional relations with an arbitrary number of relations and dimensions to be related. According to Stewart and Barnes-Holmes (2001), metaphorical comprehension involves a unidirectional equivalence relation between two elements formed through their shared formal

similarities. To understand a metaphor, one has to describe the properties belonging to each of the two elements (the hierarchical relation), discriminate similarities and dissimilarities (the distinction relation) of these properties between two elements, and form an equivalence relation by selecting the shared properties. Additionally, the behavioral processes involved in the understanding of metaphors and analogies are similar, except the equivalence relation is reversible in analogies but not in metaphors (Stewart & Barnes-Holmes, 2001). Thus, the derivation of a unidirectional relation between two elements in a metaphor is involved in metaphorical comprehension.

The behavior-analytic interpretation of metaphor provides a basis for requisite skills involved in the process of metaphorical comprehension and thus have important implications in the development of effective instruction. An intervention study by Persicke, Tarbox, Ranick, and St. Clair (2012) was the first attempt to apply RFT in the instruction of metaphorical comprehension to children with ASD. The instruction consisted of the instructor asking questions to prompt student responses for the hierarchical, distinction, and equivalence relations relevant to the target metaphor. Specifically, after reading a short story, the instructor presented the metaphor question (e.g., “If I say the cake was a rock, what do I mean?”). If the child did not respond correctly, the instructor provided questions guiding the child to describe the features of two elements (e.g., A cake is sweet, fluffy, and smells good; a rock is hard, heavy, and solid) and to discriminate the different features and the shared features between the two elements in order for the child to emit a correct response (e.g., The cake was hard). Multiple exemplar instruction was another important characteristic which involved contextual cues presented in a story (e.g., “The cake had fluffy frosting, and it smelled really good, but was hard on the inside”) to include several metaphors for an object (e.g., The cake was perfume; the cake’s frosting was a cloud; the

cake was a rock). The researchers reported that all three children acquired metaphorical comprehension and such comprehension was generalized to untaught metaphors after the instruction was completed. Of the three children, two of them required additional visual prompts in order to acquire metaphorical comprehension. The visual prompts added to the intraverbal training consisted of a paper with two columns where the student was instructed to write the names of the two elements on top and their associated features underneath each element. The child was then required to draw a line to connect the shared feature(s) between the two elements.

One of the strengths of Persicke et al.'s (2012) intervention was the use of *intraverbal training* (question-and-answer) with a supplement of visual prompts to guide the student tacting the relations between the two elements and their associated features, thereby leading to metaphorical comprehension. A potential limitation of their study concerns target selection. The target metaphors were written by the researchers without referencing the children's actual reading level or their grade equivalence. Additionally, the target metaphors were not evaluated to determine their age or developmental appropriateness. It is necessary to select target metaphors based on the children's reading levels or their grade equivalence and conduct pre-assessments to evaluate the children's prerequisite skills.

Based on the psycholinguistic tradition, Mashal and Kasirer (2011) used thinking maps to improve metaphor comprehension in children with ASD and children with learning disabilities. The thinking maps used by Marshal and Kasirer (2011) were very similar to the visual supports used in Persicke et al. (2012). Each thinking map consisted of the two elements in a metaphor written in the central bubbles with several surrounding bubbles connected to each element. The features for each element were written in the surrounding bubbles. The instructor guided the children to write the shared features in the bubbles connected to both elements in order for them

to understand the metaphor. The intervention was implemented in two sessions with a small group of two to four children. Five metaphors were taught using the thinking maps in each session. However, neither the questions asked, nor the assistive prompts provided by the instructor were clearly described. Results indicated that children with ASD maintained the target metaphors learned in the sessions but comprehension did not generalize to untaught metaphors after the intervention. It is possible that the intervention did not have sufficient metaphor exemplars for generalization to occur, as it was only implemented in two training sessions with 10 metaphors taught. In addition, no mastery criterion was required for each child to conclude the training.

Both visual support (Persicke et al., 2012) and thinking maps (Marshal & Kasirer, 2011) involved a visual structure of associated features for the two elements involved in metaphors. The instructor used intraverbal training to point out the equivalence relation between the two elements and guide the child in the selection of shared features in order to comprehend the metaphor. When children were asked to generate free associations (e.g., features of an item) with each element, additional verbal cues (e.g., the category and the conventional use of the element) may be necessary for them to emit accurate responses relevant to the metaphor, especially in the beginning stage of acquisition. Once the shared feature(s) was found, the children could summarize the equivalence relation to interpret the meaning of the metaphor.

The visual supports used in the above two studies required the children to write words, and thus, the children needed to have handwriting skills for the visual supports to be effective. Previous research on *transfer stimulus control from tact to intraverbal* suggests that it is effective in increasing intraverbal responses in children with ASD or developmental disabilities (Braam & Poling, 1983; Goldsmith, LeBlanc, & Sautter, 2007; Luciano, 1986). It is possible to present



pictures for each element involved in a metaphor to prompt for correct intraverbal responses in naming the elements and their features. Besides facilitating the transfer from tact to intraverbal, pictures provide visual cues for the identification of each element's associated features and the common feature(s) of the two elements in a metaphor. In addition, the use of pictures does not require handwriting, which is also appropriate for children who do not have handwriting skills in their repertoire.

Expanding from previous research, this study attempted to apply RFT to teach children with ASD metaphorical comprehension via intraverbal training. Generalization to understanding untaught metaphors was also evaluated before and after the instruction. The following research questions were addressed: (a) to what extent does the instruction increase the children's correct responses to target metaphors in probe sessions across baseline, instruction, and follow-up conditions?, and (b) to what extent does the instruction increase the children's correct responses to generalization metaphors in probe sessions across conditions?

