

# Investigating Relational Framing of Categorization in Young Children

Teresa Mulhern<sup>1</sup> · Ian Stewart<sup>1</sup>  · John Mc Elwee<sup>2</sup>

Published online: 18 September 2017  
© Association for Behavior Analysis International 2017

**Abstract** The aims of the current study were to measure patterns of relational framing linked with categorization in young, typically developing children and to correlate framing performance with linguistic and cognitive potential as measured by standardized instruments, including the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), the Stanford–Binet Intelligence Scales—Fifth Edition (SB5), and the Children’s Category Test (CCT). The relational protocol developed for this study assessed properties of relational framing in 3 relational domains, including nonarbitrary and arbitrary containment relations and arbitrary hierarchical relations. There were 50 participants, 10 from each of the following age ranges: 3–4, 4–5, 5–6, 6–7, and 7–8. The results provided data concerning the acquisition of relational categorization skills across childhood and also showed strong correlations between relational performance and that on each of the 3 standardized measures. The results are discussed in relation to previous research and for their implications in regard to future studies on relational framing and categorization in children.

**Keywords** Relational frame theory · Hierarchy · Containment · Intelligence · Language · Classification

Classification or categorization involves grouping stimuli into distinct classes or categories based upon common functions or physical features (Astley, Peissig, & Wasserman, 2001;

Zentall, Galizio, & Critchfield, 2002). In hierarchical classification, classes are themselves categorized into higher order classes (Greene, 1994; Murphy, 2002; Slattery & Stewart, 2014; Slattery, Stewart, & O’Hora, 2011). An example of the latter would be classifying *budgie* into the category *bird* and classifying *bird* into the category *animal*.

The bulk of research examining classification as a skill set has been conducted within the field of cognitive–developmental psychology (e.g., Arterberry & Bornstein, 2012; Blewitt, Michnick, Golinkoff, & Alioto, 2000; Bornstein & Arterberry, 2010; Deneault & Ricard, 2005, 2006; Hajibayova, 2013; Manders & Hall, 2002; Mompeán, 2006; Rostad, Yott, & Poulin-Dubois, 2012; Valentin & Chanquoy, 2012). Research within this field suggests that this skill set is acquired gradually, beginning within early childhood (3–4 months; see Arterberry & Kellman, 2016; Eimas & Quinn, 1994; Eimas, Quinn, & Cowan, 1994; Quinn, 2016; Quinn & Eimas, 1996; Song, Pruden, Michnick Golinkoff, & Hirsh-Pasek, 2016) with basic classification skills and culminating later in childhood (approximately eleven years old; see Blewitt, 1989, 1994; Carneiro, Albuquerque, & Fernandez, 2009; Deneault & Ricard, 2006) with advanced “hierarchical” classification responding.

Although cognitive–developmental research on classification has identified several crucial features of classification and has generated data regarding variables that may affect it, this approach is largely descriptive. The principal aim of such research is to gain a theoretical comprehension of the phenomenon in question rather than to isolate environment–behavior interactions that can facilitate prediction and influence over behavior. As such, theoretical accounts offered within the cognitive–developmental approach can be argued to be less than optimal for realizing practical change within the applied domain (Margolis & Laurence, 2000; Murphy, 2002; Palmer, 2002).

✉ Ian Stewart  
ian.stewart@nuigalway.ie

<sup>1</sup> School of Psychology, National University of Ireland Galway, Galway, Ireland

<sup>2</sup> Department of Special Education, Wilkes University, Wilkes-Barre, PA, USA

The approach to classification taken in the current study is based on a contextual behavioral approach referred to as relational frame theory (RFT; Dymond & Roche, 2013; Hayes et al., 2001a). This approach conceptualizes language and cognition as patterns of generalized, contextually controlled relational responding that are learned via multiple exposures to behavior–environment contingencies operating in the socioverbal environment. These patterns are known as relational frames. Acquisition of these frames begins with learning to respond to nonarbitrary or physical relations between objects in the presence of particular contextual cues (e.g., choosing the physically bigger or smaller of two objects in the presence of the cues *more* and *less*, respectively). With sufficient exemplars of the pattern, however, this responding can come under the control of the contextual cues alone and is then applicable in contexts without nonarbitrary relational support (e.g., if told that X is more than Y, a child might be able to derive that Y is less than X in the absence of any nonarbitrary or formal differences between the relata). The latter is referred to as arbitrarily applicable relational responding (AARR) because the relational response can be applied in any circumstance no matter what the actual physical relation between the relata is (Hayes et al., 2001b; O’Toole, Barnes-Holmes, Murphy, O’Connor, & Barnes-Holmes, 2009; Stewart & McElwee, 2009).

Over the last number of decades, researchers have shown that patterns of AARR (i.e., relational frames) come in a variety of forms (see Stewart, 2016, and Stewart & Roche, 2013, for an overview). Despite the variety of patterns of AARR, all are characterized by three properties: mutual entailment, combinatorial entailment, and transformation of function. *Mutual entailment* (ME) refers to the feature of relational framing by which a unidirectional relation from Stimulus A to Stimulus B in a specific context entails a second unidirectional relation from Stimulus B to Stimulus A. For instance, in the earlier example, a child told that “X is more than Y” could derive the mutually entailed relation that “Y is less than X.” *Combinatorial entailment* (CE) is the phenomenon whereby two stimulus relations are combined to allow derivation of a third relation. For example, a child told that “X is more than Y” and “Y is more than Z” might derive that “X is more than Z” and “Z is less than X.” Finally, the *transformation of functions* (ToF) is seen when the functions of a stimulus change or transform as a result of its being in a derived relation with another stimulus. For example, consider the case of the child who has derived through CE that X is more than Z. If Z was already a directly conditioned reinforcer for this child, then he or she might work for it; however, if offered a choice between X and Z, the child might choose the former (i.e., X) over the latter (i.e., Z), even though he or she might have had no previous experience of X as a reinforcer. Such a preference would demonstrate that the functions of X had been transformed through its derived relations with Z (e.g., Hayes et al., 2001b).

The generativity of relational framing illustrated by these properties is one source of evidence suggesting the link between AARR and human language and cognition. However, there is by now substantive evidence from many other quarters also showing this link and, further, showing that AARR in its multiple forms appears to be the key functional process underlying these phenomena (e.g., Cassidy, Roche, Colbert, Stewart, & Grey, 2016; Gore, Barnes-Holmes, & Murphy, 2010; Hayes & Stewart, 2016; Healy, Barnes-Holmes, & Smeets, 2000; Luciano et al., 2007; O’Hora et al., 2008; Viscaíno-Torres et al., 2015).

From this perspective, classification as a form of verbal behavior may also be approached in terms of particular patterns of framing. Perhaps the most relevant frames in this regard are those of containment (e.g., A contains B; B is within A) and hierarchy (e.g., A is a member of B; B is a class containing A; see, e.g., Hayes et al., 2001a; Slattery & Stewart, 2014). These patterns of framing, as with others, can be theorized to originate in more basic nonarbitrary relations. For example, in the case of these particular patterns of relational responding, a child might be taught to put one object physically inside another in the presence of a cue such as *in* or *inside* or to accurately state which object in a container–contained dyad is the former and which the latter. Such responding, based as it is on the presence of actually physically related objects, might then lay the foundation for relatively more abstract patterns of responding, such as classification (i.e., responding to “members” as being contained in “classes”) or hierarchical classification (classes being contained within classes) in which immediate physical relations (e.g., of containment) are no longer present.

Gil et al. (2012) provided the first empirical demonstration of hierarchical relational responding. The participants were 10 university students, and the experiment was conducted across five phases. In Phase 1, four arbitrary shape stimuli were established as *includes*, *belongs to*, *same*, and *different* contextual cues, respectively. In Phase 2, participants were trained and tested for the derivation of arbitrary sameness (i.e., “equivalence”) relations between arbitrary nonsense-word stimuli. These stimuli thus made up the lowest level of two separate hierarchical relational networks. Next, in Phase 3, responding in accordance with higher levels of these networks was induced by having participants derive relations of containment between lower level stimuli and (novel, additional) stimuli further up the network by using the *includes* and *belongs to* cues. In Phase 4, the researchers established particular functions in particular stimuli at different levels of the hierarchical network, and in the final phase they reported particular resultant patterns of ToF whereby stimuli acquired novel untrained functions by virtue of their position in the hierarchical network.

Slattery and Stewart (2014) subsequently argued that hierarchical relational framing subsumes more than one pattern of

hierarchical responding and that these patterns include both hierarchical classification—in which stimuli are related as being in classes and subclasses—as well as hierarchical containment—in which stimuli are related as simply containing and being contained, without full implications of classification being involved (see Markman & Seibert, 1976). They sought to demonstrate a model of hierarchical relational framing as classification per se. Accordingly, they used nonarbitrarily related stimuli (i.e., sharing certain physical properties, as typically seen within classification hierarchies) to establish two arbitrary shapes as contextual cues for *member of* and *includes* relational responding with four undergraduates. Thereafter, these cues were then used to establish arbitrary stimuli in particular hierarchical relations. Finally, the derivation of additional untrained hierarchical relations and ToF was assessed. Subsequent patterns of responding showed properties of unilateral property induction, transitive class containment, and asymmetrical class containment, as predicted and in keeping with properties ascribed to hierarchical classification by mainstream theorists.

