



A Systematic Review of Derived Relational Responding Beyond Coordination in Individuals with Autism and Intellectual and Developmental Disabilities

Ashley R. Gibbs¹ · Christopher A. Tullis¹ · Daniel E. Conine¹ · Andrew A. Fulton¹

Accepted: 5 March 2023 / Published online: 17 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

As interest in derived relational responding has increased, so have the number of investigations evaluating interventions to promote the emergence of derived responding for individuals with autism, as well as other intellectual and developmental disabilities. However, much of the literature has focused on the relation of sameness, and less is known about interventions to facilitate derived responding in other relations. Systematic searches identified 38 studies contained in 30 articles that met inclusion criteria. These studies were analyzed according to their participants, assessment methods, experimental design, content taught, setting, teaching procedures, derived responses, outcomes, and reliability measures. The quality of the studies was measured using the Single Case Analysis and Research Framework (SCARF). The results of the current review indicate that many learners with autism spectrum disorder and other intellectual and developmental disabilities demonstrate derived relational responding beyond the relation of coordination across varied instructional content and teaching methodologies, but the quality and rigor of the published literature requires the results be interpreted with caution, leading to recommendations for future research.

Keywords Autism spectrum disorder · Derived relational responding · Intellectual and developmental disabilities

Behavior-analytic approaches to intervention with individuals with autism spectrum disorder (ASD) and other intellectual and developmental disabilities (IDD) have consistently been supported as effective (Courtade et al., 2015; Hume et al., 2021). A feature of some of these empirically derived interventions (e.g., discrete trial training) is

✉ Christopher A. Tullis
ctullis2@gsu.edu

¹ Department of Learning Sciences, Georgia State University, PO Box 3979, Atlanta, GA 30302, USA

repeated practice opportunities under meticulous stimulus control, leading to responding that appears “rote” (Stauch et al., 2017). This is a concern, particularly when language or communication repertoires are the focus of intervention. A primary feature of language is generativity, or the ability to create and understand a potentially infinite number of sentences never previously heard or said (Ming et al., 2014). As a result, facilitating generative language has become a key area of interest in behavior analysis. Methods to promote the emergence of untrained responding have been developed out of several theoretical frameworks, such as naming theory (Horne & Lowe, 1996) and stimulus equivalence (Sidman, 1971; see Gibbs & Tullis, 2021 for an overview of these two frameworks), the current review will focus on interventions derived from relational frame theory (RFT; Hayes et al., 2001) due to its foundation in the inherently derived nature of language.

Relational Frame Theory

RFT (Hayes et al., 2001) suggests that language comes from the ability to engage in a generalized repertoire of responding to stimuli in terms of other stimuli, otherwise known as relational responding (Stewart & Roche, 2013). Relational responding can be nonarbitrary (i.e., based on the formal properties of the stimuli being related) or arbitrarily applicable (i.e., based on verbal or contextual control). Arbitrarily applicable relational responding is developed via exposure to multiple exemplars and contingencies provided by the larger socioverbal community, and it appears to form the foundation of human language (Stewart & Roche, 2013). Various patterns of arbitrarily applicable relational responding, or ‘frame families’, exist and have been cited throughout literature. *Coordination*, based on sameness or similarity, develops earliest (Hayes et al., 2001), and has been the focus of the majority of research in relational responding (Gibbs & Tullis, 2021). Although coordination has been the most commonly researched topic, other frames that have been evaluated empirically include *distinction* (difference), *opposition* (opposite), *comparison* (relativity between stimuli along a specific dimension), *hierarchical* (containment, inclusion), *temporal* (sequencing) and *deictic* (perspective-taking) (Barnes-Holmes et al., 2018). As any stimulus can be related to any other stimulus in keeping with any relational frame, arbitrarily applicable relational responding has the potential to be incredibly generative.

All relational frames are defined by the properties of mutual entailment, combinatorial entailment, and transformation of stimulus function (Rehfeldt & Barnes-Holmes, 2009). *Mutual entailment* is when relations between stimuli are bidirectional and responding in one direction entails responding in the other direction (e.g., if A is the opposite of B, then B is the opposite of A; if A contains B, then B is part of A, etc.). *Combinatorial entailment* is when two stimulus relations combine and allow a third relation to be derived (e.g., if A is more than B and B is more than C, then A is more than C; if A is the opposite of B and B is the opposite of C, then A is the same as C, etc.). *Transformation of stimulus function* is when the functions of one stimulus changes, or transforms, the functions of another stimulus based on

the derived relation between the two stimuli (Dymond & Rehfeldt, 2000). As an example, a person who likes sweet desserts derives that lemon cookies (stimulus A) are sweeter than lemons (stimulus B), and then their friend tells them that lemon bars (stimulus C) taste even sweeter than lemon cookies. Later, when given a choice between lemon cookies and lemon bars for dessert, that person chooses lemon bars despite never having eaten them before. The evaluative functions of lemon bars have now been transformed based on their relation to lemon cookies (i.e., lemon bars are sweeter than lemon cookies).

Interventions Promoting Derived Responding

Previous reviews have been conducted evaluating the emergence of derived responding within and across specific relational frames. For example, Ming et al. (2014) reviewed studies that either (a) demonstrated the establishment of derived relational responding within various frames, or (b) used existing derived relational responding repertoires to teach educationally relevant skills to individuals with ASD. They concluded that programming should focus on establishing relevant patterns of derived relational responding skills that are absent via multiple exemplar training (MET), and that if learners can demonstrate specific types of derived relational responding (e.g., equivalence, naming, etc.), then those skills can and should be used to make subsequent programming more efficient (Ming et al., 2014). They recommended future research focus on the development of a standardized tool to systematically evaluate derived relational responding abilities, highlighting some of the work that has been done with this aim in mind (i.e., the Training and Assessment of Relational Precursors and Abilities; TARPA; Moran et al., 2010, 2014). Although their review encompassed studies that taught multiple relational frames (i.e., coordination, comparison, opposition, and deictic frames), it was not a complete, systematic evaluation of the literature base, and did not include a quantitative analysis nor a measure of overall study rigor.

Ming and Stewart (2017) reviewed research evaluating the relation of distinction, or difference, with both nonarbitrary and arbitrary stimuli. At the time of the review, no studies had examined how best to establish frames of distinction in individuals who were unable to demonstrate those relational responses. Based on prior work, Ming and Stewart (2017) recommended that derived relational responding to relations of distinction be conceptualized as a continuum of responding from non-arbitrary to arbitrary stimuli. They recommended future research should determine the necessary hierarchy of component skills, the most optimal sequencing of these skills, and the most effective teaching procedures for establishing them. Although fairly comprehensive through looking at research across a variety of domains, this review did not systematically search the existing literature and also did not provide a quantitative analysis. While Ming and Stewart (2017) focused on relations of distinction, Montoya-Rodríguez et al. (2017a, b) published a bibliographical review of research published between 2001 and 2015 evaluating deictic relational responding in typically developing and atypically developing populations. Their review found

that while the number of empirical articles being published had increased, the studies were most often carried out with a typically developing participant population, and training protocols for deictic relational responding have had limited development and investigation (Montoya-Rodríguez et al., 2017a, b).

Raaymakers et al. (2019) completed a systematic review of derived verbal behavior research conducted with typical and atypical populations published between 2000 and 2017. The 52 studies included in Raaymakers et al. (2019) were required to evaluate derived verbal operants (i.e., mands, tacts, intraverbals, echoics, textual, dictation, and autoclitics) from a stimulus equivalence (Sidman, 1971) or RFT framework, excluding articles whose methodologies referenced naming theory (Horne & Lowe, 1996). Their results found marked variability in the reporting on participant prerequisite skills, existing verbal repertoires, and assessments utilized, which limited analysis of potential repertoires necessary for derived relational responding to occur. Results also indicated that different instructional procedures were most effective with different verbal operants (i.e., tact and intraverbal training were most effective for establishing derived intraverbal responses, conditional discrimination training was most effective to establish derived mand and tact responses, etc.). Mastery criterion also varied between studies, with some using a percentage correct and others using block or rolling block mastery criteria. A limitation of Raaymakers et al. (2019) is that, despite requiring included articles to evaluate derived verbal behavior from either a stimulus equivalence or RFT perspective, there was no consideration of the pattern of relational responding, or frame family, that the derived verbal operants were part of, and there was little discussion of the basic processes underlying the emergence of these responses.

Two recent citation analyses found increasing interest in the use of RFT technologies to promote derived relational responding in atypically developing populations (O'Connor et al., 2017; Belisle et al., 2020a, b), with a major limitation that the relational frame of coordination was most frequently targeted for investigation. This specific relational frame was explored further by Gibbs and Tullis (2021) in a systematic review of 47 articles published since 2013 evaluating the emergence of derived responding in accordance with coordination across the theoretical bases of naming, stimulus equivalence, and RFT in learners with IDD and ASD. Gibbs and Tullis sought to determine if individuals with IDD and ASD can demonstrate the emergence of untrained coordination relations. Questions asked included whether specific learner characteristics influenced emergence, if there were specific assessment tools to identify learners capable of demonstrating emergence, and whether particular instructional procedures facilitated emergence. The findings supported the conclusion that individuals with IDD and ASD are able to demonstrate derived coordination relations, however it was emphasized that due to the low quality and rigor of many of the studies evaluated, the results should be interpreted with caution (Gibbs & Tullis, 2021).

Gibbs and Tullis (2021) found that while the expansiveness of a learner's verbal repertoire, particularly the skill of bidirectional naming, influences the emergence of derived responding, further research is required to better determine other characteristics contributing to emergence. They concluded more research is needed to determine tools with predictive validity for derived relational responding, and instructional

procedures beyond match-to-sample (MTS) are worthy of investigation (Gibbs & Tullis, 2021). Despite a comprehensive scope, a limitation of Gibbs and Tullis (2021) is they solely focused on frames coordination. It is unknown whether investigations conducted across other relational frames could shed further light on learner characteristics, assessment tools, and instructional procedures that may facilitate derived relational responding in learners with IDD and ASD. As a result, a systematic review of the literature evaluating the emergence of derived relational responding beyond the frame of coordination in individuals with IDD and ASD is warranted.