## **Method**

### **Participants**

Recruitment of research participants was posted on the website of the university-affiliated autism research center and the WeChat (a social media app widely used in China) of a parent support group for families of children with ASD in the local area. The initial inclusion criteria included that the child had (a) an ASD diagnosis, (b) basic social communication skills, (c) difficulties in language comprehension suggested by a teacher, and (d) instructional control in the classroom. The initial screening was conducted via a phone conversation with the child's parents. The parents were invited to bring their children to gather basic information of the study and a brief interview. Parental consent was obtained during this visit.

Assessments of each participant's skills in social communication, question asking, and understanding of jokes and metaphors were conducted in a brief interview based on the Developmentally-based Behavior Assessment for Children with Autism (DBACA; Feng & Sun, 2017). The brief interview consisted of four segments. First, the experimenter began conversations by introducing herself and asking the student social questions (e.g., "What's your name?" "What do you like to do when you have free time?"). Second, the experimenter continued with the topic the student was interested and asked more topic-specific questions (e.g., "What's a movie you have seen recently?" "Who is your favorite character in the movie?"). The experimenter also guided the student to ask questions to continue the conversation (e.g., "Now you ask me some questions about movies I like.>"). Third, after the social conversation, the experimenter asked the student to tell a joke or something interesting (e.g., "Would you share something funny?" "Why do you like it?" "What's funny about it?"). Fourth, the experimenter described a metaphor and asked the student to explain the meaning. The metaphor was identical for all participants during the interview and was not included as one of the target metaphors or generalization metaphors in this study: In a racing contest, Ming said to Yuan, "You run like a turtle." "What does Ming mean?" The experimenter concluded the interview by thanking the child and parent. The entire interview lasted approximately 7 to 10 min.

The first participant, Ray, was an 8-year-old boy with ASD. According to his diagnosis documentation, his total IQ score was 87, measured by the Chinese version of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003; Zhang, 2008). His score on the Chinese version of the Childhood Autism Rating Scale (CARS; Lu, Yang, Shu, & Su, 2004; Schopler, Reichler, & Renner, 2002) was 31, indicating mild to moderate autism. He attended second grade in a general education classroom of a public elementary school. He was

able to tact common objects, actions, locations, and occupations. He could also tact the features, categories, and conventional uses of common objects. He could perform single-digit additions and subtractions with carryover. When asked WH questions, Ray was able to recount the stories he had heard. His reading ability was in the average range of his class. He could read and comprehend at least 100 Chinese characters. With regards to non-verbal communication, Ray requested preferred items by pointing, responded to his name with eye contact, and observed the behaviors of his interlocutors. With regards to verbal communication, he requested preferred items with full sentences and asked WH questions. He was able to respond to others' social interactions but rarely initiated social interactions. When asked to explain the meaning of jokes or sarcastic remarks, he could not express his understanding with words.

The second participant, Fu, was a 5-year-old boy with ASD. Fu's diagnosis record indicated that his IQ score was 115, measured by the Chinese version of the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989; Zhang, 2009). His CARS score was 31, indicating mild to moderate autism. He attended a full-day kindergarten program in a general education classroom of a privately funded preschool. Fu could tact common objects, actions, locations, and individual's occupations. When tacting common objects, he also tacted their intended uses, features, and classes. He could count objects and perform single-digit additions and subtractions. He read and comprehended approximately 50 Chinese characters. When asked WH questions about stories, he was not able to accurately recount each story's content. He was observed to engage in solitary play during free play time. When participating in group play, he did not like to take turns with others but would follow teacher directions. He did not have social initiations but would respond to others' initiations when verbally prompted to do so. He responded to and asked WH questions. When given examples of jokes and sarcastic

remarks, Fu was unable to explain their meanings.

The third participant, Yu, was a 7-year-old girl with ASD. She was enrolled in a first-grade general education classroom within a public elementary school. Her record indicated that the CARS score was 30, in the category of mild to moderate autism, but her IQ score had not been measured. According to her teacher, Yu's reading level was equivalent to typically-developing first grade readers. Yu was able to tact common objects, actions, locations, family members, and school teachers. When tacting common objects, she could also tact their common uses, features, and categories. She could read and comprehend at least 50 Chinese characters and recount stories with simple sentences. She imitated actions of others but rarely demonstrated eye contact with people. She did not respond to peers' verbal or non-verbal communications, nor did she initiate social interactions. She followed adult directions individually but did not do so in a group. She was able to answer WH questions but did not ask them. She did not comprehend others' jokes and sarcasm.

### **Setting and Materials**

**Study location.** The study was conducted in one of the individual tutoring rooms in a university-affiliated autism research center in central China. Each room was 2.5m x 2.5m in size with a reading corner, play area, and instruction area. The reading corner had a small couch and a bookshelf; the play area had two shelves with toys and a mat on the floor; the instruction area had a small table with two chairs and a stack of drawers with instructional materials in them. Each room also had a two-way mirror for observation, and a video camera was set on a tripod to record the sessions.

**Target Selection.** The target metaphors were selected from observations of children's conversations in the classroom and from children's books written for elementary-aged Chinese

children. Natural observations were conducted in a kindergarten classroom of a local private preschool. The researchers took field notes of children's conversations involving metaphors. These metaphorical sentences were categorized into (a) metaphors involving physical features of two individuals/objects and (b) metaphors involving abstract properties of two individuals/objects. Two typically developing children (one kindergartener and one first-grader, ages 5 and 6, respectively) recommended by their classroom teachers were interviewed and assessed on their comprehension of these metaphorical language samples. Only metaphors understood and explained well by both children were included in this study. We used 15 metaphors involving physical features and 15 metaphors involving abstract properties as target metaphors. See Appendix 1 for the translation of example metaphors used in this study. The Chinese version and the English translation of all metaphors used in this study are included as supplementary material (see supporting information).

The generalization metaphors were selected from children's books written for first-grade readers. The five paragraphs each included four to five full sentences, two of which were metaphorical. All 10 metaphors in these paragraphs were also tested with two typically developing children in order to ensure the materials were developmentally appropriate. See Table 1 for a sample of generalization paragraphs used in this study. All target metaphors and generalization metaphors were evaluated by a Chinese-speaking behavior analyst based on the behavior-analytic definition of metaphors and examined by a Chinese university professor with a specialty in children's reading to ensure the selected metaphors were appropriate for elementary-aged Chinese children.