The aforementioned studies constitute an important starting point with respect to the RFT induction and investigation of hierarchical relational framing. They provide evidence that humans engage in this particular pattern of contextually controlled relational responding and that it can be brought under contextual control in the laboratory. One important direction in the further empirical investigation of this pattern of relational framing is to begin to examine its historical origins. That is the aim of the current study, which seeks to measure the emergence of relational framing related to classification, including both containment and hierarchy, in young children from 3 to 8 years old and to correlate performance in this regard with performance on standardized measures of intellectual potential.

The current study involved using a protocol developed specifically to track repertoires of relational framing related to hierarchical classification from simple and concrete up to complex and abstract; specifically, nonarbitrary containment, arbitrary containment, and arbitrary hierarchy, each assessed in terms of both ME and CE, as well as, in the case of arbitrary hierarchy, ToF. We anticipated that doing so might provide information regarding the typical acquisition of relational framing repertoires relevant to classification. This might further inform researchers and practitioners regarding the development of relational framing in general as well as framing related to hierarchical classification specifically. It might provide information regarding typical hierarchically relevant framing repertoires at different ages as well as potential deficits at particular ages that might be amenable to training. As indicated earlier, the results of previous research suggest that relational framing is positively correlated with performance on measures of intellectual potential. It might be hypothesized that the framing repertoires of containment and hierarchy

specifically would be similarly correlated. Assuming that this is the case, such a finding might point to the potential importance of training relational framing skills of categorization using a protocol based on that developed for the current study.

To assess nonarbitrary containment, we presented children with differently colored boxes in which smaller boxes were physically contained inside larger boxes and asked the children whether one particular box was inside another or contained the other. To assess arbitrary containment, we presented the children with a number of differently colored circles that were physically the same size as each other. We then told them about one or more “containment” relations between the circles and probed to see whether they could correctly derive further relations in the absence of any physical containment relations being demonstrated (hence, these tasks probed for arbitrary relations). For example, for ME tasks, we told them that one particular colored circle (e.g., the green one) contained another particular colored circle (e.g., the red one) and then probed to see whether they could derive the correct entailment relation (i.e., in this case, that the red one was inside the green one). Finally, to assess arbitrary hierarchy, we presented children with real and nonsense words on a computer screen, told them about one or more “hierarchical class”-type relations between the stimuli, and probed to see whether they could correctly derive further relations. For example, for ME tasks, we told them that one particular nonsense word (e.g., *tol*) was a type of animal and then probed to see whether they could endorse the correct entailment relation (i.e., in this case, that the class *animal* contained *tols* as members). As for the previous task, this was an arbitrary relational task because the stimuli representing classes and members did not show any physical relationship of relevance to the task.

To assess whether the relational repertoires measured in this study correlated with intellectual performance, a number of standardized tests of intellectual skill were used, including the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007), the Stanford-Binet Intelligence Scales—Fifth Edition (SB5; Roid, 2003), and the Children’s Category Test (CCT; Boll, 1993). A test of Piagetian class inclusion (CI; Piaget, 1952) was also used to examine whether scoring on the relational framing protocol might predict performance on this well-known test of classification.

## Method

### Participants

Fifty typically developing children (23 female, 27 male) between 3 and 8 years of age (range of 3 years and 0 months to 7 years and 9 months) took part. The mean age and standard deviation for each group ( $n = 10$  each) were as follows: 3-year-olds— $M = 3.56$  years,  $SD = 0.26$  years; 4-year-olds— $M$

= 4.56 years,  $SD = 0.27$  years; 5-year-olds— $M = 5.54$  years,  $SD = 0.29$  years; 6-year-olds— $M = 6.31$  years,  $SD = 0.28$  years; and 7-year-olds— $M = 7.27$  years,  $SD = 0.27$  years.

Participants were recruited from rural primary schools and play schools within the local area. Prior ethical approval for recruitment of participants for this study was obtained from the research ethics committee of the host institution. Consent for conducting the study was obtained from the principal in each respective school. Parental consent was obtained for each child who participated, and verbal consent was also obtained from each of the participants.

## Materials

Participants were exposed to a number of different assessments. These included preassessment of (a) color tacting and (b) yes–no responding. They were also assessed on a number of standardized measures as follows: (a) the PPVT–4 (Dunn & Dunn, 2007), (b) the SB5 (Roid, 2003), and (c) the CCT (Boll, 1993). The assessments also included a test of CI (Piaget, 1952) and a test of relational responding involving three parts: (a) nonarbitrary containment, (b) arbitrary containment, and (c) arbitrary hierarchy.

**Color Tacting** This test involved nine paper circles, each of which was colored with a different color used during the relational assessment procedure (i.e., black, white, orange, green, blue, yellow, purple, red, or pink). During the test, the child was exposed to three arrays of three colored circles each and asked to tact each particular color displayed. Children had to answer correctly on all nine trials to proceed.

**Yes–No Responding** This test involved 10 pictures of commonly seen objects or animals (e.g., a dog). On each trial, the child was shown a picture of an object or animal and asked a yes–no question. Children were asked five questions to which the answer *yes* was appropriate (e.g., they were shown a picture of a dog and asked “Is this a dog?”) and five questions to which the answer *no* was appropriate (e.g., they were shown a picture of a circle and asked “Is this a square?”). Children had to answer correctly on all 10 trials to proceed.

**PPVT–4** This test assesses receptive vocabulary in respondents from ages 2 years and 6 months to late adulthood and is often used as a test of scholastic aptitude. It is administered individually and is presented in a multiple-choice format in which the respondent is presented with four pictures and asked to select the one that best illustrates the definition of a particular word. The stimuli include items representing up to 20 content areas (e.g., actions, vegetables, tools) and components of speech (nouns, verbs, or attributes), encompassing a broad range of difficulty. Test–retest reliability has been shown to be between .92 and .96 (Community-University Partnership for

the Study of Children, Youth, and Families, 2011), and internal consistency has been noted to be similarly high (between .94 and .95). Construct and convergent validity has been assessed by comparing the PPVT–4 to the Expressive Vocabulary Test, Second Edition (EVT–2; Williams, 2007). Correlations between the two were high ( $r = .80$ – $.84$ ) for all age groups (Community-University Partnership for the Study of Children, Youth, and Families, 2011).

**SB5** This is an intelligence test for use with individuals from ages 2 to 85 years that measures five weighted factors (fluid reasoning, knowledge, quantitative reasoning, visual–spatial processing, and working memory) and consists of both verbal and nonverbal subtests. Participants in the current study were evaluated for the full-scale intelligence quotient (FSIQ), the verbal intelligence quotient (VIQ), and the nonverbal intelligence quotient (NVIQ). Reliability coefficients have been found to be extremely high for the FSIQ ( $r = .98$ ), the NVIQ ( $r = .95$ ), and the VIQ ( $r = .96$ ), showing excellent stability. In addition, the five factor index scores were all above .90 and were higher than the subtest scales, which were, however, comparable to other cognitive tests with ranges from .84 to .89 (Roid, 2003). Validity has been assessed through comparison with other tests of cognitive and intellectual ability, including the Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI–R; Wechsler, 1989), the Wechsler Intelligence Scale for Children—Third Edition (WISC–III; Wechsler, 1991), the Wechsler Adult Intelligence Scale—Third Edition (WAIS–III; Wechsler, 1997), the Woodcock–Johnson III Test of Cognitive Ability, and the Woodcock–Johnson III Test of Achievement. Correlations ranged from .78 to .84 for both FSIQ and VIQ with comparable indices on other major IQ batteries (Roid, 2003).

**CCT** This test was designed to assess categorization skills for children between 5 and 8 years of age (Level 1) and between 9 and 16 years of age (Level 2). It is individually administered and is designed to assess concept formation and problem-solving abilities with novel material. In the current study, only Level 1 was used. This level consists of five subtests and includes 80 questions. Within each subtest, children are asked to determine the principle underlying correct performance (i.e., matching based on color, size, or proportion) using examiner feedback. Previous research has indicated that the test has adequate test–retest reliability ( $r = .75$ ) in addition to strong internal consistency for both Levels 1 ( $r = .88$ ) and 2 ( $r = .86$ ; MacNeill Horton, 1996).

**Piagetian CI** This is a classic test of categorization ability that assesses the capacity to respond to a stimulus as simultaneously belonging to both a class and a superordinate class (Thomas & Horton, 1997). It has its origins in Piagetian developmental

psychology (Piaget, 1952), where it was seen as determining if a child had reached the “concrete operational” stage of development, hypothesized to occur by age 7 or 8. In the test, the child is first shown an array of stimuli that belong to a particular category. These stimuli form two different subclasses, with a greater number of stimuli within one category than the other (e.g., five apples and two pears). The child is asked if there are more of the larger subclass or of the category—for example, “Are there more apples, or is there more fruit?” Materials used in the current study included  $5.5 \times 5.5$  cm colored flashcards depicting a variety of animals (horse, dog, pig, cow, cat, dog), fruit (strawberry, apple, pear, banana, lemon, orange), clothing (socks, skirt, T-shirt, dress, pants, coat), and vehicles (car, fire engine, truck, motorcycle, tractor, bus). Multiple examples of each of these stimuli were used.

**Relational Responding Test** The current study assessed three different patterns of relational responding, including (a) non-arbitrary containment, (b) arbitrary containment, and (c) arbitrary hierarchy. These were assessed as follows:

1. Nonarbitrary containment: Materials included both square and triangular boxes of different sizes (large, medium, or small) and colors (red, yellow, green, blue, purple, orange, black, white, and pink).
2. Arbitrary containment: Materials included same-sized circles and same-sized triangles of differing colors (red, yellow, green, blue, purple, orange, black, white, and pink).
3. Arbitrary hierarchy: This test used a laptop to show arrays of stimuli that were used as the basis for asking particular questions to probe for derived relations.