Research Questions and Statement of Purpose

The purpose of this review is to extend the work of Gibbs and Tullis (2021) by systematically synthesizing and analyzing the results of research facilitating derived relational responding beyond the frame of coordination in learners with IDD and ASD, and to make recommendations for additional areas of investigation. This review aims to answer several questions. First, is there sufficient, high-quality evidence that individuals with IDD and ASD have demonstrated derived relational responding beyond coordination in the context of empirical research? Second, is there evidence to indicate the presence of distinct learner characteristics or profiles that influence the development of this skill? Third, are there particular assessment instruments that are ideal to determine a learner's relational skill repertoire? Last, are there specific instructional procedures best suited to develop relational responding in this population?

Method

Search Procedures

Systematic searches of peer-reviewed journal articles were conducted using the APA PsycInfo®, CINAHL Plus, Proquest Central, Pubmed, and Google Scholar electronic databases. Two Boolean searches of each database were conducted, the first using *intellectual and developmental disability* AND (a) *relational frame theory*, (b) *derived relational responding*, or (c) *relational frame(s)*, and the second using *autism spectrum disorder* in combination with the previously listed search terms. When searches returned 1000 or more articles, the results were further narrowed by including the terms (a) *comparison*, (b) *opposition*, (c) *distinction*, (d) *hierarchical*, and (e) *deictic*. In addition to database searches, ancestry searches from the reference lists of identified studies, and reviews of published citation analyses were also completed. A total of 2870 articles were identified from electronic databases and evaluated for inclusion. After evaluation, 2768 articles were excluded and 67 articles were identified as duplicates identified by multiple databases, leaving 35 articles to be screened for full text review. Of those articles, 23 were identified as eligible for inclusion, with an additional seven articles found after ancestry searches and citation

analysis reviews, resulting in a total of 30 articles containing 38 studies. See Fig. 1 for a visual representation of this search.

Inclusion and Exclusion Criteria

Studies were required to be published in English in peer-reviewed journals, and have at least one participant with a formal diagnosis of ASD or IDD whose data could be disaggregated for analysis. Articles were further required to measure generalized or derived responding. Review articles (e.g., Ming & Stewart, 2017; Montoya-Rodríguez et al., 2017a, b), non-experimental articles (e.g., Kavanagh et al., 2020; McHugh & Reed, 2008), articles evaluating assessment tools without an additional evaluation of training and emergence (e.g., Pomorska et al., 2020), and articles evaluating relational responding within the frame of coordination were excluded from this review.

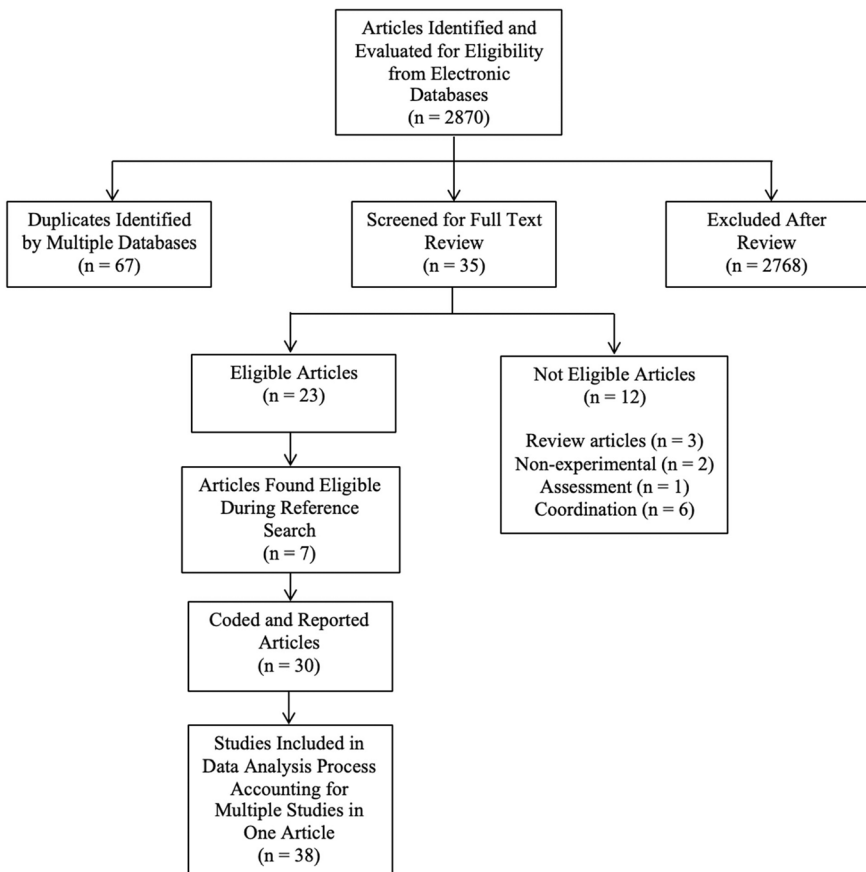


Fig. 1 Visual representation of systemic search procedures

Data Classification

Article Analysis

All articles were analyzed using the following criteria: (a) participants (chronological age, diagnosis); (b) assessments conducted and associated results (e.g., PPVT-4, WISC-IV, etc.); (c) experimental design; (d) relational frame family or families evaluated (e.g., comparison, opposition, distinction, hierarchical, deictic, temporal, or multiple frames); (e) content taught (e.g., metaphor comprehension, working memory, etc.); (f) setting (e.g., school, home, etc.); (g) teaching procedures (e.g., multiple exemplar teaching, conditional discrimination training, etc.); (h) generalized responses measured; (i) outcomes (whether participants demonstrated derived relational responding and whether responding was variable); and (j) reliability and fidelity (reporting of interobserver agreement and procedural fidelity). Similar outcome criteria to those used by Gibbs and Tullis (2021) were utilized in the current review, with outcomes classified as positive, negative, or variable based on the demonstration of derived relational responding. Positive outcomes indicated all participants demonstrated derived responding when tested, and negative outcomes indicated no participants demonstrated evidence of derived responding. Variable outcomes indicate either (a) some participants demonstrated derived responding while others did not, (b) participants demonstrated some, but not all derived responses, or (c) some participants required additional intervention to demonstrate derived responding. The results of the article analysis is presented in Table 1.

Single Case Analysis and Review Framework v2.0

The Single-Case Analysis and Review Framework (SCARF; Ledford et al., 2020), a tool to evaluate the quality, rigor, and outcomes of single case design studies, was completed for all studies which included an appropriate graphic display of data (i.e., line or bar graphs showing a minimum of two primary comparison conditions, such as baseline and intervention or pre- and posttest data). The SCARF quantifies the rigor, quality and breadth of measurement, and outcomes across 10 elements. This data provides a scatterplot representation of an article's overall quality and rigor, the extent to which generalization outcomes are internally valid, and the extent to which maintenance outcomes are measured separately in time from intervention (Ledford et al., 2020). Within each scatterplot, the majority of the included data points existing in the upper right quadrant of the graph is indicative of the best outcomes for each measure. Elements associated with the quality of a single-case design evaluated by the SCARF include (a) participant description (i.e., demographics, formal assessment results, general learner information, inclusion and exclusion criteria), (b) dependent variable descriptions (i.e., operational definitions, examples and non-examples, measurement system and utilization), (c) condition descriptions (i.e., adequate description of procedures, dosage, setting, and implementors), (d) social validity (i.e., importance of a behavior to key stakeholders and society), (e) ecological validity (i.e., the relevance and reliable implementation of an intervention outside of controlled settings), (f) generalization measurement and measures (i.e., whether

Table 1 Summary of articles evaluating derived relational responding beyond coordination in individuals with intellectual and developmental disabilities

Author	Participants	Assessment	Design	Frame family/families	Content
Balruschat et al. (2012)	Three males with ASD (6–9 years old)	AGTB 5–12 (normed and standardized German test of working memory)	Multiple baseline across participants with pretest/posttest	Temporal relations	Working memory
Barron et al. (2019)	Two children with ASD (6 years old)	PEAK-T preassessment (combined receptive and expressive deictic scores of 7 out of 32)	Multiple baseline across skills	Deictic relations	Here-there & Then-later
Belisle et al. (2016)	Three males with ASD (12–18 years old)	PEAK-DT (scores 112–135) PEAK-T assessment (draft version)	Multiple baseline across participants with embedded multiple probe	Deictic relations	Single reversal you-I
Belisle et al. (2020a, b) Experiment 1	Two males with ASD (14–19 years old)	WISC-IV (IQ scores 84–90) PEAK-E preassessment ABLLS-R (maximum score) VB-MAPP (170)	Multiple baseline across skills with an embedded multiple probe	Comparative relations	Bigger-smaller & Faster-slower
Belisle et al. (2020a, b) Experiment 2	Two males with ASD (14–19 years old)	WISC-IV (IQ scores 84–90) PEAK-E preassessment ABLLS-R (maximum score) VB-MAPP (170)	Multiple baseline across participants with embedded multiple probe	Comparative relations	Bigger-smaller
*Cassidy et al. (2011) Experiment 2	Three males and five females with educational and behavioral difficulties (11–12 years old)	WISC-IV UK (IQ scores 70–92) RAI (scores 29–60.7)	Pre-test/Post-test quasi-experimental design with no control group	Coordination/Comparison/ Opposition relations	Nonsense syllables
*Dunne et al. (2014) Study 2	Two males and two females with ASD (4–5 years old)	VB-MAPP (scores 44–139.5)	No formal design; Series of testing & training	Opposition relations	Picture stimuli that differed along one physical dimension (big-small; long-short; wet-dry; hot-cold; happy-sad; clean-dirty; empty-full; dark-light; rough-smooth; heavy-light)
*Dunne et al. (2014) Study 3	One male and one female with ASD (4–5 years old)	VB-MAPP (scores 93.5 and 139.5)	No formal design; Series of testing & training	Distinction relations	Picture stimuli that differed along one dimension (color, length, texture, shape)

Table 1 (continued)