**Materials.** The materials used in this study were PowerPoint slides presented on a laptop computer. Each metaphor had one slide consisting of (a) the metaphor presented in textual

stimuli on top, (b) a picture of the first element of the metaphor on the left side beneath the textual stimuli, and (c) a picture of the second element on the right side beneath the textual stimuli. Both pictures could be faded in or out. Textual prompt cards, sized 4 x 6 cm, were used to prompt correct responses. Each card presented a full sentence of the meaning of a metaphor. The generalization metaphors were presented in additional slides. Each slide contained one paragraph with only textual stimuli.

### **Experimental Design**

A combined design of multiple-probe across two behaviors and three participants (Ledford & Gast, 2018) was used to evaluate the effects of the instruction on the student's comprehension of metaphors. The two behaviors consisted of the metaphors with physical features of individuals/objects (Behavior 1,) and the metaphors with abstract properties of individuals/objects (Behavior 2). The sequence of the conditions included baseline, instruction, and follow-up conditions. Probe sessions for all target metaphors were conducted across all conditions. Results of probe sessions were graphed and counted toward criterion. After baseline performance was stable, the instruction was introduced to Behavior 1 of student 1. The mastery criterion for the completion of instructional sessions was that the student achieved 100% for the target metaphors in two consecutive probe sessions.

The instruction condition consisted of three phases (Teach 1, Teach 2, and Teach 3), five target metaphors were presented in each phase. A probe session for the target metaphors was conducted before each instructional session to evaluate student performance. Once the student achieved criterion in Teach 1, the instruction was introduced to Behavior 2 of student 1 and Behavior 1 of student 2. The same sequence was applied to the third student.

After mastery criterion was reached in Teach 1, three probe sessions for all 15 target

metaphors was implemented in Probe 2. If the student responded with 100% accuracy for a to-be-taught target metaphor for two probe trials, that metaphor was removed from instructional sessions.

### **Response Definition and Data Collection**

The dependent variable was the percentage of accurate responses in metaphorical comprehension in probe sessions. A 100% accurate comprehension was defined as, upon hearing the antecedent, the student provided an answer containing (a) the common feature of the two elements and (b) the first element is like the second element. If the student only provided either (a) or (b), the response was coded as 50% accuracy. If an answer did not contain (a) or (b), it was coded as 0% accuracy. For example, the instructor read, “Fang is beautiful. Her eyebrows are two willow leaves. What does this mean?” The 100% accurate comprehension was coded, if the student responded, “It means that her eyebrows are as curvy as willow leaves,” or “Her eyebrows are curvy, and the willow leaves are curvy. So, her eyebrows are like willow leaves.” If the student said, “Her eyebrows and willow leaves are curvy,” or “Her eyebrows look like willow leaves,” it was coded as 50% accuracy. If the student did not respond, repeated one of the sentences, provided irrelevant answers (e.g., “This is wrong. Her eyebrows cannot be willow leaves”), or indicated that he did not understand (e.g., “I’m not sure”), it was coded as 0% accuracy.

### **Procedure**

**Pre-experimental skills assessment.** The students’ skills were assessed using the DBACA (Feng & Sun, 2017). The DBACA is a semi-structured assessment containing five domains: cognitive/language, physical, adaptive/self-help, communication, and social/emotional. The DBACA has demonstrated the content validity evaluated by two experts in developmental

psychology, and the field tests with children ages 3 to 12 years in Taiwan had a high level of inter-scoring agreement ( $M = 92%$ , range: 90-100%). The DBACA was conducted by a trained graduate student of special education.

**Preference assessment.** A preference assessment was conducted based on the procedure described by DeLeon and Iwata (1996) to determine the student's preferred items. The results of the student's preferred activities were used as reinforcers at the end of each instructional session.

**Pre-instruction assessment and training.** After the target metaphors were selected, the students were assessed on their ability to tact the names of the objects involved in target metaphors. They were also tested on their ability to tact at least three features of each individual/object. Students who could not tact the names of the objects and at least three features for each object were taught to mastery (100% accuracy for two probe trials) before baseline data were collected. The pre-instruction assessment/training, probe sessions, and the instruction was conducted by a masters' student of special education in her second year who had basic training in applied behavior analysis for one year.

**Probe sessions across conditions.** Probe sessions for the target metaphors were conducted across baseline, instruction, and follow-up conditions. For each probe trial, the instructor first displayed the slide with only textual stimuli on it, asked the student to read the textual stimuli, and then asked, "What does it mean by saying element 1 is element 2?" or "What does this sentence mean?" The instructor waited 3 s for the student to respond. If the student responded correctly, the instructor gave specific praise (e.g., "You are right. Fang's eye brows are as curvy as willow leaves, and therefore, Fang's eye brows are willow leaves") to reinforce correct responses. If the student did not respond correctly or did not provide any answers within 3 s, the instructor gave general praise (e.g., "Thank you," "I like the way you are listening



nicely,") in order to maintain student participation.

All target metaphors and generalization metaphors were probed to obtain each child's initial performance at baseline (Probe 1). As the instruction was introduced at Teach 1, five target metaphors were presented, and a probe session of these metaphors was conducted before the next instructional session. The probe sessions were conducted in the same manner as baseline. Once these five target metaphors reached criterion performance at Teach 1 (100% accuracy for two consecutive probe sessions), all 15 target metaphors were probed at Probe 2. The same sequence was repeated (Teach 2-Probe 3-Teach 3-Probe 4) until all 15 target metaphors reached criterion performance at Probe 4 (100% for three consecutive probe sessions).

Probe sessions for generalization metaphors were conducted (a) at baseline, (b) immediately after the completion of the instruction, and (c) in the follow-up condition. The probe trials for each generalization metaphor were conducted in the same manner as the probe trials conducted in baseline and instruction conditions, except the instructor displayed a slide, required the student to read the metaphorical sentences, and then asked the meaning of each metaphorical sentence separately.