## Procedure

In the case of each school setting, the research was conducted in a separate classroom or resource room within the school building. In the first (screening) session, children were exposed to the color tacting and yes–no pretests, which together took a total of 10 min per child. All the children passed these tests. Thereafter, they were tested individually on the main assessments in sessions lasting between 30 and 45 min, with the length of a session depending on the test being administered and the age of the child. These sessions involved breaks every 10 or 15 min, also depending on the age of the child. Children were exposed to the main assessments in the following order, with the number of sessions involved for each measure and the range of testing time per session, including breaks, given in parentheses:

1. Relational Responding Test 1 (three sessions lasting 40–45 min each, one for nonarbitrary containment, one for arbitrary containment, and one for arbitrary hierarchy).

2. SB5 (two to three sessions, with 40 min for Session 1, 25–30 min for Session 2, and 20–25 min for Session 3, with the number of sessions and exact duration depending on the age of the child).
3. PPVT–4 (one session, 30–40 min).
4. CCT (one session, 30 min).
5. CI (included in the same session as the CCT, 10 min).
6. Relational Responding Test 2 (number and length of sessions were the same as for Relational Responding Test 1).

Testing sessions took place over two phases. Phase 1, which included Assessments 1–5, involved approximately seven to eight days of testing per child spread out over a duration of approximately 2 weeks. Phase 2, which included Assessment 6 (i.e., the second relational responding test), took place 2 weeks after the initial phase ended and involved a successive 3 days of testing.

**Relational Responding Test** In their first session after passing the screening tests, participants were assessed for the three relational responding repertoires seen as involved in classification: nonarbitrary containment (i.e., relating stimuli on the basis of an observed physical containment relationship), arbitrary containment (i.e., relating stimuli on the basis of an abstract or arbitrary containment relationship), and arbitrary hierarchy (i.e., relating stimuli on the basis of an abstract hierarchical class relationship). These three relational repertoires were assessed over the course of three sessions (one repertoire per session) based on 128 questions (32 for each of the first two repertoires and 64 for the third).

In the case of nonarbitrary and arbitrary containment, relational framing was assessed in terms of ME (16 questions) and CE (16 questions; see Tables 1 and 2, respectively). On all trials, participants were first shown the relevant stimuli, and then the experimenter described the relationship between the stimuli. For the assessment of nonarbitrary containment responding, the experimenter further demonstrated the relationship by manipulating the stimuli (e.g., following the description “A red box is inside a blue box,” the experimenter would then place a red box inside a blue box). Participants were then asked a question about the relationship between certain stimuli that could typically be answered in the form of a *yes* or *no* response (more specific details regarding format are provided later in subsections corresponding to each repertoire). No contingent feedback was provided at any time.

In the case of arbitrary hierarchy, relational framing was assessed in terms of ME (16 questions), ToF through ME (16 questions), CE (16 questions), and ToF through CE (16 questions). On all trials, participants were first shown an array of stimuli on a computer screen, and then the experimenter described the relationship between the stimuli, elaborating slightly for ToF questions. As in previous sections, participants were then presented with a follow-up question that could

**Table 1** Trial Types Used in the Nonarbitrary Containment Phase

Property	Statement	Question [answer]
Mutual entailment (two stimuli)	The [A] box is inside the [B] box.	Is the [A] box inside the [B] box? [Yes]
	The [A] box is inside the [B] box.	Is the [B] box inside the [A] box? [No]
	The [A] box is inside the [B] box.	Does the [A] box contain the [B] box? [No]
	The [A] box is inside the [B] box.	Does the [B] box contain the [A] box? [Yes]
	The [B] box is inside the [A] box.	Is the [A] box inside the [B] box? [No]
	The [B] box is inside the [A] box.	Is the [B] box inside the [A] box? [Yes]
	The [B] box is inside the [A] box.	Does the [A] box contain the [B] box? [Yes]
	The [B] box is inside the [A] box.	Does the [B] box contain the [A] box? [No]
	The [A] box contains the [B] box.	Is the [A] box inside the [B] box? [No]
	The [A] box contains the [B] box.	Is the [B] box inside the [A] box? [Yes]
	The [A] box contains the [B] box.	Does the [A] box contain the [B] box? [Yes]
	The [A] box contains the [B] box.	Does the [B] box contain the [A] box? [No]
	The [B] box contains the [A] box.	Is the [A] box inside the [B] box? [Yes]
	The [B] box contains the [A] box.	Is the [B] box inside the [A] box? [No]
	The [B] box contains the [A] box.	Does the [A] box contain the [B] box? [No]
	The [B] box contains the [A] box.	Does the [B] box contain the [A] box? [Yes]
Combinatorial entailment (three stimuli)	The [A] box is inside the [B] box; the [B] box is inside the [C] box.	Is the [A] box inside the [C] box? [Yes]
	The [A] box is inside the [B] box; the [B] box is inside the [C] box.	Is the [C] box inside the [A] box? [No]
	The [A] box is inside the [B] box; the [B] box is inside the [C] box.	Does the [A] box contain the [C] box? [No]
	The [A] box is inside the [B] box; the [B] box is inside the [C] box.	Does the [C] box contain the [A] box? [Yes]
	The [C] box is inside the [B] box; the [B] box is inside the [A] box.	Is the [A] box inside the [C] box? [No]
	The [C] box is inside the [B] box; the [B] box is inside the [A] box.	Is the [C] box inside the [A] box? [Yes]
	The [C] box is inside the [B] box; the [B] box is inside the [A] box.	Does the [A] box contain the [C] box? [Yes]
	The [C] box is inside the [B] box; the [B] box is inside the [A] box.	Does the [C] box contain the [A] box? [No]
	The [A] box contains the [B] box; the [B] box contains the [C] box.	Is the [A] box inside the [C] box? [No]
	The [A] box contains the [B] box; the [B] box contains the [C] box.	Is the [C] box inside the [A] box? [Yes]
	The [A] box contains the [B] box; the [B] box contains the [C] box.	Does the [A] box contain the [C] box? [Yes]
	The [A] box contains the [B] box; the [B] box contains the [C] box.	Does the [C] box contain the [A] box? [No]
	The [A] box contains the [B] box; the [B] box contains the [C] box.	Is the [A] box inside the [A] box? [Yes]
	The [C] box contains the [B] box; the [B] box contains the [A] box.	Is the [A] box inside the [C] box? [Yes]
	The [C] box contains the [B] box; the [B] box contains the [A] box.	Is the [C] box inside the [A] box? [No]
	The [C] box contains the [B] box; the [B] box contains the [A] box.	Does the [A] box contain the [C] box? [No]
The [C] box contains the [B] box; the [B] box contains the [A] box.	Does the [C] box contain the [A] box? [Yes]	

The codes [A], [B], and [C] signify a set of different colors used within a particular trial. The actual colors of the box stimuli (e.g., red, green, blue, yellow, pink, purple, and so on) varied quasirandomly across trials

**Table 2** Trial Types Used in the Arbitrary Containment Phase

Property	Statement	Question [answer]
Mutual entailment (two stimuli)	The [A] circle is inside the [B] circle.	Is the [A] circle inside the [B] circle? [Yes]
	The [A] circle is inside the [B] circle.	Is the [B] circle inside the [A] circle? [No]
	The [A] circle is inside the [B] circle.	Does the [A] circle contain the [B] circle? [No]
	The [A] circle is inside the [B] circle.	Does the [B] circle contain the [A] circle? [Yes]
	The [B] circle is inside the [A] circle.	Is the [A] circle inside the [B] circle? [No]
	The [B] circle is inside the [A] circle.	Is the [B] circle inside the [A] circle? [Yes]
	The [B] circle is inside the [A] circle.	Does the [A] circle contain the [B] circle? [Yes]
	The [B] circle is inside the [A] circle.	Does the [B] circle contain the [A] circle? [No]
	The [A] circle contains the [B] circle.	Is the [A] circle inside the [B] circle? [No]
	The [A] circle contains the [B] circle.	Is the [B] circle inside the [A] circle? [Yes]
	The [A] circle contains the [B] circle.	Does the [A] circle contain the [B] circle? [Yes]
	The [A] circle contains the [B] circle.	Does the [B] circle contain the [A] circle? [No]
	The [B] circle contains the [A] circle.	Is the [A] circle inside the [B] circle? [Yes]
	The [B] circle contains the [A] circle.	Is the [B] circle inside the [A] circle? [No]
	The [B] circle contains the [A] circle.	Does the [A] circle contain the [B] circle? [No]
	The [B] circle contains the [A] circle.	Does the [B] circle contain the [A] circle? [Yes]
Combinatorial entailment (three stimuli)	The [A] circle is inside the [B] circle; the [B] circle is inside the [C] circle.	Is the [A] circle inside the [C] circle? [Yes]
	The [A] circle is inside the [B] circle; the [B] circle is inside the [C] circle.	Is the [C] circle inside the [A] circle? [No]
	The [A] circle is inside the [B] circle; the [B] circle is inside the [C] circle.	Does the [A] circle contain the [C] circle? [No]
	The [A] circle is inside the [B] circle; the [B] circle is inside the [C] circle.	Does the [C] circle contain the [A] circle? [Yes]
	The [B] circle is inside the [A] circle; the [B] circle is inside the [C] circle.	Is the [A] circle inside the [C] circle? [No]
	The [B] circle is inside the [A] circle; the [B] circle is inside the [C] circle.	Is the [C] circle inside the [A] circle? [Yes]
	The [B] circle is inside the [A] circle; the [B] circle is inside the [C] circle.	Does the [A] circle contain the [C] circle? [Yes]
	The [B] circle is inside the [A] circle; the [B] circle is inside the [C] circle.	Does the [C] circle contain the [A] circle? [No]
	The [A] circle contains the [B] circle; the [B] circle contains the [C] circle.	Is the [A] circle inside the [C] circle? [No]
	The [A] circle contains the [B] circle; the [B] circle contains the [C] circle.	Is the [C] circle inside the [A] circle? [Yes]
	The [A] circle contains the [B] circle; the [B] circle contains the [C] circle.	Does the [A] circle contain the [C] circle? [Yes]
	The [A] circle contains the [B] circle; the [B] circle contains the [C] circle.	Does the [C] circle contain the [A] circle? [No]
	The [B] circle contains the [A] circle; the [B] circle contains the [C] circle.	Is the [A] circle inside the [C] circle? [Yes]
	The [B] circle contains the [A] circle; the [B] circle contains the [C] circle.	Is the [C] circle inside the [A] circle? [No]
	The [B] circle contains the [A] circle; the [B] circle contains the [C] circle.	Does the [A] circle contain the [C] circle? [No]
	The [B] circle contains the [A] circle; the [B] circle contains the [C] circle.	Does the [C] circle contain the [A] circle? [Yes]