Author	Participants	Assessment	Design	Frame family/families	Content
*Dunne et al. (2014) Study 4	One male and one female with ASD (4–5 years old)	VB-MAPP (scores 93,5 and 139.5)	No formal design; Series of testing & training	Comparative relations	More-less
Gale and Stewart (2020)	Three males with ASD (6 years old)	VABS receptive (age equivalencies <3y–3y4m) VABS expressive (age equivalencies <3y–3y8m) PPVT-IV (age equivalencies 4y6m–4y9m) TARPA	Multiple probe across embedded multiple probe across stimulus sets	Comparative relations	More-less
Gilroy et al. (2015)	Three males with ASD (8–13 years old)	No formal assessment	Multiple probe across participants	Deictic relations	Simple you-I, reversed you-I, and double reversed you-I and now-then relations
*Gorham et al. (2009) Experiment 2	Two males with ASD (6–9 years old)	No formal assessment	No formal design; Series of testing & training	Comparative relations	More-less
Gould et al. (2011)	Two males and one female with ASD (3–5 years old)	No formal assessment	Multiple probe across participants	Deictic relations	Perspective taking
Grannan and Rehfeldt (2012)	Two males with ASD (5 years old)	PLS (age equivalencies 3y6m–3y7m) VBMAP (Level 3 learners)	Multiple probe across participants	Hierarchical relations	Class inclusion
Greer and Yuan (2008) Experiment 1	Three males and one female with language delays, health impairments, developmental delays and ASD (6–7 years old)	VABS (2 participants, adaptive behavior composite 73), Stanford-Binet Intelligence Scale (1 participant, IQ score 76), EOWPVT (1 participant, total score 66)	Time-lagged multiple probe across participants	Deictic relations	Past tense verbs in an autoclitic function
Greer and Yuan (2008) Experiment 2	Three children with developmental delays (4 years old)	VABS (1 participant, adaptive behavior composite 54), Stanford-Binet Intelligence Scale (1 participant, IQ score 83), PLS (1 participant, auditory comprehension 57, expressive communication 72)	Time-lagged multiple probe across participants	Deictic relations	Past tense verbs in an autoclitic function

Table 1 (continued)

Author	Participants	Assessment	Design	Frame family/families	Content
Jackson et al. (2014)	Three males with ASD (5–6 years old)	CARS (17.5–30) ToM tests	ABACADA multiple probe with pretest/posttest	Deictic relations	Simple, single reversal, and double reversal I-you, Here-then, and Now-then
*Kent et al. (2017) Study 2	11 males with ASD (4–13 years old)	PPVT (scores < 40–89) K-BIT (scores 20–101)	No formal design; Series of testing & training	Coordination/ Distinction/ Comparison/ Opposition relations	Nonarbitrary and arbitrary picture stimuli
*Kent et al. (2017) Study 3	Four males with ASD (3–4 years old)	PPVT (scores 80–104) K-BIT (scores 73–98)	No formal design; Series of testing & training	Coordination/ Distinction/ Comparison/ Opposition relations	Nonarbitrary and arbitrary picture stimuli
Lee et al. (2019)	Two males and one female* with ASD (5–8 years old) *did not complete study	WISC-IV and WPPSI-R (2 participants, IQ scores 87 and 115) CARS (scores 30–31) DBACA	Multiple probe across behaviors and participants	Hierarchical/ Distinction/ Coordination relations	Metaphor
Lovett and Rehfeldt (2014)	Three young adults with ASD (17–18 years old)	SLDT-A (score 67) ToMI (scores 16.5–20)	Multiple probe across participants	Deictic relations	Perspective taking
Luke et al. (2011) Experiment 1	Three males and one female with ASD (5–7 years old)	New York State Curricular Standards, Preschool Inventory of Repertoires for Kindergarten, Verbal Capabilities Checklist	Pretest/Posttest	Spatial relations, Deictic relations	Autoclitics
Ming et al. (2018)	Three individuals with ASD (8–19 years old)	PPVT-4 (age equivalencies 6y5m–7y 11m)	Nonconcurrent multiple baseline across participants	Hierarchical relations	Class inclusion
**Molina-Cobos and Amador-Castro (2010) Experiment 1	Nine males and six females with Down's syndrome and developmental delay (5–35 years old)	No formal assessment	ABA design	Deictic relations	Preferences of another
***Molina-Cobos and Amador-Castro (2010) Experiment 2	Three males and three females with Down's syndrome and developmental delay (7–35 years old)	No formal assessment	ABCA design	Deictic relations	Preferences of another

Table 1 (continued)

Author	Participants	Assessment	Design	Frame family/families	Content
Montoya-Rodríguez et al. (2017a)	One adult male with Down syndrome (24 years old)	No formal assessment	AB with pretest/posttest	Deictic relations	Simple, reversed, and double reversed I-you, Here-there, and Now-then
**Montoya-Rodríguez and Molina-Cobos (2019)	Three males and two females with Down's syndrome (18–34 years old)	WAIS-IV (IQ scores 47–55)	ABA multifactorial	Deictic relations	Perspective taking (Interpersonal [I-he], spatial [here-there], and temporal [now-then] relations)
*Murphy and Barnes-Holmes (2009) Experiment 1	Two males and one female with ASD (7–11 years old)	Preschool Inventory of Repertoires for Kindergarten	No formal design; Series of testing & training	Comparative relations	More-less mands
*Murphy and Barnes-Holmes (2009) Experiment 2	Two males with ASD (7–11 years old)	Preschool Inventory of Repertoires for Kindergarten	No formal design; Series of testing & training	Comparative relations	More-less mands
Murphy and Barnes-Holmes (2010)	Four males with ASD (14 years old) *only one completed study	No formal assessment	No formal design; Series of testing & training	Comparative relations	More-less mands
**Newsome et al. (2014)	One male with Down syndrome, one female with ASD, one female with language difficulties due to a seizure disorder, one female with comprehension difficulties due to fetal drug exposure, and one male with an IEP due to comprehension difficulties (9–12 years old)	AIMSweb™ QRI-4 (comprehension questions and story retell) Expressive writing measure	ABC with embedded multiple probe	Hierarchical/ Coordination/ Distinction relations	Reading comprehension
Pérez-González et al. (2007)	Two children with PDD (6–8 years old)	PPVT (age equivalencies 2.92–6 years old)	Multiple baseline across participants	Opposition relations	Intraverbals
Persicke et al. (2012)	Three males with ASD (5–6 years old)	No formal assessment	Multiple baseline across participants	Hierarchical/ Distinction/ Coordination relations	Metaphor
Persicke et al. (2013)	Three males with ASD (6–7 years old)	No formal assessment	Multiple baseline across participants	Opposition relations	Sarcasm

Table 1 (continued)

Author	Participants	Assessment	Design	Frame family/families	Content
Ranick et al. (2013)	Three males with ASD (6–9 years old)	No formal assessment	Multiple baseline across participants	Distinction relations	Deceptive statements
Speckman et al. (2012) Experiment 1	Two males with language delays (4–5 years old)	REEL (one participant, expressive age equivalence 12 months, receptive age equivalence 14 months) Rosetti Infant Toddler Language Scale (one participant, skills from 15–27 months)	Delayed multiple probe across participants	Comparative relations	Autoclitics
Speckman et al. (2012) Experiment 2	One male with language delay (4 years old)	CELF-2 (average expressive and receptive language skills, poor pragmatic skills and dysfluent speech)	Nonconcurrent multiple probe across participants	Comparative relations	Autoclitics
Zagrabaska-Swiatkowska et al. (2020)	Three males with ASD (23–26 years old)	PPVT-IV (age equivalencies 10y4m–13y4m)	Multiple baseline across participants	Hierarchical relations	Class inclusion
Setting ^a	Teaching procedures ^b	Generalized responses		Outcomes	IOA and fidelity
H	MET	Digit-span backwards task with novel stimuli AGTB 5–12 posttest scores	Positive	Interobserver agreement 87% No procedural fidelity reported	
UC	MET	Here-there and Then-later responses with novel stimuli Novel behaviors in the context of deictic cues (drawing, writing)	Positive	Interobserver agreement 100% No procedural fidelity reported	
S	MET	Single reversal you responses with novel stimuli, emergence of single reversal I responses	Variable	Interobserver agreement 100% No procedural fidelity reported	
S	RT	A-C and C-A (Bigger-smaller and faster-slower) combinatorially entailed relations Transformation of stimulus functions of entailed A-C and C-A (bigger-smaller and faster-slower) relations	Positive	Interobserver agreement 100% No procedural fidelity reported	
S	RT	A-C, A-D, A-E, B-D, B-E, and C-E (Bigger-smaller) combinatorially entailed relations Transformation of stimulus functions of entailed A-E and E-A (bigger-smaller) relations	Positive	Interobserver agreement 100% No procedural fidelity reported	

Table 1 (continued)

Setting ^a	Teaching procedures ^b	Generalized responses	Outcomes	IOA and fidelity
Not specified	MET	Same/opposite, and more than/less than responses with novel stimuli WISC-IV UK and RAI posttest scores	Positive	No interobserver agreement or procedural fidelity reported
S	MET	Oppositional relational responding with novel stimuli	Variable	No interobserver agreement or procedural fidelity reported
S	MET	Arbitrary responding to combinatorially entailed A-C (same/different) relations with novel stimuli	Variable	No interobserver agreement or procedural fidelity reported
S	MET	Arbitrary responding to combinatorially entailed A-C (more/less) relations with novel stimuli	Variable	No interobserver agreement or procedural fidelity reported
S	TARPA	Arbitrary mutually and combinatorially entailed more-less responses with novel stimuli	Positive	Interobserver agreement 98.6% No procedural fidelity reported
H	RT within a story reading context	Deictic relations emitted with a novel experimenter using novel stimuli	Positive	Interobserver agreement 98.9% Procedural fidelity 100%
S	MET	Derived transitive inference of more-less relations	Variable	Interobserver agreement 97.7% No procedural fidelity reported
H	MET	Perspective taking with untrained picture stimuli	Variable	Interobserver agreement 100%
C	TT, MTS	Perspective taking in the natural environment with real people Derived categorization intraverbal responses	Variable	No procedural fidelity reported Interobserver agreement 99% Procedural fidelity 99–100%
S	SEI, MEI	Derived “-ed” suffixes to novel regular and irregular verbs	Variable	Interobserver agreements 100% Procedural fidelity 100%
S	SEI, MEI	Derived “-ed” suffixes to novel regular and irregular verbs	Variable	Interobserver agreement 97–100% Procedural fidelity 100%
H, UC	MET	ToM test performance	Negative	Interobserver agreement 91% Procedural fidelity 92%
S	MET	Coordination/ distinction/ comparison/ opposition relational responses with novel stimuli	Variable	No interobserver agreement or procedural fidelity reported
S	MET	Coordination/ distinction/ comparison/ opposition relational responses with novel stimuli	Positive	No interobserver agreement or procedural fidelity reported
UC	IT	Understanding of novel metaphors	Positive	Interobserver agreement 97% Procedural fidelity 95–97%