**Instruction.** Each instructional session consisted of five target metaphors, and each metaphor was presented with three instructional trials randomly rotated with other metaphors. Thus, each instructional session had a total of 15 instructional trials.

There were three segments involved in an instructional trial. Segment 1: The instructor presented a target metaphor (i.e., textual stimuli) on a PowerPoint slide. The instructor required the student to read the sentences aloud. If the student omitted or misread any textual stimulus on the slide, the instructor provided an echoic prompt for the student to continue. After the student read the metaphor, the instructor asked, "What are the two things in the sentences?" and waited 3

s for the student to respond. If the student responded correctly (e.g., “Eyebrows and willow leaves”), the instructor gave praise. If not, the instructor provided an echoic prompt and required the student to repeat correct responses.

Segment 2: The instructor asked, “What is the common feature of the two things?” The instructor then waited for 3 s. If the student responded correctly (e.g., “Her eyebrows and willow leaves are curvy”), the instructor gave praise. If not, the following prompt steps were implemented. Step 1: The instructor added a picture of the first element involved in the metaphor. Step 2: The student was required to tact at least three features of the first element (e.g., “Can you describe the eye brows?” The student’s correct answers were reinforced with praise. The instructor pointed to the picture of the first element and provided echoic prompts if the student did not provide three features (e.g., The eyebrows are curvy, thin, and dark in color). Step 3: The instructor repeated Step 1 and Step 2 with the second element. Step 4: The instructor provided a verbal prompt, “Did you see anything in common for these two things?” and waited 3 s for the student to respond. Again, the instructor provided praise for a correct response and an echoic prompt (e.g., “They both are curvy”) for an incorrect response or no response.

Segment 3: The instructor asked, “What does it mean by saying, ‘Fang’s eyebrows are willow leaves?’” The student’s correct responses were reinforced with praise. If the student did not provide a correct answer or did not respond, the instructor presented a textual prompt card containing a full sentence except for an empty space where the key metaphorical feature would go (e.g., Both Fang’s eye brows and willow leaves are \_\_\_\_\_, so this means that Fang’s eye brows are willow leaves.) to evoke a correct response.

### **Procedural Integrity and Interobserver Agreement**

All sessions were videotaped for the purpose of assessing the procedural integrity and

interobserver agreement (IOA). The procedural integrity and IOA were assessed for 30% sessions randomly selected from each condition for each student using the recording form (Appendix 2). Two graduate students of special education who were experienced in implementing discrete trials served as assessors. The percentage of procedural integrity was calculated using this formula: accurate steps of implementation  $\div$  total steps of implementation  $\times$  100. The procedural integrity of the probe trials across all conditions averaged 95% with a range from 91% to 100%. The procedural integrity of the instructional trials in instructional sessions across the students averaged 97% with a range from 94% to 100%.

The percentage of point-to-point IOA was calculated using the number of agreements divided by the total number of agreements and disagreements and multiplying by 100. The Kappa coefficient ( $k$ ) was calculated with the following formula:  $(P_o - P_c) \div (1 - P_c)$  where  $P_o$  is the proportion of observed agreement and  $P_c$  is the proportion of chance agreement (Ledford, Lane, & Gast, 2018; Watkins & Pacheco, 2000).  $P_c =$

$$\frac{\text{Number of occurrences for observer 1} + \text{Number occurrences for observer 2}}{\text{Total number of trials}^2} +$$

$$\frac{\text{Number of nonoccurrences for observer 1} + \text{Number nonoccurrences for observer 2}}{\text{Total number of trials}^2}. \text{ According to}$$

Cicchetti (1994), the level of agreement is poor, fair, good, and excellent when reliability coefficient is below .40, between .40 and .59, between .60 and .74, and between .75 and 1.00, respectively. The mean IOA for the probe trials across conditions was 97% ( $k = .92$ ) with a range from 87% to 100% (Range  $k$ : .78-1) for all sessions assessed.

### **Social Validity**

Social validity was assessed via a questionnaire and interview questions developed by the researchers of the study. We solicited responses to the questionnaire from each child's parent and interviewed each child and his homeroom teacher. The questionnaire for parents consisted of 15

items, including instruction acceptability (Items 1 to 4, Item 11), perceived helpfulness (Items 5 to 7), feasibility (Items 8 to 10), and satisfaction (Items 12 to 14). Questions about acceptability included the importance, developmental appropriateness, cultural relevance of the instruction, adequacy of the instructional format, and the instructor's skill sets. Questions related to parental perceived helpfulness contained whether the parent perceived that their child had a better understanding of metaphors commonly used in real life and in readings, whether their child learned to explain new metaphors, and whether the instruction was helpful for the parents to discuss metaphors with their children. The parents were asked to consider whether the frequency, location, and schedule of weekly sessions were feasible. Finally, the parents were asked whether they were satisfied with the results, whether they would recommend the instruction to other parents, and whether they perceived their children liked the instruction. Each item was rated on a 5-point Likert scale (1 = not satisfied or not helpful to 5 = very satisfied or very helpful). The last question was an open-ended question for parents to provide suggestions for program improvement, comments for the instruction, and to share their experiences participating in the study. The questionnaire for parents is included as supplementary material (see supporting information).

We also solicited responses from each child in regard to their overall satisfaction of the instruction. There were five yes/no questions, including "Do you like the instructional materials," "Do you like the instructor," "Are you happy that you've learned all these," "If we were to continue, would you like to come," and "Would you recommend your friends to come?"

We also conducted brief interviews with each child's homeroom teacher, none of whom were aware of the research. The interview questions were as follows: In the past few months, "did you notice any difference in the child's overall language comprehension," "any change in

the child's interactions and communication with others," "any change in the child's reading comprehension," and "any other change of the child in other areas of learning?"