The codes [A], [B], and [C] signify a set of different colors used within a particular trial. The actual colors of the box stimuli (e.g., red, green, blue, yellow, pink, purple, and so on) varied quasirandomly across trials

typically be answered in the form of a *yes* or *no* response (again, more specific details regarding format are provided in later sections), and no contingent feedback was provided at any time.

As described previously, for all three relational repertoires, questions were scored in groups of 16. In the case of each item in each 16-question section, participants received either 1 (*correct*) or 0 (*incorrect*) for that item, and thus they received a score from 0 to 16 for that section. A score of 13/16 or higher was deemed a pass for that section (it was calculated that achieving a score within this range by chance was approximately 1 in 100). Section final scores were added to give cumulative scores for each relational repertoire and an overall relational framing score.

Participants were first assessed for nonarbitrary containment responding across 32 trials. Table 1 shows a generic representation of the trial types involved. In tests for ME, the participant was presented with two differently colored boxes (one inside another), and all questions within this phase of assessment were focused on the relationship between these two stimuli. A total of 16 questions (including eight questions that could be answered *yes* and eight that could be answered *no*) were presented in random order during this phase. In tests for CE, the participant was presented with three differently colored boxes (one inside a second and the second inside a third), and all 16 questions in this phase focused on the relationship between the first and third boxes.

Participants were next assessed for arbitrary containment responding across 32 trials. The trial types for this assessment were similar to those for the assessment of nonarbitrary containment, but this assessment used identically sized colored-circle stimuli between which no physical containment relationship was demonstrated (see Table 2 for a generic representation of the trial types involved).

In tests for ME of arbitrary containment, the participant was presented with two shapes (e.g., circles) of equal size, but different color, and all questions in this assessment phase were focused on the relationship between these two stimuli. A total of 16 questions, presented in random order, were presented during this phase. In tests for CE, the participant was presented with three shapes (designated as A, B, and C) of equal size, but different color, and all 16 questions in this phase focused on the derivation of a relation between Stimuli A and C based on the combination of given relations between A and B and between B and C, respectively.

In the third session of relational testing, participants were assessed for arbitrary hierarchical relational responding across a total of 64 questions (see Table 3 for a list of the trial types involved).

In all tests for arbitrary hierarchy, the participant was presented (on a laptop computer) with on-screen textual descriptions of the relationship(s) among one or more pairs of stimuli (see Fig. 1), and these descriptions were also read aloud. The

questions in this phase were presented in groups of eight items each (see Table 3); the questions involved were presented in random order in each. The groups were presented in the following order:

1. ME questions including the cue *type of*.
2. ME questions including the cue *type of* that assessed ToF.
3. ME questions including the cue *contains*.
4. ME questions including the cue *contains* that assessed ToF.
5. CE questions including the cue *type of*.
6. CE questions including the cue *type of* that assessed ToF.
7. CE questions including the cue *contains*.
8. CE questions including the cue *contains* that assessed ToF.

In each item in the basic ME phases (i.e., Groups 1 and 3), an initial statement involving a hierarchical relational cue was presented (e.g., “A *tol* is a type of animal”), and the participant was then required to respond correctly to a follow-up question based on this provided statement (see Table 3). Correctly answered items in Groups 1 and 3 were combined to give a score for ME. In each item in the ME ToF phases (i.e., Groups 2 and 4), an initial statement involving a hierarchical relational cue was presented, a function (e.g., “big eyes”) was (verbally) associated with one or other of the two stimuli, and the participant was required to respond correctly to a follow-up question based on this. Correctly answered items in Groups 2 and 4 were combined to give a score for ME with ToF. In each item in the basic CE phases (i.e., Groups 5 and 7), two statements involving a hierarchical relational cue were presented, and the participant was then required to respond correctly to a follow-up question based on the combination of those statements. Correctly answered items in Groups 5 and 7 were combined to give a score for CE. In each item in the CE ToF phases (i.e., Groups 6 and 8), two statements involving a hierarchical relational cue were presented, a function was associated with one of the stimuli involved, and the participant was required to respond correctly to a follow-up question based on this. Correctly answered items in Groups 6 and 8 were combined to give a score for CE with ToF.

As noted previously in the first paragraph of the procedure, two versions of the relational responding test were given: Relational Responding Tests 1 and 2. Test 2 was similar to Test 1, except that a different set of stimuli was involved. More specifically, in the nonarbitrary and arbitrary containment tests, differently shaped boxes and different shapes, respectively, were used, whereas in the arbitrary hierarchy test, a different set of textual stimuli (i.e., nonsense and real words) was presented (see Table 8 for the full set of arbitrary hierarchy retest questions).



**Table 3** Trial Types Used in the Arbitrary Hierarchy Phase

Property	Statement	Question [answer]
ME (two stimuli)	A tol is a type of animal.	Is a tol a type of animal? [Yes] Is an animal a type of tol? [No]
	A blorg is a type of animal.	Are all blorgs animals? [Yes] Are all animals blorgs? [No]
	A zimp is a type of animal.	Are some animals zimps? [Yes] Does the class of zimps contain all of the animals? [No]
	A quig is a type of animal.	Does the class of animals contain quigs? [Yes] Are all animals quigs? [No]
ToF (two stimuli)	A tol is a type of gip. Gips have bones.	Do all tols have bones? [Yes] Does the class of tols contain members without bones? [No]
	A blorg is a type of grap. Graps have freckles.	Does the class of blorgs contain members that have freckles? [Yes] Does the class of graps contain members without freckles? [No]
	A zimp is a type of crint. Zimps have spots.	Do all crints have spots? [Cannot know] Do some crints have spots? [Yes]
	A quig is a type of donk. Quigs have purple tongues.	Does the class of donks contain members with purple tongues? [Yes] Do all donks have purple tongues? [Cannot know]
ME (two stimuli)	The class of animals contains tols.	Are all tols animals? [Yes] Are all animals tols? [No]
	The class of animals contains blorgs.	Does the class of animals contain blorgs? [Yes] Does the class of blorgs contain all of the animals? [No]
	The class of animals contains zimps.	Does the class of animals contain zimps? [Yes] Are all animals zimps? [No]
	The class of animals contains quigs.	Are some animals quigs? [Yes] Does the class of quigs contain all of the animals? [No]
ToF (two stimuli)	The class of crints contains zimps. Crints are furry.	Does the class of zimps contain members that are furry? [Yes] Does the class of crints contain members that aren't furry? [No]
	The class of donks contains quigs. Donks have square eyes.	Do all quigs have square eyes? [Yes] Does the class of quigs contain members without square eyes? [No]
	The class of gips contains tols. Tols have big green eyes.	Does the class of gips contain members with big green eyes? [Yes] Do some gips have big green eyes? [Yes]
	The class of graps contains blorgs. Blorgs have whiskers.	Do all graps have whiskers? [Cannot know] Do some graps have whiskers? [Yes]
CE (three stimuli)	A bik is a type of gip; a gip is a type of animal.	Is a bik a type of animal? [Yes] Is an animal a type of bik? [No]
	A grap is a type of plak; a plak is a type of animal.	Are all graps animals? [Yes] Are all animals graps? [No]
	A crint is a type of wark; a wark is a type of animal.	Are some animals crints? [Yes] Does the class crint contain all of the animals? [No]
	A donk is a type of frik; a frik is a type of animal.	Does the class of animals contain donks? [Yes] Are all animals donks? [No]