Table 1 (continued)

Setting ^a	Teaching procedures ^b	Generalized responses	Outcomes	IOA and fidelity
S	MET	Perspective taking with novel video presentations SLDT-A and ToMI posttest scores	Variable	Interobserver agreement 100% No procedural fidelity reported
S	MEI	Listener and speaker responses to four spatial relations in autoclitic frames (on/on top, under, left, right)	Positive	Interobserver agreement 97–100% Procedural fidelity 93–100%
S	MET	Class inclusion with novel stimuli	Positive	Interobserver agreement 87.5–100% Procedural fidelity 87.5–100%
S/DC	O	Posttest of other person's preferences	Variable	Interobserver agreement 100% No procedural fidelity reported
S/DC	O, DT	Posttest of other person's preferences	Positive	Interobserver agreement 100% No procedural fidelity reported
DC	MET within interactive activities	Reversed and double reversed deictic relations with novel stimuli	Variable	Interobserver agreement 95% No procedural fidelity reported
DC	False belief training protocol	ToM and simple, reversed, and double reversed deictic relational responses posttest scores	Variable	Interobserver agreement 100% No procedural fidelity reported
S	CDT	Derived transfer of more-less relational mand functions	Variable	Interobserver agreement 100% No procedural fidelity reported
S	CDT without programmed mand consequences during testing	Derived transfer of more-less relational mand functions	Variable	Interobserver agreement 100% No procedural fidelity reported
S	MTS and CDT	Derived transformation of more-less relational mand functions	Positive	Interobserver agreement 95% No procedural fidelity reported
C	MET within PT	Performance on untrained HCSD tasks QRI-4 posttest scores	Variable	Interobserver agreement for pre and posttest QRI-4 comprehension questions 91% No procedural fidelity reported
S	IT, RIT	Expressive writing measure posttest scores Reverse intraverbal responses	Variable	Interobserver agreement 99.8% No procedural fidelity reported
H	MET	Understanding of novel metaphors	Variable	Interobserver agreement 97% No procedural fidelity reported

Table 1 (continued)

Setting ^a	Teaching procedures ^b	Generalized responses	Outcomes	IOA and fidelity
H, CM	Rules/video training and in vivo training	Detection and responding to novel sarcastic statements	Positive	Interobserver agreement 99.4% No procedural fidelity reported
H	MET	Detection and responding to deceptive statements delivered by a peer/sibling	Positive	Interobserver agreement 99.6% No procedural fidelity reported
S	SEI, MEI	Detection and responding to novel deceptive statements	Variable	Interobserver agreement 100% Procedural fidelity 100%
S	SEI, MEI	Derived “-er” suffixes to novel adjectives	Variable	Interobserver agreement 100% Procedural fidelity 100%
DC	MET	Derived “-er” suffixes to novel adjectives Class inclusion with novel stimuli	Positive	Interobserver agreement 90–100% Procedural fidelity 100%

^aNot included in SCARF analysis based on study design

^{**}Not included in SCARF analysis due to absence of graphical display or inability to disaggregate data from graphical display

^aC clinic, CM community, DC day center, H home, S school, UC university clinic or laboratory

^bCDT conditional discrimination training, DT discrimination training, HCS D hierarchical relating under contextual cues of ‘same’ or ‘different’, HRT hierarchical relation training, IT intraverbal training, MEI multiple exemplar instruction, MET multiple exemplar training, PT precision teaching, MTS match to sample, O observation, REP relational evaluation procedure, RT relational training, RIT reverse intraverbal training, SEI single exemplar instruction, TARPA training and assessment of relational precursors and abilities, TT tact training

stimulus or response generalization occurred and how it was measured), and (g) maintenance measurement (i.e., whether evidence of continued behavior change occurred and how it was measured). Elements associated with the rigor of a single-case design evaluated by the SCARF include (a) dependent variable reliability (i.e., interobserver agreement, or the extent to which independent observers measure the same behavior), (b) implementation fidelity (i.e., degree to which experimental procedures are implemented as intended), and (c) sufficiency of data (i.e., data allow for analysis of level, trend, and variability within and across conditions).

Of the 30 articles within this review, 21 could be assessed with the SCARF. Of the remaining nine, six were excluded from analysis because they did not use a formal single-case design and three were excluded secondary to lacking a graphical display, or because the graphical display did not allow for disaggregation of the data (e.g., Molina-Cobos & Amador-Castro, 2010). Among the six articles that did not use a formal single-case design, Cassidy et al. (2011) utilized a pre-test/post-test quasi-experimental design with no control group, while Dunne et al. (2014), Gorham et al. (2009), Kent et al. (2017), and Murphy and Barnes-Holmes (2009, 2010) utilized a series of pre-post tests without a clear experimental design. We are unaware of any tools suited to assess the quality of such designs. The articles excluded from the SCARF analysis, as well as the reason for their exclusion, are noted in Table 1. As in Gibbs and Tullis (2021), each directly trained relation was evaluated as its own comparison and received its own entry ($N=68$), and the emergence of derived responding was categorized in the SCARF as instances of response generalization (i.e., the measurement of different specific behaviors than the specific behaviors taught in the study), or both response and stimulus generalization (i.e., the measurement of a target behavior performed with materials separate from those used in teaching).

Interrater Reliability

Interrater reliability for the descriptive article coding and the SCARF was assessed for 31% of the articles selected for inclusion by the fourth author. Reliability was calculated for each article by dividing the total number of agreements by the total number of agreements and disagreements, then multiplying the quotient by 100 to calculate the percent agreement. Mean reliability for the descriptive coding was 93% (range, 81%–100%), while mean reliability for the SCARF was 95% (range, 80% – 100%). Interrater reliability was unable to be assessed for the systematic database searches, which is a limitation that should be addressed in future reviews.

Results

Article Variables

Participants

The chronological age ranges, gender, and diagnoses of participants in each study are included in Table 1, and are summarized in Table 2. A total of 122 participants

Table 2 Demographic characteristics of participants across studies

Characteristic	<i>N</i>	Percentage
Gender		
Male	91	74.6%
Female	17	13.9%
Not reported	14	11.5%
Race/Ethnicity		
Not reported	122	100%
Age		
3–5 years	17	13.9%
5–9 years	40	32.8%
9–13 years	20	16.4%
13–18 years	9	7.4%
18–34 years	10	8.2%
Not reported for each participant	26	21.3%
Diagnosis		
ASD	82	67.2%
Down syndrome + developmental delay	15	12.3%
Down syndrome	7	5.7%
Educational and/or behavioral difficulties	8	6.6%
Language difficulties	4	3.3%
Developmental delay	3	2.5%
PDD	2	1.6%
Comprehension difficulties	2	1.6%

(accounting for multiple experiments with the same individuals) were included in the 30 articles evaluated and completed all study procedures. Gender was reported for 108 participants, with the majority of participants being male (74.6%, $N=91$) as compared to female (13.9%, $N=17$). Data on the race and ethnicity of study participants were not reported in any of the 30 articles. The majority of participants had a diagnosis of ASD (67.2%, $N=82$). Other diagnoses reported include Down's syndrome and developmental delay (12.3%, $N=15$), Down's syndrome (5.7%, $N=7$), educational and/or behavioral difficulties (6.6%, $N=8$), language difficulties (3.3%, $N=4$), developmental delay (2.5%, $N=3$), pervasive developmental disorder (1.6%, $N=2$), and comprehension difficulties (1.6%, $N=2$). Participant ages in the included studies ranged between three and 35 years old.

Over half (71.1%, $N=27$) of the included 38 studies included three or fewer participants, while 21.1% ($N=8$) included between four and six participants, 2.6% ($N=1$) of studies included between seven and 10 participants, and 5.3% ($N=2$) included more than 10 participants.

Assessments

Assessment information for each study is provided in Table 1, and a more detailed breakdown of the frequency with which each assessment was used is provided in Table 3. A majority of the 38 studies included in this review (73.7%, $N=28$) provided

Table 3 Assessments used

Assessment	<i>N</i>	Percentage
Norm-Referenced		
WISC-IV	3	7.9%
WISC-IV UK	1	2.6%
Vineland Adaptive Behavior Scales	3	7.9%
PPVT-IV	6	15.8%
PEAK-DT	1	2.6%
AGTB 5–12	1	2.6%
PLS	2	5.3%
Stanford-Binet	2	5.3%
Expressive One Word Picture Vocabulary Test	1	2.6%
Childhood Autism Rating Scales	2	5.3%
K-BIT	2	5.3%
WPPSI-R	1	2.6%
SLDT-A	1	2.6%
ToMI	1	2.6%
WAIS-IV	1	2.6%
AIMSweb	1	2.6%
REEL	1	2.6%
CELF-2	1	2.6%
Criterion-Referenced		
VB-MAPP	6	15.8%
ABLLS-R	2	5.3%
PEAK-E	2	5.3%
PEAK-T	2	5.3%
PIRK	3	7.9%
TARPA	1	2.6%
Verbal Capabilities Checklist	1	2.6%
QRI-4	1	2.6%
Rosetti Infant-Toddler Language Scale	1	2.6%
DBACA	1	2.6%
RAI	1	2.6%

participant assessment information, such as scores or learner characteristics based on assessment results. Of the studies that included assessment information, 31.6% ($N=12$) reported using exclusively norm-referenced measures, and also provided participant information such as intelligence quotient scores, percentile ranks, and age equivalencies. Nine of studies (23.7%) reported using exclusively criterion-referenced measures, and provided participant information such as criterion achieved and existing language repertoires. Seven of the studies (18.4%) reported using both norm- and criterion-referenced measures.

Experimental Design

The experimental design used most frequently in the included 38 studies was a concurrent or nonconcurrent multiple baseline or multiple probe design across participants (42.1%, $N=16$). The second most frequently used arrangement was not a formal experimental design, but rather phases of training and testing relations (23.1%, $N=9$). One study (2.6%) used a multiple baseline design across skills, one study (2.6%) used a pre- and post-intervention probe design across participants, and one study (2.6%) used a pre-test/post-test quasi-experimental design with no control group. An additional five studies (13.2%) used a multiple baseline design with an embedded pretest and posttest or multiple probe to measure the effect of the intervention on the emergence of derived responding. Four studies (10.5%) employed a variation of a reversal design (e.g., A-B-A, A-B-C-A, or A-B-A-C-A-D-A), and of the remaining two studies, one (2.6%) utilized an A-B design with a pretest and posttest, and one (2.6%) utilized a multiple probe with an embedded A-B-C design.