### **Results**

Figure 1 depicts the percentage of correct responses for all target metaphors in behaviors 1 and 2 across conditions for the three students.

#### **Acquisition of Target Metaphors**

**Ray.** At Probe 1, Ray had a low level of correct responses for Behavior 1 (range: 5-20%). When the instruction was introduced to five target metaphors for Behavior 1 at Teach 1, his correct responses immediately increased to a mid-level with a rapid ascending trend to criterion in the fourth session (range: 40-100%). At Probe 2, the five taught metaphors were at 100% accuracy, while the 10 other untaught metaphors remained at a low level (range: 15-15%). Similarly, Ray reached criterion for another five target metaphors during three instructional sessions for Teach 2 and maintained at 100% accuracy for Probe 3. He reached the criterion for two untaught target metaphors for Probe 3, and therefore, these two metaphors were removed from Teach 3 instructional sessions. The accuracy of the remaining three target metaphors immediately increased to a high level (range:75-100%) and achieved criterion in three sessions at Teach 3, and maintained at 100% for Probe 4. He also maintained criterion performance of all metaphors in one week, one month, and two months follow-up probe sessions.

Ray's Behavior 2 displayed a similar pattern as Behavior 1. Behavior 2 remained low at baseline (range: 0-20%) but immediately increased to 40% and continued with an ascending trend until reaching criterion in the fifth session of Teach 1. During Probe 2, two of the 10 remaining metaphors achieved criterion and were removed from the instruction. The accuracy of untaught metaphors increased to a mid-level during Probe 2 (range: 40-55%) and a relatively

high level during Probe 3 (range 70-70%). The five target metaphors presented in Teach 2 reached criterion in four sessions, and the three metaphors presented in Teach 3 also reached criterion in four sessions. He maintained all 15 target metaphors in follow-up probe sessions.

**Fu.** Fu had a low level of accuracy for all 15 target metaphors during Behavior 1 at baseline (range: 0%-16%). When the instruction was introduced (Teach 1), he immediately reached criterion for five metaphors in two sessions (range: 100-100%). For Probe 2, he was able to accurately describe 7 of the 10 untaught target metaphors to criterion. Thus, only three metaphors were taught during Teach 2. All 15 metaphors were maintained at a high level for two months following the completion of the instruction for Behavior 1.

Similarly, Fu's accurate responses for the target metaphors during Behavior 2 were at a low level at baseline (range: 0-20%) but immediately increased to a high level and reached criterion after three instructional sessions of Teach 1 (range: 80-100%). At Probe 2, he was able to provide accurate responses for two of the 10 untaught metaphors following Teach 1. After the instruction of another five target metaphors during Teach 2, he also reached criterion for one untaught metaphor during Probe 3. Thus, only two metaphors were presented for Teach 3. Fu maintained all 15 target metaphors at a high level in 2-week and 1-month follow-up probe sessions (range: 91-100%). The accuracy for 12 target metaphors which underwent instruction were maintained at 100% in 2-month follow-up probe sessions but decreased slightly for the three metaphors reaching criterion without instruction for 2-month follow-up probe sessions (range: 75-87%).

**Yu.** For Behavior 1, Yu had a low level of accuracy of all metaphors at baseline (range: 0-23%). After instruction was introduced in Teach 1, the accuracy immediately increased to a mid-level (40%) the first two sessions but then increased to 100% criterion performance in the

4<sup>th</sup> session. During Probe 2, she maintained the acquired five metaphors at 100% and described five untaught metaphors to criterion performance. The remaining five target metaphors were taught in Teach 2 and achieved criterion in two sessions, all 15 target metaphors remained at 100% accuracy at Probe 3 immediately after Teach 2. However, follow-up data were not obtained.

For Behavior 2, Yu's baseline data had a low level of accuracy (range: 0-40%), but the accuracy immediately increased to a relatively high level and reached criterion in five sessions for Teach 1 (range: 80-100%). During Probe 2, she maintained five taught metaphors and accurately described three of the 10 untaught metaphors to criterion performance. Unfortunately, she did not continue the instruction for the remaining metaphors for Behavior 2 due to family matters.

### **Generalization Metaphors**

Figure 2 displays the percentage of correct responses for generalization metaphors before and after the instruction for Ray and Fu who completed the entire study.

**Ray.** Ray had a low level of accuracy for generalization metaphors at baseline (range: 4-13%) but the accuracy increased to a relatively high level (range: 83-86%) after the completion of the instruction. Seven of the generalization metaphors reached criterion (100% accuracy for two probe trials) without instruction. These metaphors were maintained at a similar level for 1-week and 1-month follow-up probe sessions (range: 79-83%) but increased slightly for 2-month follow-up (range: 91-91%), with one generalization metaphor reaching criterion.

**Fu.** Fu had zero accuracy for generalization metaphors at baseline. However, he achieved criterion performance for six metaphors in post-instruction probe sessions immediately following the completion of the instruction (range: 83-88%) and maintained at a similar level for

1-week, 1-month, and 2-month follow-up sessions (range: 66-83%), with two other generalization metaphors reaching criterion.

### **Social Validity**

Two parents completed the social validity questionnaire. The average ratings were 5 ( $SD = 0$ ) on the acceptability, 4.62 ( $SD = 0.48$ ) on the perceived helpfulness, 4.88 ( $SD = 0.33$ ) on the feasibility, and 5 ( $SD = 0$ ) on the satisfaction of the instruction. Ray's mother shared that Ray began to understand some jokes. For example, when she said, "Daddy has a baby in his belly." Ray said, "Daddy's belly is big." Fu's mother mentioned that Fu had applied some metaphors learned from the instruction in their daily conversations several times. Both parents suggested that we continue our study by teaching their children to create metaphorical expressions.

Both children said they liked the instruction and the instructor, were happy about the instruction, would have liked to continue, and would recommend the instruction to their friends. Ray's mother added that Ray was always excited about coming to the instruction and requested to make up if he had to miss any scheduled sessions.