**Table 3** (continued)

Property	Statement	Question [answer]
ToF (three stimuli)	A bik is a type of gip; a gip is a type of timp. Timps like sweet food.	Do all biks like sweet food? [Yes] Does the class of timps contain members who don't like sweet food? [No]
	A rab is a type of grap; a grap is a type of plak. Plaks have scales.	Does the class of rabs contain members that have scales? [Yes] Does the class of rabs contain members without scales? [No]
	A dring is a type of crint; a crint is a type of wark. Drings have tails.	Do all warks have tails? [Cannot know] Do some warks have tails? [Yes]
	A slib is a type of donk; a donk is a type of frik. Slibs have spikes.	Does the class of frik contain members with spikes? [Yes] Do all friks have spikes? [Cannot know]
CE (three stimuli)	The class of animals contains gips; the class of gips contains biks.	Are all biks animals? [Yes] Are all animals biks? [No]
	The class of animals contains plaks; the class of plaks contains graps.	Is a grap a type of animal? [Yes] Is an animal a type of grap? [No]
	The class of animals contains warks; the class of warks contains crints.	Does the class of animals contain crints? [Yes] Are all animals crints? [No]
	The class of animals contains friks; the class of friks contains donks.	Are some animals donks? [Yes] Does the class donk contain all of the animals? [No]
ToF (three stimuli)	The class of warks contains crints; the class of crints contains drings. Warks have gold eyes.	Does the class of drings contain members with gold eyes? [Yes] Does the class of warks contain members without gold eyes? [No]
	The class of friks contains donks; the class of donks contains slibs. Friks are cold.	Are all slibs cold? [Yes] Does the class of slibs contain members that aren't cold? [No]
	The class of timps contains gips; the class of gips contains biks. Biks have small noses.	Does the class of timps contain members with small noses? [Yes] Do some timps have small noses? [Yes]
	The class of plaks contains graps; the class of graps contains rabs. Rabs have big ears.	Do all plaks have big ears? [Cannot know] <i>Do some plaks have big ears? [Yes]</i>

*ME* mutual entailment, *ToF* transformation of functions, *CE* combinatorial entailment

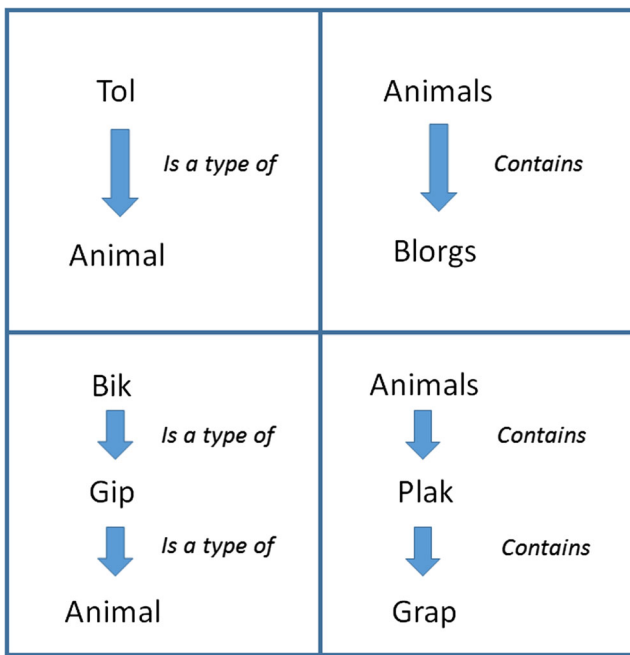
**CI** This test involved presenting eight questions focused on CI as previously described. These included the following question types:

1. Are there more [smaller subclass] or more [category]?
2. Are there more [larger subclass] or more [category]?
3. Are there more [category] or more [smaller subclass]?
4. Are there more [category] or more [larger subclass]?
5. Are there less [smaller subclass] or less [category]?
6. Are there less [larger subclass] or less [category]?
7. Are there less [category] or less [smaller subclass]?
8. Are there less [category] or less [larger subclass]?

These questions were presented in random order based on the shuffling of eight question cards. Participants received noncontingent reinforcement for taking part.

### **Interobserver Agreement (IOA) and Procedural Fidelity**

IOA was conducted by research assistants in the case of 20% of participants and was collected during nonarbitrary containment, arbitrary containment, and arbitrary hierarchy assessment sessions. Prior to data collection, IOA collectors were trained in data collection until they reached 100% accuracy. Mean IOA was calculated as 99.48% (range of 96.88% to 100%). Procedural fidelity was conducted for 20% of participants and was collected during nonarbitrary containment, arbitrary containment, and arbitrary hierarchy assessment sessions. A trained research assistant collected procedural fidelity measures. Prior to data collection, correct experimenter procedural responding was modeled to the data collectors for exemplars of procedural integrity, and data collectors were also



**Fig. 1** Sample stimuli layouts for mutual entailment (top panels) and combinatorial entailment (bottom panels) trials in the arbitrary hierarchy testing phases

provided with a printed copy of the assessment protocol and a checklist of the steps in the assessment procedure. Mean procedural integrity was calculated as 99.56% (range of 93.75% to 100%).

## Results

### Correlations

Table 4 shows a correlation matrix of Spearman’s  $\rho$  correlations among the experimental measures and some of their key subscales.

This table shows a highly significant correlation between age in months and overall relational framing score ( $\rho = .852, p < .001$ ), as well as between age and each of the three specific relational repertoires, including nonarbitrary containment ( $\rho = .826, p < .001$ ), arbitrary containment ( $\rho = .806, p < .001$ ), and arbitrary hierarchy ( $\rho = .789, p < .001$ ). The data also show highly significant correlations between the overall and specific relational framing scores and those of the other measures, as well as, in the case of the SB5, its subscales (i.e., SB5, SB5 Verbal, SB5 Nonverbal, PPVT-4, CCT, and CI). In general, the highest correlations are seen for the measures of general intellectual performance (i.e., SB5 and subscales) and language (PPVT-4), with slightly lower correlations for the measures of categorization (i.e., CCT and CI).

Regarding the correlations among the nonrelational framing measures themselves, we can see that the

**Table 4** Matrix of Spearman’s Rho Correlations for All Measures Administered

Test	Total RF	NAC	AC	AH	AM	SB5-T	SB5-V	SB5-NV	PPVT-4	CCT	CI
Total RF	—										
NAC	.915 <sup>***</sup>	—									
AC	.930 <sup>***</sup>	.799 <sup>***</sup>	—								
AH	.960 <sup>***</sup>	.804 <sup>***</sup>	.870 <sup>***</sup>	—							
AM	.852 <sup>***</sup>	.826 <sup>***</sup>	.806 <sup>***</sup>	.789 <sup>***</sup>	—						
SB5-T	.832 <sup>***</sup>	.784 <sup>***</sup>	.758 <sup>***</sup>	.770 <sup>***</sup>	.844 <sup>***</sup>	—					
SB5-V	.778 <sup>***</sup>	.747 <sup>***</sup>	.705 <sup>***</sup>	.721 <sup>***</sup>	.822 <sup>***</sup>	.966 <sup>***</sup>	—				
SB5-NV	.845 <sup>***</sup>	.790 <sup>***</sup>	.801 <sup>***</sup>	.771 <sup>***</sup>	.850 <sup>***</sup>	.927 <sup>***</sup>	.822 <sup>***</sup>	—			
PPVT-4	.785 <sup>***</sup>	.745 <sup>***</sup>	.679 <sup>***</sup>	.731 <sup>***</sup>	.828 <sup>***</sup>	.918 <sup>***</sup>	.891 <sup>***</sup>	.850 <sup>***</sup>	—		
CCT	.615 <sup>***</sup>	.551 <sup>***</sup>	.605 <sup>***</sup>	.556 <sup>***</sup>	.679 <sup>***</sup>	.817 <sup>***</sup>	.802 <sup>***</sup>	.756 <sup>***</sup>	.769 <sup>***</sup>	—	
CI	.634 <sup>***</sup>	.544 <sup>***</sup>	.674 <sup>***</sup>	.621 <sup>***</sup>	.605 <sup>***</sup>	.432 <sup>***</sup>	.401 <sup>***</sup>	.415 <sup>***</sup>	.285 <sup>***</sup>	.285 <sup>***</sup>	—

All scores are raw (unstandardized) scores. Total RF = total relational framing score; NAC = total nonarbitrary containment score; AC = total arbitrary containment score; AH = total arbitrary hierarchy score; AM = age in months; SB5-T = Stanford-Binet Intelligence Scales—Fifth Edition total score; SB5-V = Stanford-Binet Intelligence Scales—Fifth Edition total verbal score; SB5-NV = Stanford-Binet Intelligence Scales—Fifth Edition total nonverbal score; PPVT-4 = Peabody Picture Vocabulary Test, Fourth Edition, total score; CCT = Children’s Category Test total score; CI = total class inclusion score

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

standardized measures of intellectual performance and language correlate very highly with each other, as might be expected (e.g., SB5 and PPVT-4 show a correlation of  $\rho = .918$ ). Each of these measures also correlates well with the CCT (e.g., PPVT-4:  $\rho = .769$ ; SB5:  $\rho = .817$ ) and not quite as well with the CI (e.g., PPVT-4:  $\rho = .415$ ; SB5:  $\rho = .432$ ). In regard to the two latter (categorization) measures: (a) The CCT shows much higher levels of correlation with the measures of intellectual and language performance than the CI, (b) both tests show comparable levels of correlation with the relational framing measures, and (c) the level of correlation between these tests themselves is the lowest in the table (i.e.,  $\rho = .285$ ).

### Relational Framing Performance per Age Cohort

Table 5 shows the average number of correct responses and corresponding percentage of correct responses per age group and per relational framing pattern, including both relational frames (nonarbitrary containment, arbitrary containment, and arbitrary hierarchy) and relational properties (ME, CE, and ToF). The broad patterns seen are in accordance with what might have been predicted in that:

1. For each relational property (and thus also all relations), all age cohorts perform in the aggregate at least as well as any younger cohort (with one exception, which is for ME of nonarbitrary containment relations as shown by the 5- to 6-year-old and 6- to 7-year-old groups, and in that case

the difference is relatively small). In general, aggregate performance on all indices improves with age.

2. For all cohorts and for all three relations, the aggregate score for the CE test is at most as high and is typically lower than the score for the corresponding ME test.