Relational Frame and Content Taught

The relational frames and content taught within the included studies can be found in Table 4. Across these studies, the deictic relational frame was most commonly targeted for teaching ($N=12$, 31.6%). The second most commonly targeted relational frame was comparison ($N=10$, 26.3%). Three studies targeted opposition relations (7.9%) to teach content including understanding sarcasm. An additional two studies (5.3%) focused on distinction relations, three studies (7.9%) targeted hierarchical relations, and one study (2.6%) targeted temporal relations. Three studies (7.9%) targeted relational framing itself, teaching multiple relational frames in sequence using stimuli such as nonsense syllables and nonarbitrary and arbitrary pictures. The remaining four studies (10.5%) targeted multiple relational frames while teaching specific content.

Setting

The majority of studies (86.8%; $N=33$) were conducted in participants' natural environments (e.g., home, community, day program, school). Of the remaining four studies, one (2.6%) was conducted in a university-based clinic setting (Barron et al., 2019), one (2.6%) used a combination of home and university-based settings for different participants (Jackson et al., 2014), two (5.3%) utilized a clinic setting (Grannan & Rehfeldt, 2012), and one (2.6%) did not specify its setting (Cassidy et al., 2011).

Teaching Procedures

Half (50.0%, $N=19$) of the included 38 studies used multiple exemplar training (MET) to teach content across relational frames and evaluate the emergence of derived responding. Other teaching procedures included: (a) single exemplar instruction (SEI) and multiple exemplar instruction (MEI) (10.5%); (b) relational training consisting of reinforcement,

Table 4 Relational frame and content taught

Relational frame/Content	<i>N</i>	Percentage
Deictic	12 / 38	31.6%
Simple, reversed, and/or double reversed here-there, then-later, you-I, I-he, and now-then relations	5 / 12	41.7%
Perspective taking	3 / 12	25%
Identifying preferences of another	2 / 12	16.7%
“-ed” suffixes as autoclitics	2 / 12	16.7%
Comparison	10 / 38	26.3%
Bigger-smaller, faster-slower, more-less relations	8 / 10	80%
“-er” suffixes as autoclitics	1 / 10	10%
More-less mands	1 / 10	10%
Opposition	3 / 38	7.9%
Understanding sarcasm	1 / 3	33.3%
Intraverbals	1 / 3	33.3%
Physical dimensions	1 / 3	33.3%
Distinction	2 / 38	5.3%
Physical dimensions	1 / 2	50%
Understanding and responding to deceptive statements	1 / 2	50%
Hierarchical	3 / 38	7.9%
Class inclusion (e.g., animals and cats)	2 / 3	66.7%
Categorization	1 / 3	33.3%
Temporal relations	1 / 38	2.6%
Working memory		
Relational framing	2 / 38	5.7%
Nonsense syllables	1 / 2	50%
Nonarbitrary and arbitrary pictures	1 / 2	50%
Multiple frames	4 / 38	10.5%
Understanding metaphors	2 / 4	50%
Reading comprehension	1 / 4	25%
Autoclitic frames for spatial relations	1 / 4	25%

prompting, and error correction as needed (7.9%); (c) conditional discrimination training (CDT) either individually or in combination with match to sample (MTS) procedures (7.9%); (d) MET within a precision teaching (PT) instructional paradigm (2.6%); (e) observation either individually or in combination with discrimination training (5.3%); (f) tact training in combination with MTS procedures (2.6%); (g) intraverbal training alone or in combination with reverse intraverbal training (5.3%); (h) an adaptation of the TARPA (Moran et al., 2010, 2014) (2.6%); (i) a false belief training protocol (2.6%); and (j) a training package of providing rules, modeling, practice, and feedback followed by in vivo training (2.6%). See Fig. 2 for a visual representation of the frequencies of different teaching procedures used per relational frame.

Derived Responses

Studies included in this review evaluated the emergence of derived responding, which varied across investigations. The vast majority of the included studies (89.5%, $N=34$) investigated the emergence of (a) mutually entailed relations, (b) combinatorially entailed

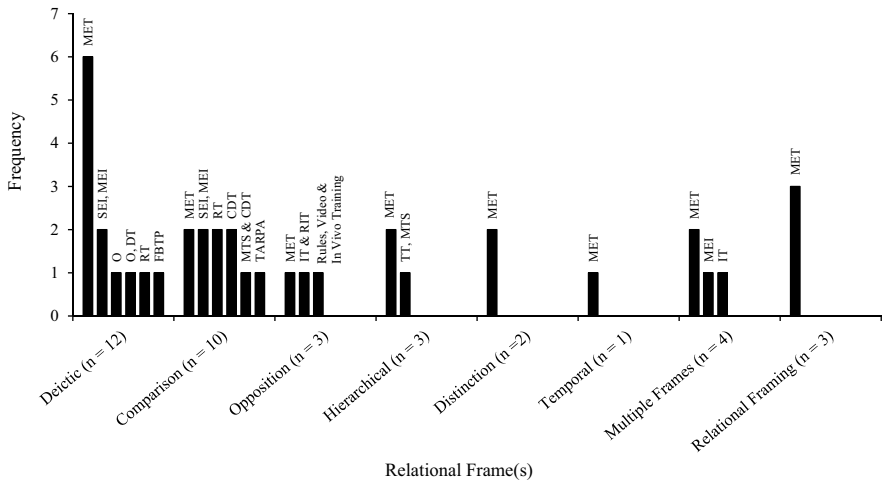


Fig. 2 Visual representation of the frequencies of interventions used to teach each relational frame or frames. CDT = conditional discrimination training, DT = discrimination training, FBTP = false belief training protocol, IT = intraverbal training, MEI = multiple exemplar instruction, MET = multiple exemplar training, MTS = match to sample, O = observation, RT = relational training, RIT = reverse intraverbal training, SEI = single exemplar instruction, TARPA = training and assessment of relational precursors and abilities, TT = tact training

relations, and/or (c) transformation of stimulus function following the direct training of specific relations. Additional outcome variables were assessed in several studies, including the emergence of novel drawing and writing behaviors in the presence of deictic cues (Barron et al., 2019), as well as scores on various assessments administered as pretests and posttests to evaluate the role the acquisition of derived relational responding has on broad outcomes such as working memory (Baltruschat et al., 2012), intelligence quotient (Cassidy et al., 2011), theory of mind (ToM; Jackson et al., 2014; Lovett & Rehfeldt, 2014), and reading comprehension (Newsome et al., 2014).

Study Outcomes

Of the 38 studies evaluated, 42.1% reported positive outcomes ($N=16$), while 55.3% reported variable outcomes ($N=21$). The main reason for this classification was that some participants required additional or more extensive intervention to demonstrate derived responding (66.6%, $N=14$), while the remaining seven (33.3%) article outcomes were classified as variable due to some participants demonstrating emergence of some, but not all derived relations. Only one study, Jackson et al. (2014), reported a negative outcome, with all participants requiring explicit training to demonstrate mastery of complex deictic relations, and no improvements seen in posttest ToM scores.

Reliability and Fidelity

The majority of the 38 studies in this review (84.2%, $N=32$) measured and reported interobserver reliability during a proportion of all teaching and testing sessions, with

agreement ranging between 87–100%. Of the six studies that did not report inter-observer agreement (15.8%), four reported using electronic devices (e.g., computer, laptop, tablet, etc.) and automated data collection procedures (Cassidy et al., 2011; Dunne et al., 2014 experiments two, three, and four). Procedural fidelity or treatment integrity data was only reported in 28.9% of the included studies ($N = 11$), with fidelity ranging between 87.5–100%.

SCARF Data Analysis

Primary Data Measurement

Figure 3 depicts the results of the primary data measurement for the 68 relations across the 21 articles able to be analyzed. As stated previously, the data displayed in this graph depicts the outcomes of the direct training done in each study, not the outcomes of the emergence of untrained relations, which is assessed in the generalization measurement graph. The scatterplot is designed such that the x-axis designates overall study quality and rigor, with quality increasing as data points move toward the right, and the y-axis designates primary outcomes, with improved effects as data points move up the axis. Low rigor was classified as a score of one or below, moderate rigor was classified as a score of between one and three, and high rigor was classified as a score of three and above (Ledford et al., 2020). Nine studies (13.6%) had lower quality evidence of minimal or weak effects, while one study (1.5%) had lower quality evidence of positive effects. Fifteen studies (22.1%) had moderate quality evidence of minimal or weak effects, while 39 studies (59.1%) had moderate quality evidence of positive effects. Four studies (6.1%) had higher quality evidence of positive effects. Although the majority of the studies evaluated reported positive effects

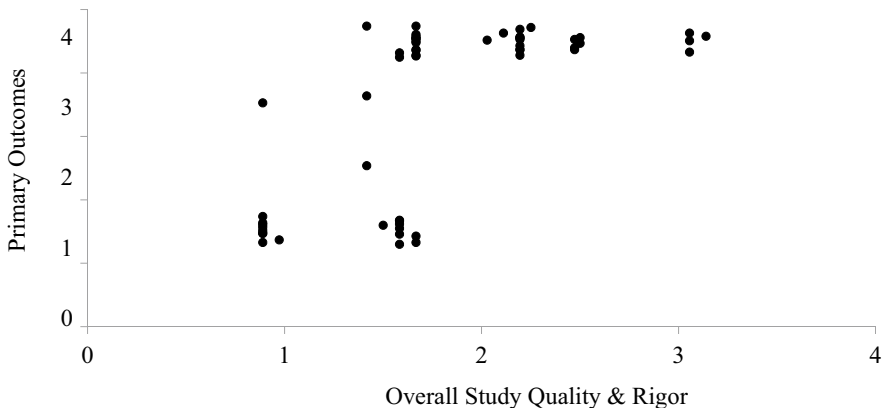


Fig. 3 Results for primary data measurement. Study quality and rigor increases as data points move down the x-axis, while primary outcomes are increasingly positive as data points move up the y-axis. While many of the studies included for review had positive outcomes, as indicated by the clustering of data points at the top of the graph, many of the studies had weak to moderate overall quality and rigor, as indicated by the clustering of data points on the left-hand side of the graph

(64.7%, $N=44$), 94.1% ($N=64$) of the included studies had low (15.2%, $N=10$) to moderate (77.9%, $N=54$) quality and rigor, and thus their results should be accepted with reservations.