Ray's teacher did not notice any obvious change in Ray's overall language comprehension but reported that Ray had more interactions with the teachers and peers in school. She also said that Ray's reading comprehension was improved but he still needed to improve his expressive writing. Fu's teacher also reported that Fu had more interactions with others and understood others' words better, but his expressions were not always clear. The teacher reported that Fu's improvement in reading comprehension and writing was significant. For example, Fu started to write about his imagination and use metaphors in his writing, which was never observed before.

### **Discussion**



This study applied the analysis of metaphors from the perspective of RFT to teach children with ASD metaphorical comprehension. Overall results indicated that the instruction was effective in increasing accurate responses of metaphorical comprehension in target metaphors for three students. Generalization to untaught metaphors occurred to the two students who completed the study. Results of social validity also indicated that the instruction was acceptable, helpful, and feasible, and both parents and students were satisfied with the results of the instruction.

### **Target Metaphors**

The results of this study were consistent with previous research that children with ASD can acquire metaphorical comprehension and interpret metaphors accurately (Mashal & Kasirer, 2011; Persicke et al., 2013). Different from previous interventions, the instructor did not read the metaphorical sentences but asked the student to read. Next, the instructor used echoic prompts for the two elements in the metaphor, picture prompts for relevant features of the two elements, and textual prompts for the meaning of metaphors. Therefore, the student was guided by the prompts and did not have to handwrite these features as in previous studies. Requiring the student to read the sentences could enhance student attention to the textual stimuli presented without asking additional questions to confirm that the student had attended to the metaphorical sentences, as in Persicke et al. (2013). After the student read the metaphorical sentences, s/he was asked to verbalize the common features of the two elements as well as the unidirectional relation indicated in metaphor (i.e., “Element 1 is like element 2,” but not the reversed direction). The picture prompts provided visual information for the student to first discriminate formal properties of the two elements, and then facilitate the processes of selecting identical feature(s) and excluding non-identical features through the guiding questions provided by the instructor.

That is, picture prompts served to transfer stimulus control from tact to intraverbal, which established and strengthened intraverbal responses to the guiding questions. Following the selection of shared feature(s), textual prompts were used to establish and strengthen the target intraverbal responses consisting of the shared feature(s) and the direction of the relation between two elements. Thus, the textual prompts served to transfer stimulus control from textual to intraverbal responses in the acquisition of metaphorical comprehension. Fostering intraverbal responding is fundamental for adequate social interactions, as many children with ASD lack such skills to initiate, respond, and maintain social conversations with others. Therefore, the design of the instruction included various types of verbal operants (e.g., tact, textual, and echoic) to establish desired intraverbal responding involved in metaphorical comprehension through which a child could potentially improve and engage in more meaningful social interactions.

The success of the instruction was attributed to (a) the analysis of metaphorical comprehension by Skinner's verbal behavior and RFT, (b) the instruction involving intraverbal training with echoic, picture, and textual prompts to establish intraverbal responses involved in metaphors, (c) sufficient exemplars provided during instruction with the requirement to criterion performance, (d) and the target selection based on children's conversations and reading materials.

Consistent with the theoretical background described by Persicke et al. (2013), we also found that the behavior-analytic approach to metaphorical comprehension had practical utility in the development of the instruction. After the student read the sentences, the first question guided the student to tact the two elements in a metaphor (i.e., "What are the two things involved in the target metaphorical sentence?"). Tacting the two elements involved in a metaphor set the occasion upon which the target intraverbal responses in the following guiding questions were

more likely to occur, thereby providing the basis for the extended metaphorical tact (Skinner, 1957). The second question (i.e., “What is the common feature between them?”) was designed based on the equivalence relation leading to comprehension (Stewart & Barnes-Holmes, 2001). When the student did not provide an accurate response, s/he was directed to speak of three associated features for each element with picture prompts. The student then responded intraverbally to the second question regarding the shared feature of the two elements. As discussed, picture prompts facilitated the discriminations between the two elements’ similar and dissimilar features, and established intraverbal responses for the students. Finally, the third question asked the meaning of the metaphor (i.e. “What do you mean by saying the target metaphorical sentence?”). Textual prompts were used to evoke the target intraverbal responses stating the shared feature and the equivalence relation between the two elements. The textual prompts set the occasion to evoke intraverbal responses in regard to the shared feature and the unidirectional equivalence relation between the two elements leading to metaphorical comprehension. All three children needed picture and textual prompts to establish and strengthen target intraverbal responding to the questions in the beginning stage of instruction, suggesting that picture and textual prompts were important components in the acquisition of metaphorical comprehension. It is possible that visual stimuli were necessary for children with ASD to establish intraverbal responding under verbal stimulus control provided by the instructor, which is a prerequisite skill for metaphorical comprehension. Such visual prompts were no longer needed once the desired intraverbal responding was established.

The guiding questions consisted of (a) tacts of the elements and (b) the relational tact of the shared feature, which were identified as critical components in analogical reasoning (Miguel et al., 2015). The behavioral processes of relating the relations in metaphors and analogies are

similar. One difference between metaphors and analogies is that metaphors involve a unidirectional relation while analogies involve a bidirectional relation between two elements (Stewart & Barnes-Holmes, 2001). In this study, the unidirectional relation between two elements in a metaphor was established via the guiding question (i.e., “What do you mean by saying element 1 is element 2?”) and textual prompts.

The emergence of metaphorical comprehension for untaught metaphors in Behavior 1 and Behavior 2 occurred after acquiring 5 to 10 target metaphors in the same category for all three children, suggesting the occurrence of generalization. However, such generalization was limited as the number of untaught target metaphors reaching criterion remained low, and the overall accuracy for the untaught metaphors was lower than the criterion level as shown in Probe 2 and Probe 3. Instruction was necessary to increase accurate responses for the remaining target metaphors in both categories. Therefore, it was necessary to include at least 15 target metaphors in each category.