Table 6 shows the aggregate number of passes achieved within each age cohort for each of the relational properties assessed for each of the three patterns of relational framing (also see Fig. 2). The broad patterns seen are similar to those seen for the raw data in that:

1. For each relational property (and thus also all relations), all cohorts perform at least as well as any younger cohort and, in general, performance on all indices improves with age.
2. For all cohorts and all three relations, the score for CE is at most as high and is typically lower than that for ME (with one slight exception: ToF for 6- to 7-year-olds)

### Relational Framing Test–Retest Scores

Table 7 shows correlations between scores for the initial test and the retest in the case of all subsections of each relation, for each relational total, and for the overall relational framing test total. All the tests showed very highly significant test–retest correlations.

**Table 5** Average Correct Response Number and Percentage per Age Group and Relational Framing Pattern (Including Frames and Properties)

Age cohort	Nonarbitrary containment		Arbitrary containment			Arbitrary hierarchy					
	ME (%)	CE (%)	Total (%)	ME (%)	CE (%)	Total (%)	ME (%)	ToF–M (%)	CE (%)	ToF–C (%)	Total (%)
3–4 years	8 (50%)	7.6 (48%)	15.6 (49%)	6.7 (42%)	5.8 (36%)	12.5 (39%)	6 (37%)	5.9 (37%)	5.6 (35%)	3.8 (24%)	21.3 (33%)
4–5 years	11.2 (70%)	9.2 (58%)	20.4 (64%)	9.8 (61%)	8.3 (52%)	18.1 (57%)	8.3 (52%)	7.1 (44%)	6.7 (42%)	5.2 (33%)	27.2 (43%)
5–6 years	14.2 (89%)	11.6 (73%)	25.8 (81%)	10.4 (65%)	8.4 (53%)	18.8 (59%)	9.4 (59%)	7.6 (47%)	8.4 (53%)	7.5 (47%)	32.8 (51%)
6–7 years	13.6 (85%)	13.3 (83%)	26.9 (84%)	12.8 (80%)	9.8 (61%)	22.6 (71%)	9.8 (61%)	8.7 (54%)	8.5 (53%)	8.1 (51%)	35.1 (55%)
7–8 years	16 (100%)	15.6 (98%)	31.6 (99%)	14.2 (89%)	13 (81%)	27.2 (85%)	13.5 (84%)	11 (69%)	12.3 (77%)	10.9 (68%)	47.7 (75%)

ME = mutual entailment (number of correct responses out of 16); CE = combinatorial entailment (number of correct responses out of 16); total = total number of correct entailment responses for a particular frame (number of correct responses out of 32 or 64); ToF–M = transformation of function of mutual entailment (number of correct responses out of 16); ToF–C = transformation of function of combinatorial entailment (number of correct responses out of 16)

**Table 6** Number of Passes for Each Relational Test per Age Cohort

Age cohort	Nonarbitrary containment		Arbitrary containment		Arbitrary hierarchy			
	ME	CE	ME	CE	ME	ToF-M	CE	ToF-C
3–4 years	0	0	0	0	0	0	0	0
4–5 years	2	0	0	0	0	0	0	0
5–6 years	7	4	2	0	2	0	0	0
6–7 years	7	7	7	0	2	0	1	1
7–8 years	10	10	10	5	6	2	5	1

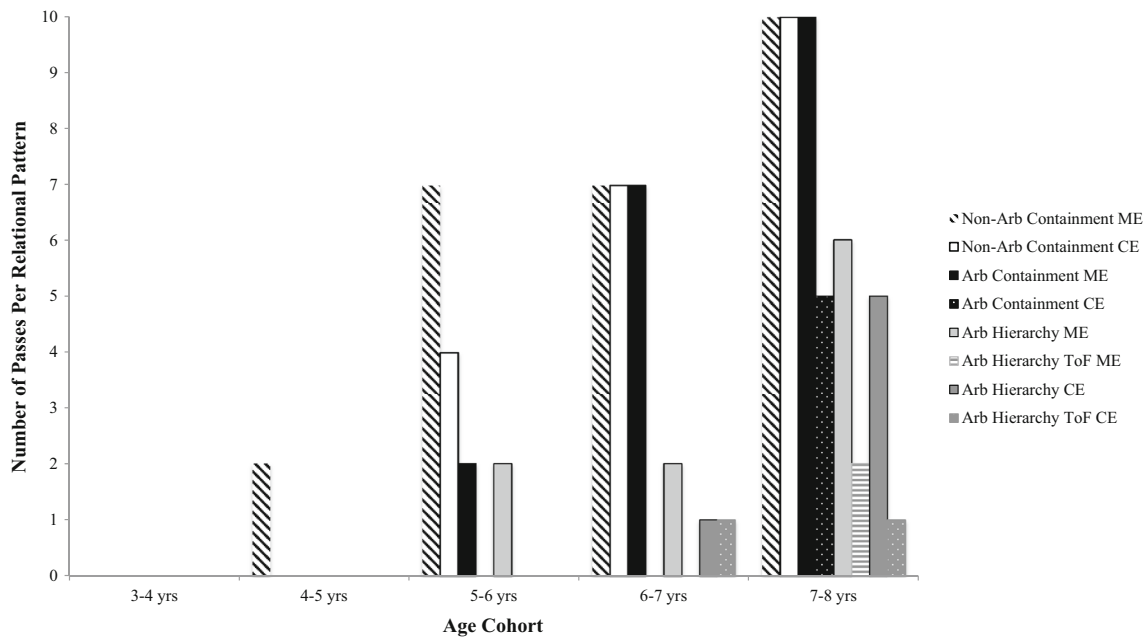
ME mutual entailment, CE combinatorial entailment, ToF-M transformation of function for mutual entailment, ToF-C transformation of function for combinatorial entailment

**Discussion**

The current study aimed to investigate relational framing related to hierarchical classification in children. To this end, it involved administering to 50 children ranging from 3 to 8 years of age a custom-made protocol designed to assess relational framing presumed to be relevant to hierarchical classification from relatively simple and concrete to relatively more complex and abstract—specifically, nonarbitrary containment, arbitrary containment, and arbitrary hierarchy, each assessed through testing of ME and CE, as well as, in the case of arbitrary hierarchy, through testing of ToF. This was done with a view to gathering information regarding typical acquisition of relational framing repertoires relevant to classification; furthermore, the protocol was administered more than once to check for reliability. All the children were also tested

using a number of other measures, including (a) assessment of generalized intellectual (cognitive and linguistic) skill (i.e., the SB5 and the PPVT-4) and (b) assessment of classification skill specifically (i.e., the CCT and the Piagetian CI test). The tests of generalized intellectual performance were administered chiefly to examine the correlation of relational framing of categorization with intellectual performance more generally, whereas the tests of categorization were administered to probe the validity of the claim that the relational framing repertoires examined herein are indeed relevant to classification responding.

The data showed first that there was a developmental trend in terms of the acquisition of the overall relational framing repertoire involved as well as with respect to each of the more specific relational repertoires. This is evident in the correlational data (Table 4), as shown by the high correlation between age and each of the repertoires, and in the descriptive data (Table 5) and the data showing the numbers of passes (Table 6) for each cohort for various patterns of relational framing (frames and properties). These patterns suggest that, as might have been predicted, relational framing of categorization is acquired gradually over the course of several years. The youngest age group (3–4 years) shows little or no capacity in any of the three specific relational repertoires assessed, including even nonarbitrary containment, whereas data for the oldest age group (7–8 years) indicate at least some emergence on each of the three repertoires. All members of this oldest group passed both sections of nonarbitrary containment. However, their performance on the two arbitrary sections and particularly the arbitrary hierarchy section was relatively poor, even though they performed better than the



**Fig. 2** Number of passes per relational frame/property and age group. ME = mutual entailment; CE = combinatorial entailment; ToF = transformation of functions

**Table 7** Test–Retest Correlations for Relational Framing Test and Subtests

Relational test or subtest	Spearman's rho
Nonarbitrary containment	
Mutual entailment	.981*
Combinatorial entailment	.981*
Total score	.991*
Arbitrary containment	
Mutual entailment	.974*
Combinatorial entailment	.966*
Total score	.989*
Arbitrary hierarchy	
Mutual entailment	.982*
ToF of mutually entailed relations	.803*
Combinatorial entailment	.946*
ToF of combinatorially entailed relations	.821*
Total score	.993*
Total relational framing	.993*

All scores are raw (unstandardized) scores. ToF = transformation of function

\*  $p < .001$

younger groups. These data suggest that children of all age groups might potentially benefit from training in relational frames of categorization and, given the correlations seen between these patterns of framing and performance on standardized measures of language and cognition, that such training might be intellectually beneficial. The data for number of passes also show that the order of sequencing of relational framing repertoires in the current protocol (i.e., nonarbitrary containment, arbitrary containment, and arbitrary hierarchy) was correct and that this is indeed the order of acquisition and would thus also be an appropriate order for assessment and training. Finally, scores on the second administration of the protocol also correlated very highly with scores for the first administration (Table 7), thus suggesting the reliability of the protocol.

Regarding the data showing correlations between relational framing and performance on other measures, a number of points in particular are important. First, as previously mentioned, there were high levels of correlation between relational framing and performance on standardized intellectual measures, including both cognitive and linguistic. These correlations were seen for overall relational framing capacity as well as for specific relational repertoires. These findings further extend a pattern seen in previous RFT studies showing strong correlations between relational framing and intellectual performance (e.g., Moran, Walsh, Stewart, McElwee, & Ming, 2015; O'Toole et al., 2009; Ruiz & Luciano, 2011). Second, although there were strong correlations seen between scores for the relational framing repertoires and those for the

standardized tests of categorization specifically, these correlations were not as high as those found between the former and those found for the broad measures of intellectual potential.