Generalization Measurement

Figure 4 depicts the results of the generalization measurement for the 68 relations analyzed. The scatterplot indicates that most of the studies reported positive (72.1%, $N=49$) or moderate (22.1%, $N=15$) generalization effects, with only four studies (5.9%) reporting negative effects. Generalization was measured with pre- and post-tests (45.6%, $N=31$) or posttests only (25.0%, $N=17$) in over half of the included studies, with fewer studies measuring generalization intermittently throughout intervention (23.5%, $N=16$) or experimentally (i.e., at least three times per condition; 5.9%, $N=4$), reducing confidence in the internal validity of the reported effects.

Maintenance Measurement

Figure 5 depicts the results of the maintenance measurement for the 68 relations analyzed. Over three quarters of the included studies did not measure maintenance outcomes (79.4%, $N=54$). Of the few studies that did evaluate maintenance outcomes (20.6%, $N=14$), most measured maintenance at or beyond one month after intervention (71.4%, $N=10$), while two studies (14.3%) evaluated maintenance between one and three weeks after intervention (Gould et al., 2011; Zagrabska-Swiatkowska et al., 2020), and two studies (14.3%) measured maintenance immediately following

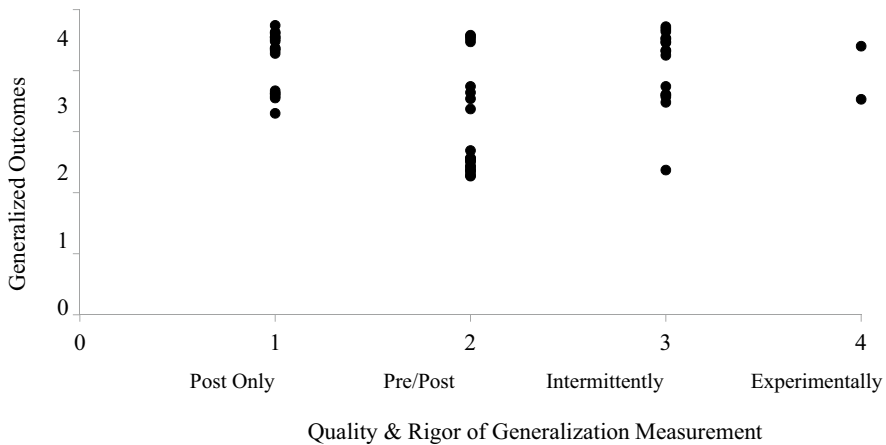


Fig. 4 Results for generalization measurement. As data points move down the x-axis, internal validity increases, while generalized outcomes are increasingly positive as data points move up 0 1 2 3 4 Primary Outcomes Overall Study Quality & Rigor 4 3 2 1 0 0 1 2 3 4 Quality & Rigor of Generalization Measurement Generalized Outcomes Post Only Pre/Post Intermittently Experimentally 4 3 2 1 0 the y-axis. While many of the studies included for review had positive generalized outcomes, as indicated by the clustering of data points at the top of the graph, most data points being plotted on the left-hand side of the graph indicates that internal validity was relatively weak for many of the included studies

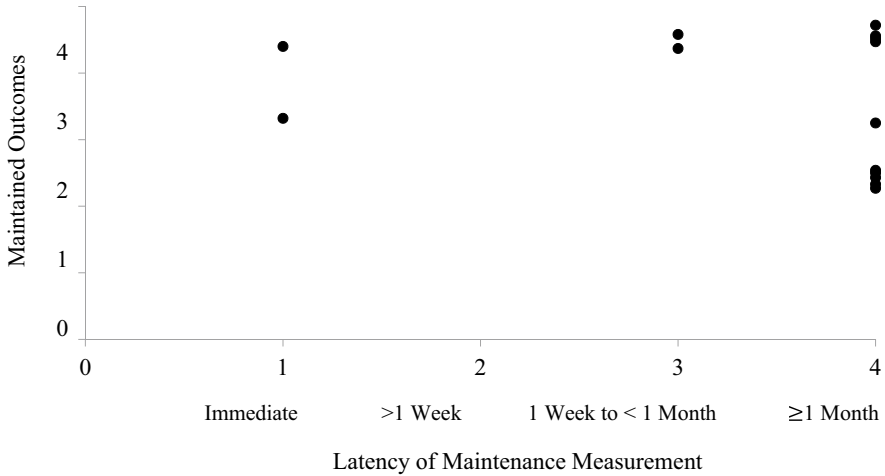


Fig. 5 Results for maintenance measurement. As data points move down the x-axis, the time between intervention and maintenance measurement increases, while maintained outcomes are increasingly positive as data points move up the y-axis. Maintained outcomes were generally positive, which is evidenced by the clustering of data points at the top of the graph. However, there are fewer data points represented within the graph as few of the included studies measured participant maintenance

intervention (Baltruschat et al., 2012; Persicke et al., 2012). Although each of the studies that measured maintenance reported moderate to positive outcomes immediately following intervention and over one month after intervention, because so many did not measure maintenance at all, limited conclusions can be made as to the retention of derived responding in the content areas taught.

Discussion

The current review sought to synthesize and analyze the results of research facilitating the emergence of derived relational responding beyond the frame of coordination. Systematic searches identified 30 articles comprised of 38 studies that met criteria for inclusion, supporting the findings of recent citation analyses by O'Connor et al. (2017) and Belisle et al. (2020a, b) that investigations into the use of technologies borne from the theoretical foundation of RFT with learners with IDD and ASD have increased in frequency over the last 15 years. The present review was focused on four key questions. The first was whether sufficient, high quality evidence suggests that learners with IDD and ASD can demonstrate derived relational responding beyond coordination. The second was whether evidence supported the presence of distinct learner characteristics or profiles that influence the development of derived relational responding beyond coordination. The third was whether there were particular assessment instruments ideal for determining a learner's relational skill repertoire. The final question was whether there are specific instructional procedures best suited for developing relational responding in learners with IDD and ASD.

Derived Relational Responding in Individuals with IDD and ASD

A total of 38 studies containing 122 participants were identified as having investigated the emergence of derived relational responding in accordance with a variety of relational frames beyond coordination. The analysis of the results of these studies found that over half of the participants (54.1%, $N=66$) demonstrated the emergence of derived relational responding in a variety of relational frames using several different teaching procedures (e.g., MET). An additional 38 participants (31.1%) demonstrated derived relational responding following additional intervention (e.g., additional prompting, remedial training, reverse intraverbal training, etc.). These findings support the assertion that learners with IDD and ASD can acquire the skills to demonstrate more complex derived relational responding beyond coordination relations. However, due to the quality and rigor of many of the included studies, these results must be interpreted with caution.

Learner Characteristics

The results of the current review support and extend the conclusions of previous work evaluating coordination relations that a learner's verbal repertoire influences their ability to demonstrate derived relational responding (Gibbs & Tullis, 2021; O'Connor et al., 2009; Lee et al., 2015). Two studies asked specific questions in an effort to determine how an individual's verbal repertoire influenced derived relational responding beyond coordination. Dunne et al. (2014) evaluated the effects of a program of testing and training on the emergence of derived relational responding in the frames of opposition (study two), distinction (study three) and comparison (study four) with learners with ASD. Across all three studies, the learners with higher VB-MAPP scores required considerably less training than learners with lower scores. In another set of studies, Kent et al. (2017) reported similar outcomes. Seven of 11 learners with ASD whose PPVT and K-BIT scores indicated more significant receptive and expressive language limitations were unable to progress through the entire test protocol of nonarbitrary and arbitrary coordination, distinction, comparison, and opposition relations even after extensive training trials (Kent et al., 2017, study two). In contrast, four learners with ASD who demonstrated receptive and expressive language delays, but whose PPVT and K-BIT scores indicated less significant limitations, were able to complete the entire test protocol following no more than 44 training trials (Kent et al., 2017, study three). These results bolster the hypothesis that a more expansive verbal repertoire influences the training requirements and test performances for derived relational responding across relational frames, but additional research is still required for more conclusive answers.

Furthermore, it has previously been suggested that bidirectional naming (BiN; Miguel, 2016) may be an essential repertoire for learners to demonstrate emergent equivalence relations (Howarth et al., 2015; Kobari-Wright & Miguel, 2014; Morgan et al., 2020), but the necessity of a BiN repertoire for the emergence of derived relations beyond coordination is less clear. Lovett and Rehfeldt (2014) taught perspective-taking skills to adolescents with ASD and assessed generalization of these skills to a more natural language

situation. For two of the three participants whose performance on reversed and double reversed deictic relations during generalization probes did not meet mastery criteria, the tact for each emotion experienced in the natural language situation was stated vocally following the video presentation. This resulted in an increase in response accuracy to mastery criterion levels for reversed relations, and improved response accuracy for double reversed relations, though still below mastery criterion. Lovett and Rehfeldt's (2014) results suggest generalization of perspective-taking skills may require not only a deictic relational responding repertoire, but also a tact repertoire. While certainly interesting, these findings are by no means conclusive, and there is currently little research investigating the relationship between BiN and different relational frames. Additional inquiry is required to better clarify the influence BiN has on the emergence of derived responding beyond coordination.

Another potential requirement for the acquisition of derived relational responding is the ability to first demonstrate nonarbitrary relational responding. As discussed previously, nonarbitrary relational responding is based on the physical properties of the target stimuli (Hayes et al., 2001), and it has been suggested that a repertoire of nonarbitrary relational responding may be a necessary prerequisite skill to demonstrate, or at least facilitate, derived relational responding (Barnes-Holmes et al., 2004; Berens & Hayes, 2007; Dunne et al., 2014; Kent et al., 2017). This concept was supported by Gale and Stewart (2020), whose three participants with ASD all demonstrated derived relational responding in accordance with comparison, and all had nonarbitrary relational responding in their repertoire. However, this suggestion has recently been brought into question. While not included in the current review secondary to only evaluating an assessment protocol, the findings of Pomorska et al. (2020) suggest the relationship between nonarbitrary and derived relational responding may be more dynamic than linear, with both skills potentially emerging in tandem. More research is necessary to better determine how nonarbitrary relational responding potentially influences derived relational responding.