### **Generalization Metaphors**

The results of the pre- and post-instruction probe sessions on generalization metaphors for Ray and Fu indicated that a generalized understanding of metaphors in both categories occurred after the completion of the instruction. Further, generalization metaphors were maintained for two months after the instruction. The generalized understanding of metaphors occurred only after a sufficient amount of exemplars were established, which was consistent with previous studies on multiple exemplar instruction (Greer & Speckman, 2009; Greer & Ross, 2008). In this study, generalization was demonstrated for two children after 15 metaphors with physical features and 15 metaphors with abstract properties were acquired, suggesting that a total of 30 metaphors were required for generalized understanding of metaphors.

Target selections based on descriptive observations and children's books are developmentally appropriate because metaphors commonly used and understood by younger children are different from older children. This approach also ensures that the selected target metaphors have suitable cultural content the children are more likely to encounter in their environment.

There are some potential limitations of this study. First, the students' skills in word associations or providing multiple responses to a single question were not evaluated. Generating multiple features for each element is an intraverbal behavior requiring divergent control, a repertoire that is weak in many children with ASD and need to be specially targeted in instruction (Michael, Palmer, & Sundberg, 2011). If children fail to generate multiple word associations, they would not find the shared associations required to form a non-arbitrary equivalence relation between the two elements. However, this speculation is not verified in this study and thus warrants further investigations. Secondly, the experimental design does not allow component analysis, and therefore, we are not certain whether echoic, picture, or textual prompts were necessary components responsible for metaphorical comprehension. Thirdly, the students were instructed to select the shared feature between the two elements and form an equivalence relation between two elements to understand metaphors. However, the instruction did not include the skills of identifying a more salient property for element 1 and transferring it to element 2, which is an important ability to comprehend more complex metaphors and metaphorical creations (Cortes, Cobos, & Tarbox, 2018). Future researchers can consider including such skills in the instruction to enhance metaphorical comprehension. Finally, only two students completed the entire study. The third student did not complete the instruction, and therefore, the acquisition and generalization effects of the instruction for this student were not evaluated.

In summary, this study used a behavioral analytic approach containing intraverbal training with echoic, picture, and textual prompts to teach metaphorical comprehension to children with ASD. The results indicated that children with ASD were able to acquire the skills to understand metaphorical language used in typically developing children's daily conversations and children's books. It is important to target metaphorical comprehension to strengthen effective social communication in interventions for children with ASD.

### **Compliance with Ethical Standards**

#### **Disclosure of Potential Conflict of Interest**

All authors involved in this study do not have any interests that might be interpreted as influencing the research.

#### **Ethical Approval**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

#### **Informed Consent**

Informed consent was obtained from the parents of the children included in the study.

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington, DC: Author.
- Braam, S. J., & Poling, A. (1983). Development of intraverbal behavior in mentally retarded individuals through transfer of stimulus control procedures: classification of verbal responses. *Applied Research in Mental Retardation*, 4(4), 279–302.
- Cameron, L. (2003). *Metaphor in educational discourse*. London ; New York: Continuum.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.
- Cortes, A. I. R., Cobos, F. J. M., & Tarbox, J. (2018). Teaching children to create metaphorical expressions. *International Journal of Psychology and Psychological Therapy*, 18(1), 27–38.
- DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis*, 29(4), 519–533.  
<https://doi.org/10.1901/jaba.1996.29-519>
- Feng, H., & Sun, W.-C. (2017). *Tzu-pi cheng fa chan pen wei ping liang hsi tung [Developmentally-based behavior assessment system for children with autism]*. Taipei, Taiwan: Hua-Teng Publisher.
- Goldsmith, T. R., LeBlanc, L. A., & Sautter, R. A. (2007). Teaching intraverbal behavior to children with autism. *Research in Autism Spectrum Disorders*, 1(1), 1–13.  
<https://doi.org/10.1016/j.rasd.2006.07.001>
- Greer, R. Douglas, & Speckman, J. (2009). The integration of speaker and listener responses: A



- theory of verbal development. *Psychological Record*, 59(3), 449–488.
- Greer, Robert Douglas, & Ross, D. E. (2008). *Verbal Behavior Analysis: Inducing and expanding new capacities in children with language delays*. Boston, Mass.; London: Allyn and Bacon.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: a post-Skinnerian account of human language and cognition*. New York: Kluwer Academic/Plenum Publishers.
- Ledford, J. R., & Gast, D. L. (2018). Combination and other designs. In J.R. Ledford and D. L. Gast (Eds.) *Single Case Research Methodology: Applications in special education and behavioral sciences* (3rd ed., pp. 334–364). New York: Routledge.
- Ledford, Jennifer R., Lane, J. D., & Gast, D. L. (2018). Dependent variables, measurement, and reliability. In J. R. Ledford and D. L. Gast (Eds.) *Single Case Research Methodology: Applications in special education and behavioral sciences* (3rd ed., pp. 97–132). New York, NY: Routledge.
- Lu, J. P., Yang, Z. W., Shu, M. Y., & Su, L. Y. (2004). Er tong gu du zheng liang biao ping ding de xin du xiao du fen xi [Reliability and validity analysis of the Childhood Autism Rating Scale]. *Zhong Guo Xian Dai Yi Xue Za Zhi*, 14(3), 119–123.
- Luciano, M. C. (1986). Acquisition, maintenance, and generalization of productive intraverbal behavior through transfer of stimulus control procedures. *Applied Research in Mental Retardation*, 7(1), 1–20.
- Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities*, 32(6), 2045–2054. <https://doi.org/10.1016/j.ridd.2011.08.012>