An original version of the current protocol assessed ToF not only for arbitrary hierarchy but also for all three framing patterns; indeed, data on ToF for all three frames were collected for the current cohort. However, the items assessing ToF in the case of the two containment relational patterns were problematic, and thus the data were omitted. Hence, the data reported herein for these two frames focus just on derived ME and CE alone and do not assess ToF. Despite this, the correlations between the overall relational framing protocol and the alternative measures used were strong. This is perhaps not surprising, as (a) ToF data were collected for arbitrary hierarchy, which might be argued to be particularly important in terms of the repertoire of classification, and (b) RFT protocols that focus on derived relations alone have previously proven to be excellent predictors of intellectual performance (see, e.g., Cassidy, Roche, & Hayes, 2011). Therefore, the data provided for the containment frames would still have been useful in this respect. At the same time, from an RFT point of view, ToF is a key property of relational framing, and thus collecting it for all three frames would likely boost predictive power. Engaging in the latter as an extension of the current study would be a useful future research direction.

Another possible critique of the categorization framing protocol used in the current study might be that it did not provide as thorough or comprehensive a test of arbitrary hierarchical framing as might be desirable. For example, although framing appropriately in response to the cues of *type of* and *contains* is an important part of this pattern, there are other aspects of this pattern that might also be tested. For instance, when two stimuli are framed as being part of the same class, then they should also both be framed as being different from other stimuli framed within a different class and equivalent to each other, in at least some contexts, independent of their physical properties. Such a pattern of responding might be expected from someone with a sufficiently advanced repertoire of hierarchical classification. The current protocol does not involve testing for such relations. Adjusting it so as to do so might be expected to improve its reliability and validity as a test of classification. This would be one useful direction for future work.

Another broader issue for consideration is that when using a relational skills test as a proxy for other skills (i.e., such as categorization) by way of identifying which skills deficits may need ameliorating in a relational skill training intervention, it is assumed that the relational skills levels are distributed more or less normally within each age cohort. In other words, if one does not know how skills are distributed across larger samples, one cannot know the significance of a relational test score that is deviating from the mean. Analysis of this kind would not have been possible with the current data set, as it is too small, but this may be an important consideration for

researchers interested in using relational tests as either proxies for other intellectual skills or as screening tests to assess deficits prior to intervention.

Another relevant direction for future studies might be the exploration of the capacity for establishing and/or strengthening repertoires of classification framing. If the current protocol (or a refined version of it) is seen as providing a reasonable measure of relational framing relevant to classification, as seems possible based on these data, then in future studies, it might be used to identify deficits in this repertoire as a precursor to the training of the relevant relational skills. An intervention for one or more aspects of classification framing might be expected to boost not only this repertoire itself but also—analogue to the outcomes seen in the case of other relational frame training interventions—intellectual performance more generally. Indeed, given the centrality of classification to the average person's intellectual repertoire, perhaps such training might have a more substantial effect in this regard than training involving other frames.

One further possible direction for future research might be the examination of the relationship and/or interaction between classification framing and other framing repertoires. Such work might investigate various phenomena, including correlation between classification and various other patterns of framing, as well as the effect of training particular frames on performance in accordance with classification framing. For example, as suggested previously, sameness and difference

frames may be particularly important precursors and/or accompaniments of hierarchical framing, suggesting that training fluency and/or flexibility on these frames might be particularly beneficial with respect to the acquisition and/or strengthening of hierarchical framing and categorization framing more broadly.

In summary, this is a promising initial result for the assessment of classification framing. The pattern of data acquired from the protocol used to measure this repertoire in combination with that from a suite of other measures of intellectual functioning, including ones focused on classification per se, might suggest that the former is indeed tapping into important aspects of categorization. This extends research on relational framing in general and provides an impetus for further work into classification per se from an RFT point of view. Further research might usefully extend the current study by incorporating additional testing of features of categorization as a pattern of responding as well as by facilitating the training of this repertoire in typically developing children or children with developmental delays who show relevant deficits.

**Compliance with Ethical Standards** This research was supported by a scholarship awarded to Teresa Mulhern. On behalf of all authors, the corresponding author states that there are no conflicts of interest. This research was conducted at the National University of Ireland, Galway, and was supported by the National University of Ireland, Galway, Doctoral Research Scholarship Scheme, which was awarded to Teresa Mulhern.

## Appendix

**Table 8** Trial Types Used in the Arbitrary Hierarchy Phase of the Second Administration of the Relational Test

Property	Statement	Question [answer]
ME (two stimuli)	A bento is a type of animal.	Is a bento a type of animal? [Yes]
	A stak is a type of animal.	Is an animal a type of bento? [No] Are all staks animals? [Yes]
	A zum is a type of animal.	Are all animals staks? [No] Are some animals zums? [Yes]
	A vink is a type of animal.	Does the class of zums contain all of the animals? [No] Does the class of animals contain vinks? [Yes]
ToF (two stimuli)	A bento is a type of jin. Jins have lungs.	Are all animals vinks? [No] Do all bentos have lungs? [Yes]
	A stak is a type of wilk. Wilks have silver ears.	Does the class of bentos contain members without lungs? [No] Does the class of staks contain members that have silver ears? [Yes]
	A zum is a type of taga. Zums have feathers.	Does the class of wilks contain members without silver ears? [No] Do all tagas have feathers? [Cannot know]

**Table 8** (continued)

Property	Statement	Question [answer]
ME (two stimuli)	A vink is a type of ropa. Vinks have claws.	Do some tagas have feathers? [Yes] Does the class of ropas contain members with claws? [Yes] Do all ropas have claws? [Cannot know]
	The class of animals contains bentos.	Are all bentos animals? [Yes] Are all animals bentos? [No]
	The class of animals contains staks.	Does the class of animals contain staks? [Yes] Does the class of staks contain all of the animals? [No]
	The class of animals contains zums.	Does the class of animals contain zums? [Yes] Are all animals zums? [No]
ToF (two stimuli)	The class of animals contains vinks.	Are some animals vinks? [Yes] Does the class of vinks contain all of the animals? [No]
	The class of tagas contains zums. Tagas have beady eyes.	Does the class of zums contain members that have beady eyes? [Yes] Does the class of tagas contain members without beady eyes? [No]
	The class of ropas contains vinks. Ropas have three eyes.	Do all vinks have three eyes? [Yes] Does the class of vinks contain members without three eyes? [No]
	The class of jins contains bentos. Bentos like salty food.	Does the class of jins contain members that like salty food? [Yes] Do some jins like salty food? [Yes]
CE (three stimuli)	The class of wilks contains staks. Staks eat apples.	Do all wilks eat apples? [Cannot know] Do some wilks eat apples? [Yes]
	A tang is a type of jin; a jin is a type of animal.	Is a tang a type of animal? [Yes] Is an animal a type of tang? [No]
	A yalt is a type of wilk; a wilk is a type of animal.	Are all yalts animals? [Yes] Are all animals yalts? [No]
	An unda is a type of taga; a taga is a type of animal.	Are some animals undas? [Yes] Does the class unda contain all of the animals? [No]
ToF (three stimuli)	A sibe is a type of ropa; a ropa is a type of animal.	Does the class of animals contain sibes? [Yes] Are all animals sibes? [No]
	A tang is a type of jin; a jin is a type of farro. Farros have tails.	Do all tangs have tails? [Yes] Does the class of farros contain members without tails? [No]
	A wob is a type of yalt; a yalt is a type of wilk. Wilks are fluffy.	Does the class of wobs contain members that are fluffy? [Yes] Does the class of wobs contain members that aren't fluffy? [No]
	A drub is a type of unda; an unda is a type of taga. Drubs are smelly.	Are all tagas smelly? [Cannot know] Are some tagas smelly? [Yes]
CE (three stimuli)	A clun is a type of sibe; a sibe is a type of ropa. Cluns have blue eyes.	Does the class of ropas contain members with blue eyes? [Yes] Do all ropas have blue eyes? [Cannot know]
	The class of animals contains jins; the class of jins contains tangs.	Are all tangs animals? [Yes] Are all animals tangs? [No]
	The class of animals contains wilks; the class of wilks contains yalts.	Is a yalt a type of animal? [Yes] Is an animal a type of yalt? [No]
	The class of animals contains tagas; the class of tagas contains undas.	Does the class of animals contain undas? [Yes] Are all animals undas? [No]
ToF (three stimuli)	The class of animals contains ropas; the class of ropas contains sibes.	Are some animals sibes? [Yes] Does the class sibe contain all of the animals? [No]
	The class of tagas contains undas; the class of undas contains drubs. Tagas are hairy.	Does the class of drubs contain members that are hairy? [Yes] Does the class of taga contain members that aren't hairy? [No]
ToF (three stimuli)	The class of ropas contains sibes; the class of sibes contains cluns. Ropas have tentacles.	Do all cluns have tentacles? [Yes] Does the class of cluns contain members without tentacles? [No] Does the class of farros contain members that eat oranges? [Yes]



**Table 8** (continued)

Property	Statement	Question [answer]
	The class of farros contains jins; the class of jins contains tangs. Tangs eat oranges.	Do some farros eat oranges? [Yes]
	The class of wilks contains yalts; the class of yalts contains wobs. Wobs are purple.	Are all wilks purple? [Cannot know] Are some wilks purple? [Yes]

Note. ME = mutual entailment; ToF = transformation of functions; CE = combinatorial entailment