Though a well-developed verbal repertoire inclusive of BiN and the ability to engage in nonarbitrary relational responding may be influential in facilitating derived relational responding across relational frames, the possibility that distinct precursor skills are required for the emergence of derived responding in each frame individually cannot be ruled out. For example, Molina-Cobos and Amador-Castro (2010) reported participants with Down syndrome and developmental delay were unable to successfully report on the preferences of another until discrimination training was utilized to teach participants to discriminate between themselves and the character being used in the study. As no assessment results were reported, the participants' overall discrimination skills are unknown, which precludes suggestions related to the required strength of a discrimination repertoire for deictic responding. While the previously described studies have contributed to our understanding of what skill repertoires may be necessary to support derived relational responding in this population of learners, few studies have specifically sought to answer this fundamental question. Future research should continue to investigate whether broad skill repertoires and/or specific precursor skills, and what ones, are required to facilitate derived relational responding within and across relational frames.

Assessment Tools

Similar to the findings of Gibbs and Tullis (2021), many of the studies in the current review reported the results of a variety of norm- (e.g., WISC-IV) and criterion-referenced (e.g., PIRK) assessments in an effort to describe participants cognitive and language skills. However, the use of specific assessment tools in an effort to determine learner's relational skill repertoires occurred less frequently. The preassessments associated with the Promoting the Emergence of Advanced Knowledge: Equivalence (PEAK-E; Dixon, 2015) and Transformation (PEAK-T; Dixon, 2016) modules were used in four studies. Positive results were reported when participants demonstrated mutual and combinatorial entailment on the PEAK-E (Belisle et al., 2020a, b) and PEAK-T (Belisle et al., 2016) preassessments. Barron et al. (2019) also reported positive outcomes following the ability to demonstrate nonarbitrary deictic relations on the PEAK-T receptive and expressive preassessments.

Two other assessments were used in an attempt to ascertain learner relational abilities. Cassidy et al. (2011) devised and administered the Relational Abilities Index (RAI), a computer-based assessment that measures proficiency in coordination, opposition, and comparison relational responding. While the RAI has been cited as an acceptable surrogate measure of IQ in the RFT literature (Colbert et al., 2017), in Cassidy et al. (2011) its purpose was to evaluate participants relational skill repertoires before and after intervention to determine if MET was responsible for changes in relational abilities. Since the development and extension (Cassidy et al., 2016) of the RAI, further work has been done to create a more expansive assessment of relational responding. The Relational Abilities Index + (RAI+; Colbert et al., 2020) assesses relational performance across distinction, temporality, and analogy in addition to coordination, opposition, and comparison. However, as no studies have yet been published on the use of the RAI+ with learners with IDD and ASD, more research is needed to determine its utility in assessing their relational repertoires.

Gale and Stewart (2020) utilized the TARPA (Moran et al., 2010, 2014), a hierarchical testing and training protocol based on simple and conditional discrimination skills in which the first stage tests an individual's ability to learn simple discriminations, the second stage tests an individual's ability to engage in nonarbitrary relational responding, and the final stage tests an individual's ability to engage in derived relational responding. Gale and Stewart (2020) first used the TARPA as a testing method to assess participants' abilities to learn the repertoires necessary to derive comparative relations, then as a training procedure to establish comparative relational responding, which was successful with all three participants with ASD. Given the TARPA's ability to assess and train prerequisite and relational responding skills, further research is warranted on its utility in clinical practice with learners with ASD as well as learners with IDD.

Teaching Procedures

Multiple exemplar training (MET) remains the most frequently implemented teaching procedure and it was utilized across relational frames and content, however several studies utilized multiple exemplar instruction (MEI). While sometimes used interchangeably,

MET and MEI are considered two different procedures, with MET often used in RFT approaches to derived responding, and MEI utilized in investigations grounded in BiN (LaFrance & Tarbox, 2020). Briefly, MET presents different exemplars while teaching a target response topography that serves the same function, while MEI rotates instructions targeting different responses to produce interdependence between speaker and listener repertoires (see LaFrance & Tarbox, 2020 for a more detailed discussion of each methodology). In the current review, both MET and MEI resulted in the acquisition of derived relational responding in various content areas, but participants sometimes required additional instruction in both procedures. For example, one participant in Belisle et al.'s (2016) investigation of teaching single reversal I-you deictic frames to learners with ASD via MET required additional mixed deictic frame training after failure to demonstrate derived relations after mastery of the initial training content. Similarly, in Greer and Yuan's (2008) evaluation of MEI to teach past tense verbs in an autoclitic function to learners with developmental delays, some participants demonstrated derived responding following MEI with one set of stimuli, and others needed intervention with an additional set. Future research should seek to determine if there is an optimal number of exemplars, or level of explicit instruction, required to facilitate derived responding with either of these two procedures.

Other teaching procedures used less frequently in the current review, but with positive results on derived relational responding, include a tabletop version of the TARP (Gale & Stewart, 2020), intraverbal training (Lee et al., 2019), match-to-sample and conditional discrimination training (Murphy & Barnes-Holmes, 2010), and a combination intervention of observation and discrimination training (Molina-Cobos & Amador-Castro, 2010, experiment two). The use of term 'observation' rather than 'observational learning' to describe this final intervention procedure is deliberate, as observational learning is characterized by observation of a modeled response and a subsequent consequence (Masia & Chase, 1997). In both experiments, Molina-Cobos and Amador-Castro (2010) had participants observe a photo of a named character engaging in an activity he was reported to prefer, but there was no subsequent consequence observed. Thus, the procedure does not fully meet the common definition of observational learning. While observation alone in experiment one was not successful for participants to learn the preferences of another, as previously discussed, all participants demonstrated the target response after observation was paired with discrimination training so participants could effectively discriminate between themselves and the character (Molina-Cobos & Amador-Castro, 2010). Given the relative simplicity of the interventions, additional research on the effects of a combination of observation and discrimination training on the emergence of derived relational responding is certainly warranted.

Design Rigor

The rigor of the studies included in the current review was measured using the SCARF (Ledford et al., 2020), and as reported in the results, while 67% of the studies reported positive outcomes for participants, only four of those with positive outcomes had high quality and rigor. Of the three SCARF elements associated with design rigor, sufficiency

of data and implementation fidelity were the two biggest contributors to the classification of many studies as low rigor, which is similar to the findings of Gibbs and Tullis (2021). Sufficiency of data refers to the presence of an adequate number of demonstrations of effect, and an adequate number of data points in each condition to allow for analysis of level, trend, and variability within and across conditions. The standard for sufficiency of data set by the SCARF requires at least three data points per condition, and at least three potential demonstrations of effect (Ledford et al., 2020), a standard that only 16.7% of the included studies met. The majority of the included studies failed to meet this standard due to having fewer than three data points per condition, which occurred most frequently in baseline (e.g., Jackson et al., 2014), and some studies also had fewer than three potential demonstrations of effect (Belisle et al., 2020a, b, experiments one and two; Grannan & Rehfeldt, 2012). A limitation of the current paper is that, as some studies were not amenable for evaluation of quality using the SCARF, it is unknown whether or not these articles may have strengthened or weakened the rigor of the overall body of work in this area, or had a significant degree of influence on the findings as a whole.

While certain quality indicator checklists put forth less rigorous standards for data sufficiency (e.g., Horner et al., 2005; Tate et al., 2008), in addition to the SCARF, quality standards set forth by the Council for Exceptional Children (CEC, 2014), the What Works Clearinghouse (WWC, 2020), and others (Reichow et al., 2008; Smith, 2012) require a minimum of three data points in each experimental condition. Valid arguments do exist for conditions in which fewer than three data points are appropriate to determine a pattern of responding (e.g., severe problem behavior, when an extended baseline would withhold necessary education from a learner, etc.). However, this did not appear to be the case in the included studies. Many of the studies that did not meet criteria utilized multiple probe designs or they did not use a discernable experimental design at all, but rather a series of testing and training phases. Two potential arguments could be made specifically to DRR literature. First, because relational pretesting can require an extensive number of trials (e.g., a 36-trial deictic framing protocol used for pre and posttest probes; Lovett & Rehfeldt, 2014), three data points in baseline might require too much time. Second, when arbitrary stimuli are used within a stimulus class and the established stimulus relations are determined by the experimenter, it is incredibly unlikely for a learner to have a history of reinforcement for responding to such relations. However, this point must be balanced with the need to establish via repeated observations that a learner truly lacks a derived relational responding repertoire to improve the believability of the intervention methods, especially given the evidence that learners with IDD and ASD may present with splinter skills, and engage in responding beyond what their developmental level or IQ score suggests they are capable of (Mayes et al., 2012).

In addition to sufficient data to support the presence of a functional relation, a high level of implementation fidelity, or the degree to which the independent variable is implemented as intended, is considered necessary by the SCARF as well as several other quality assessment tools (CEC, 2014; Horner et al., 2005; Tate et al., 2016) for confidence that changes in participant behavior are the result of the intervention being applied. Only 28.9% of the included studies reported procedural fidelity data. This is similar to the results of Gibbs and Tullis (2021), though the percentage is slightly improved from their reported 22.6%. While very few studies

measured maintenance, many of those that did reported maintained effects over one month after intervention. However, given that 78.8% of the included studies reported no maintenance at all, whether derived relational responding is a skill that maintains over time is an area requiring additional investigation.

Areas of Future Research

As previously reported, data on the race and ethnicity of research participants was not included in any of the 30 studies within this review, an unsurprising finding given the results of a systematic review by Steinbrenner et al. (2022) which found only 25% of articles published since 1990 examining evidence based practices in autism intervention reported on participant race and ethnicity. The absence of this data raises the question of whether interventions designed to facilitate derived relational responding have similar efficacy across racial and ethnic groups. Similarly, of the 30 studies, Lee et al. (2019) was the only one to clearly state participants' native language, despite several other studies being conducted internationally and in non-English speaking countries. In order to function as communication between speaker and listener, language requires context which includes cultural background knowledge, among other factors (Garten et al., 2019). This may be especially true when teaching complex relational responding such as comprehending metaphors or idioms, as the verbally mediated meaning in one language may not transfer to another (e.g., in German when one says, "Ich habe einen Kater," it means, "I have a hangover," and not the literal translation, "I have a tomcat"). Future studies should include this information explicitly to better determine whether race, ethnicity, and native language of participants influences intervention efficacy, and to aid in the development of more nuanced research questions regarding complex language generativity.