- Melogno, S., D’Ardua, C., Pinto, M. A., & Levi, G. (2012). Explaining metaphors in high-functioning Autism Spectrum Disorder children: A brief report. *Research in Autism Spectrum Disorders*, 6(2), 683–689. <https://doi.org/10.1016/j.rasd.2011.09.005>
- Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior*, 27(1), 3–22.
- Miguel, C. F., Frampton, S. E., Lantaya, C. A., LaFrance, D. L., Quah, K., Meyer, C. S., ... Fernand, J. K. (2015). The effects of tact training on the development of analogical reasoning. *Journal of the Experimental Analysis of Behavior*, 104(2), 96–118. <https://doi.org/10.1002/jeab.167>
- Persicke, A., Tarbox, J., Ranick, J., & St. Clair, M. (2012). Establishing metaphorical reasoning in children with autism. *Research in Autism Spectrum Disorders*, 6(2), 913–920. <https://doi.org/10.1016/j.rasd.2011.12.007>
- Rundblad, G., & Annaz, D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, 14(1), 29–46. <https://doi.org/10.1177/1362361309340667>
- Schopler, E., Reichler, R. J., & Renner, B. R. (2002). *The childhood autism rating scale (CARS)*. Los Angeles: Western Psychological Services.
- Semino, E., & Demjén, Z. (Eds.). (2016). *The Routledge handbook of metaphor and language*. Milton Park, Abingdon, Oxon; New York, NY: Routledge.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Stewart, I., & Barnes-Holmes, D. (2001). Understanding metaphor: A relational frame perspective. *The Behavior Analyst*, 24(2), 191–199.
- Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research:

Importance and calculation. *Journal of Behavioral Education*, 10(4), 205–212.

<https://doi.org/10.1023/A:1012295615144>

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children® - Fourth Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, David. (1989). *Wechsler Preschool and Primary Scale of Intelligence-Revised*. San Antonio, TX: Psychological Corporation.

Zhang, H. C. (2008). *Wei shi er tong zhi li liang biao-di si ban [The Wechsler Intelligence Scale for children--4th edition]*. Zhuhai, Guangdong: King-May Psychological Assessment Technology Development, Ltd.

Zhang, H. C. (2009). Wei shi you er liang biao xiu gai ban de xin li xue ce liang fen xi [Psychometrical analysis of Chinese version of Wechsler Preschool and Primary Scale of Intelligence-Revised]. *Chinese Journal of Psychological Science*, 15(5), 1–13.

Appendix 1. *Examples of metaphors used in this study*

<b>Metaphors</b>	<b>Element 1</b>	<b>Element 2</b>	<b>Common feature</b>	<b>Meaning</b>
<b>Metaphors involving physical features</b>				
Fang Fang is very beautiful. Her eyebrows are two willow leaves.	eyebrows	willow leaves	curvy	Fang Fang’s eyebrows and willow leaves are curvy. Her eyebrows are like willow leaves.
It is dark outside, the curved moon hung in the sky with lots of twinkling stars. Doudou said that the stars are gems.	stars	gems	twinkling	The stars are twinkling, and the gems are also twinkling. The stars look like gems.
The class is still very noisy after the bell; some people were chatting and others were arguing. The teacher came in and said, "How did the classroom become a marketplace?"	classroom	marketplace	noisy	The classroom is very noisy, and the marketplace is usually very noisy. The teacher feels that the classroom is as noisy as the marketplace.
<b>Metaphors involving abstract properties</b>				
Mumu is a warm-hearted primary school student. His classmates usually ask him for help when they have difficulties. Everyone says he is a Superhero.	Mumu	Superman	helps others	Mumu helps others. Superheroes also helps others. Mumu is very much like a Superhero.

Zhou likes to punch people, and his classmates are scared of him. When he comes to school, his classmates shout, "The big wolf is coming!"	Zhou	big wolf	scary	Zhuang Zhuang often beats his classmates; everyone is scared of him. Wolves also hurt people so people are scared of them. Zhuang Zhuang is a wolf.
Spring has arrived, the ice and snow have melted, the grass has emerged from the ground, and the earth turned green. Spring is a magician.	spring	magician	changing things	When spring comes, there will be many changes. Magicians can also make many changes. Spring is very much like a magician.
<b>Generalization Metaphors</b>				
The round sun is about to set, and the fallen leaves are butterflies that slowly fall from the tree. The little mermaid emerged lightly from the surface of the water. She saw the happy dolphins flipping their arms, and the huge whales spewed water from the nostrils and turned into a moving fountain.	fallen leaves	butterflies	falling or flying around the trees	The leaves fall from the tree. The butterflies dance between the trees. As they are both very light, fallen leaves look like butterflies.
	whale	water fountain	sprays water	Whales can spray water from their nostrils; fountains can also spray water from the nozzle. As both the whale and fountain have this ability, they are similar.

Appendix 2. *Procedural integrity and interobserver agreement data collection form*

	Antecedent (A), student response, Consequence (C)	Trials (+ correct; - incorrect)									
		1	2	3	4	5	6	7	8	9	10
Probe trials	A: Present metaphor, Ask student to read, “What does this sentence mean?” Wait 3 s										
	Student response										
	C: Praise correct responses, Ignore incorrect responses										
Instructional trials	Segment 1										
	A: Present metaphor, ask student to read. Echoic prompts for incorrect responses, if any “What are the two things?” Wait 3 s										
	Student response										
	C: Praise correct responses. Echoic prompts for incorrect responses										
	Segment 2										
	A: “What do these two things have in common?” Wait 3 s										
	Student response										
	C: Praise correct response Prompts for incorrect responses: See *										
	Segment 3										
	A: “What does it mean by saying ‘_ is _’ in this sentence?”										
	Student response										
	C: Praise correct responses Textual prompt for incorrect responses										
*First, present picture of element 1, “Describe it.” Provide echoic prompts for three features. Have the student echo correct responses. Second, present picture of element 2, “Describe it.” Provide echoic prompts for three features. Have the student echo correct responses. Third, present a verbal prompt, “Did you see anything in common?” Wait 3 s. Provide an echoic prompt for the common feature. Have the student echo correct responses.											

*Note. The A and C codes were used to calculate the percentage of accurate implementations to assess procedural fidelity. The student response codes were used to calculate interobserver agreement. Only independent student responses were recorded as “+,” all prompted student responses were recorded as “-.”*