## References

- Arterberry, M. E., & Bornstein, M. H. (2012). Categorization of real and replica objects by 14- and 18-month-old infants. *Infant Behavior and Development, 35*, 606–612. <https://doi.org/10.1016/j.infbeh.2012.04.007>.
- Arterberry, M. E., & Kellman, P. J. (2016). *Development of perception in infancy: The cradle of knowledge revisited*. New York: Oxford University Press.
- Astley, S. L., Peissig, J. J., & Wasserman, E. A. (2001). Superordinate categorization via learned stimulus equivalence: Quantity of reinforcement, hedonic value, and the nature of the mediator. *Journal of Experimental Psychology: Animal Behavior Processes, 27*, 252–268. <https://doi.org/10.1037/0097-7403.27.3.252>.
- Blewitt, P. (1989). Categorical hierarchies: Levels of knowledge and skill. *Genetic Epistemologist, 17*, 21–29.
- Blewitt, P. (1994). Understanding categorical hierarchies: The earliest levels of skill. *Child Development, 65*, 1279–1298. <https://doi.org/10.1111/j.1467-8624.1994.tb00817.x>.
- Blewitt, P., Michnick Golinkoff, R., & Alioto, A. (2000). Do toddlers have label preferences? A possible explanation for word refusals. *First Language, 20*, 253–272. <https://doi.org/10.1177/014272370002006002>.
- Boll, T. (1993). *Children's Category Test*. San Antonio: Psychological Corporation.
- Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology, 46*, 350–365. <https://doi.org/10.1037/a0018411>.
- Carneiro, P., Albuquerque, P., & Fernandez, A. (2009). Opposite developmental trends for false recognition of basic and superordinate names. *Memory, 17*, 411–427. <https://doi.org/10.1080/09658210902758847>.
- Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences, 47*, 222–235. <https://doi.org/10.1016/j.lindif.2016.03.001>.
- Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *The Psychological Record, 61*, 173–198.
- Community-University Partnership for the Study of Children, Youth, and Families. (2011). *Review of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)*. Edmonton: Author.
- Deneault, J., & Ricard, M. (2005). The effect of hierarchical levels of categories on children's deductive inferences about inclusion. *International Journal of Psychology, 40*, 65–79. <https://doi.org/10.1080/00207590444000032>.
- Deneault, J., & Ricard, M. (2006). The assessment of children's understanding of inclusion relations: Transitivity, asymmetry, and quantification. *Journal of Cognition and Development, 7*, 551–570. [https://doi.org/10.1207/s15327647jcd0704\\_6](https://doi.org/10.1207/s15327647jcd0704_6).
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)*. Minneapolis: Pearson.
- Dymond, S., & Roche, B. (2013). *Advances in relational frame theory: Research and application*. Oakland: Context Press.
- Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development, 65*, 903–917. <https://doi.org/10.2307/1131427>.
- Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology, 58*, 418–431. <https://doi.org/10.1006/jecp.1994.1043>.
- Gil, E., Luciano, C., Ruiz, F. J., & Valdivia-Salas, S. (2012). A preliminary demonstration of transformation of functions through hierarchical relations. *International Journal of Psychology and Psychological Therapy, 12*, 1–19.
- Gore, N. J., Barnes-Holmes, Y., & Murphy, G. (2010). The relationship between intellectual functioning and relational perspective-taking. *International Journal of Psychology and Psychological Therapy, 10*, 1–17.
- Greene, T. R. (1994). What kindergartners know about class containment hierarchies. *Journal of Experimental Child Psychology, 57*, 72–88. <https://doi.org/10.1006/jecp.1994.1004>.
- Hajibayova, L. (2013). Basic-level categories: A review. *Journal of Information Science, 39*, 676–687. <https://doi.org/10.1177/0165551513481443>.
- Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *British Journal of Educational Psychology, 86*, 397–411. <https://doi.org/10.1111/bjep.12114>.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001a). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Plenum Press.
- Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001b). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 21–50). New York: Plenum Press.
- Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior, 74*, 207–227. <https://doi.org/10.1901/jeab.2000.74-207>.
- Luciano, C., Gómez Becarra, I., & Rodríguez Valverde, M. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior, 87*, 349–365. <https://doi.org/10.1901/jeab.2007.08-06>.
- MacNeill Horton, A. (1996). Book and test reviews. *Archives of Clinical Neuropsychology, 11*, 171–173.
- Manders, K., & Hall, D. G. (2002). Comparison, basic-level categories, and the teaching of adjectives. *Journal of Child Language, 29*, 923–937. <https://doi.org/10.1017/S0305000902005391>.

- Margolis, E., & Laurence, S. (2000). *Concepts: Core readings*. Cambridge: MIT Press.
- Markman, E. M., & Seibert, J. (1976). Classes and collections: Internal organization and resulting holistic properties. *Cognitive Psychology*, 8, 561–577. [https://doi.org/10.1016/0010-0285\(76\)90019-0](https://doi.org/10.1016/0010-0285(76)90019-0).
- Mompeán, J. A. (2006). The phoneme as a basic-level category: Experimental evidence from English. *International Journal of English Studies*, 6, 141–172.
- Moran, L., Walsh, L., Stewart, I., McElwee, J., & Ming, S. (2015). Correlating derived relational responding with linguistic and cognitive ability in children with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, 19, 32–43. <https://doi.org/10.1016/j.rasd.2014.12.015>.
- Murphy, G. (2002). *The big book of concepts*. Cambridge: MIT Press.
- O’Hora, D., Peláez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, 58, 569–584.
- O’Toole, C., Barnes-Holmes, D., Murphy, C., O’Connor, J., & Barnes-Holmes, Y. (2009). Relational flexibility and human intelligence: Extending the remit of Skinner’s *Verbal Behavior*. *International Journal of Psychology and Psychological Therapy*, 9, 1–17.
- Palmer, D. C. (2002). Psychological essentialism: A review of E. Margolis and S. Laurence (Eds.), *Concepts: Core Readings*. *Journal of the Experimental Analysis of Behavior*, 78, 597–607. <https://doi.org/10.1901/jeab.2002.78-597>.
- Piaget, J. (1952). *The child’s conception of number*. New York: Norton.
- Quinn, P. C. (2016). Establishing cognitive organization in infancy: From perceptual grouping of objects to social classification of faces. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues*, 3rd ed (pp. 79–104). New York: Routledge.
- Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, 63, 189–211. <https://doi.org/10.1006/jecp.1996.0047>.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition*. Itasca: Riverside Publishing.
- Rostad, K., Yott, J., & Poulin-Dubois, D. (2012). Development of categorization in infancy: Advancing forward to the animate/inanimate level. *Infant Behavior and Development*, 35, 584–595. <https://doi.org/10.1016/j.infbeh.2012.05.005>.
- Ruiz, F. J., & Luciano, C. (2011). Cross-domain analogies as relating derived relations among two separate relational networks. *Journal of the Experimental Analysis of Behavior*, 95, 369–385. <https://doi.org/10.1901/jeab.2011.95-369>.
- Slattery, B., & Stewart, I. (2014). Hierarchical classification as relational framing. *Journal of the Experimental Analysis of Behavior*, 101, 61–75. <https://doi.org/10.1002/jeab.63>.
- Slattery, B., Stewart, I., & O’Hora, D. (2011). Testing for transitive class containment as a feature of hierarchical classification. *Journal of the Experimental Analysis of Behavior*, 96, 243–260. <https://doi.org/10.1901/jeab.2011.96-243>.
- Song, L., Pruden, S. M., Michnick Golinkoff, R., & Hirsh-Pasek, K. (2016). Prelinguistic foundations of verb learning: Infants discriminate and categorize dynamic human actions. *Journal of Experimental Child Psychology*, 151, 77–95. <https://doi.org/10.1016/j.jecp.2016.01.004>.
- Stewart, I. (2016). The fruits of a functional approach for psychological science. *International Journal of Psychology*, 51, 15–27. <https://doi.org/10.1002/ijop.12184>.
- Stewart, I., & McElwee, J. (2009). Relational responding and conditional discrimination procedures: An apparent inconsistency and clarification. *The Behavior Analyst*, 32, 309–317.
- Stewart, I., & Roche, B. (2013). Relational frame theory: An overview. In S. Dymond & B. Roche (Eds.), *Advances in relational frame theory: Research and application* (pp. 51–71). Oakland: Context Press.
- Thomas, H., & Horton, J. J. (1997). Competency criteria and the class inclusion task: Modeling judgements and justifications. *Developmental Psychology*, 33, 1060–1073. <https://doi.org/10.1037/0012-1649.33.6.1060>.
- Valentin, D., & Chanquoy, L. (2012). Olfactory categorization: A developmental study. *Journal of Experimental Child Psychology*, 113, 337–352. <https://doi.org/10.1016/j.jecp.2012.05.007>.
- Viscaíno-Torres, R. M., Ruiz, F. J., Luciano, C., López-López, J. C., Barbero-Rubio, A., & Gil, E. (2015). The effect of relational training on intelligence quotient: A case study. *Psicothema: Revista Anual de Psicología*, 27, 120–127. <https://doi.org/10.7334/psicothema2014.149>.
- Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence—Revised*. San Antonio: Psychological Corporation.
- Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition*. San Antonio: Psychological Corporation.
- Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale—Third Edition (WAIS—III)*. San Antonio: Psychological Corporation.
- Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition (EVT-2)*. Circle Pines: Pearson.
- Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior*, 78, 237–248. <https://doi.org/10.1901/jeab.2002.78-237>.