Although only one study examining the effect of improved derived relational responding has on IQ scores in learners with IDD was included in the current review (Cassidy et al., 2011), there has been increased interest in this area of research. Studies have reported significant IQ increases following an intervention to derive coordination, opposition, and comparison relations in typically developing learners (Cassidy et al., 2016; Colbert et al., 2018). An initial study by Dixon et al. (2021) indicates that similar increases in IQ score might be achieved with learners with IDD and/or ASD, but questions remain whether an increase in IQ score correlates with an increase in adaptive functioning following this type of intervention. Although general intelligence is by no means unimportant, adaptive functioning (e.g., daily living skills, interpersonal skills) may arguably be of more concern for learners with IDD and ASD, as it encompasses skills needed for everyday functioning (McQuaid et al., 2021). Previous research has demonstrated that among learners with IDD and ASD, a lower IQ score is correlated with more deficits in adaptive skills (Kanne et al., 2011; Matson & Shoemaker, 2009). However, research has also indicated that adaptive functioning is much lower than what would be expected based on IQ score in learners with ASD without a co-occurring intellectual disability (Bradshaw et al., 2019; Klin et al., 2007). Given that interventions to facilitate derived relational responding have the potential advantage of instructional efficiency, it is key to

determine whether or not the positive effect of these interventions extends to additional populations and beyond IQ score.

Similar to Gibbs and Tullis (2021), a limitation of the literature contained in the current review is that they lack social validity measures. This limitation is one that can and should be remedied with additional research. While 89.5% of the included studies took place in participants' natural environments (e.g., school, home, etc.), Lee et al. (2019) was the only study to directly measure social validity. Participants' parents filled out a questionnaire, and investigators interviewed the participants and their teachers (who were not aware of the research) to determine whether the intervention in metaphor comprehension facilitated improvements in communication, social, or reading comprehension skills in the classroom environment. As derived relational responding may be a potential avenue to more generative language, it behooves researchers to include generative measures outside of the experimental context with others who are blind to the goal of an investigation. Further, derived relational responding offers a more efficient treatment option, with intervention in only a few relations resulting in the derivation of other relations without explicit teaching. However, most social validity measures focus on stakeholder satisfaction, and few consider issues such as the time and cost-effectiveness associated with an intervention (Callahan et al., 2017), two factors that may influence more widespread adoption of these technologies. Future research should not only include objective measures of social validity to determine stakeholder satisfaction with interventions to facilitate derived relational responding, but also the instructional and resource efficiency of these interventions.

Practical Implications

The result of the current review that learners with IDD and ASD demonstrated derived relational responding beyond the frame of coordination has pronounced implications for skill acquisition and education, particularly language generativity, in this population. The evidence suggests a more expansive verbal repertoire, inclusive of BiN, and the ability to engage in nonarbitrary derived relational responding may influence this population's ability to demonstrate derived relational responding. This may assist practitioners in determining those learners who have the necessary prerequisite skills to begin this type of intervention. Additionally, the identification of assessments such as the PEAK-E and T assessments, RAI, and TARPA to determine current learner repertoires and in some cases guide intervention (i.e., PEAK modules, TARPA) is important. Finally, while MET was the most frequently utilized teaching procedure in the included studies, previous research has found the use of technology aided instruction such as the Teaching Implicit Relational Assessment Procedure (T-IRAP; Kilroe et al., 2014) may be more instructionally efficient than traditional tabletop presentations in learners with ASD. Technology-based instructional arrangements certainly have appeal, especially when viewed in light of the COVID-19 pandemic which resulted in the loss of face-to-face instruction for many students with and without disabilities around the world. However, regardless of whether instruction is taking place in person, finding the most efficient way to teach, and to reach the greatest possible number of learners certainly merits continued attention and research.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no relevant financial or non-financial interests to disclose.

Ethical Approval and Informed Consent The current project is a literature review that did not involve human subjects, and no institutional review board or informed consent was required.

References

- Baltruschat, L., Hasselhorn, M., Tarbox, J., Dixon, D. R., Najdowski, A., Mullins, R. D., & Gould, E. (2012). The effects of multiple exemplar training on a working memory task involving sequential responding in children with autism. *The Psychological Record*, *62*, 549–562. <https://doi.org/10.1007/BF03395820>
- Barnes-Holmes, Y., Barnes-Holmes, D., & McEntegart, C. (2018). Relational frame theory: Description, evidence, and clinical applications. In P. Lucena-Santos, S. Carvalho, J. Pinto-Gouveia, M. Silva Oliveira, & J. Pistorello (Eds.), *International ACT practical handbook*. TBC Press.
- Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. *International Journal of Psychology & Psychological Therapy*, *4*, 531–558.
- Barron, B. F., Verkuylen, L., Belisle, J., Paliliunas, D., & Dixon, M. R. (2019). Teaching “Then Later” and “Here-There” relations to children with autism: An evaluation of single reversals and transformation of stimulus function. *Behavior Analysis in Practice*, *12*, 167–175. <https://doi.org/10.1007/s40617-018-0216-1>
- Belisle, J., Dixon, M. R., Stanley, C. R., Munoz, B., & Daar, J. H. (2016). Teaching foundational perspective-taking skills to children with autism using the PEAK-T curriculum: Single reversal “I-You” deictic frames. *Journal of Applied Behavior Analysis*, *49*, 965–969. <https://doi.org/10.1002/jaba.324>
- Belisle, J., Paliliunas, D., Lauer, T., Giamanco, A., Lee, B., & Sickman, E. (2020a). Derived relational responding and transformations of function in children: A review of applied behavior-analytic journals. *The Analysis of Verbal Behavior*, *36*, 115–145. <https://doi.org/10.1007/s40616-019-00123-z>
- Belisle, J., Stanley, C. R., Schmick, A., Dixon, M. R., Alholail, A., Galliford, M. E., & Ellenberger, L. (2020b). Establishing arbitrary comparative relations and referential transformations of stimulus function in individuals with autism. *Journal of Applied Behavior Analysis*, *53*, 938–955. <https://doi.org/10.1002/jaba.655>
- Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for a relational operant. *Journal of Applied Behavior Analysis*, *40*, 45–71. <https://doi.org/10.1901/jaba.2007.7-06>
- Bradshaw, J., Gillespie, S., Klaiman, C., Klin, A., & Saulnier, C. (2019). Early emergence of discrepancy in adaptive behavior and cognitive skills in toddlers with autism spectrum disorder. *Autism*, *23*, 1485–1496. <https://doi.org/10.1177/1362361318815662>
- Callahan, K., Hughes, H. L., Mehta, S., Toussaint, K. A., Nichols, S. M., Ma, P. S., Kutlu, M., & Wang, H. T. (2017). Social validity of evidence-based practices and emerging interventions in autism. *Focus on Autism and Other Developmental Disabilities*, *32*, 188–197. <https://doi.org/10.1177/1088357616632446>
- 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. <https://doi.org/10.1007/BF03395755>
- 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>
- Colbert, D., Dobutowitsch, M., Roche, B., & Brophy, C. (2017). The proxy-measurement of intelligence quotients using a relational skills ability index. *Learning and Individual Differences*, *57*, 114–122. <https://doi.org/10.1016/j.lindif.2017.03.010>
- Colbert, D., Malone, A., Barrett, S., & Roche, B. (2020). The relational abilities index+: Initial validation of a functionally understood proxy measure for intelligence. *Perspectives on Behavior Science*, *43*, 189–213. <https://doi.org/10.1007/s40614-019-00197-z>

- Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART training really increase intelligence? A replication study. *Journal of Behavioral Education, 27*, 509–531. <https://doi.org/10.1007/s10864-018-9302-2>
- Council for Exceptional Children (CEC). (2014). Council for exceptional children: Standard of evidence-based practices in special education. *Teaching Exceptional Children, 46*, 206–212. <https://doi.org/10.1177/0040059914531389>
- Courtade, G. R., Test, D. W., & Cook, B. G. (2015). Evidence-based practices for learners with severe intellectual disability. *Research and Practice for Persons with Severe Disabilities, 39*, 305–318. <https://doi.org/10.1177/2F1540796914566711>
- Dixon, M. R. (2015). *The PEAK relational training system: Equivalence module (PEAK-E)*. Shawnee Scientific Press.
- Dixon, M. R. (2016). *The PEAK relational training system: Transformation module (PEAK-T)*. Shawnee Scientific Press.
- Dixon, M. R., Pailiunas, D., Barron, B. F., Schmick, A. M., & Stanley, C. R. (2021). Randomized controlled trial evaluation of ABA content on IQ gains in children with autism. *Journal of Behavioral Education, 30*, 455–477. <https://doi.org/10.1007/s10864-019-09344-7>
- Dunne, S., Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., & Murphy, C. (2014). Facilitating repertoires of coordination, opposition, distinction, and comparison in young children with autism. *Behavioral Development Bulletin, 19*, 37–47. <https://doi.org/10.1037/h0100576>
- Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus function. *The Behavior Analyst, 23*, 239–254. <https://doi.org/10.1007/BF03392013>
- Gale, L., & Stewart, I. (2020). Assessing and training comparative relations in children with autism spectrum disorder. *Journal of European Psychology Students, 11*, 1–14. <https://doi.org/10.5334/jeps.487>
- Garten, J., Kennedy, B., Sagae, K., & Dehghani, M. (2019). Measuring the importance of context when modeling language comprehension. *Behavior Research Methods, 51*, 480–492. <https://doi.org/10.3758/s13428-019-01200-w>
- Gibbs, A. R., & Tullis, C. A. (2021). The emergence of untrained relations in individuals with autism and other intellectual and developmental disabilities: A systematic review of the recent literature. *Review Journal of Autism and Developmental Disorders, 8*, 213–238. <https://doi.org/10.1007/s40489-020-00211-0>
- Gilroy, S. P., Lorah, E. R., Dodge, J., & Fiorello, C. (2015). Establishing deictic repertoires in autism. *Research in Autism Spectrum Disorders, 19*, 82–92. <https://doi.org/10.1016/j.rasd.2015.04.004>
- Gorham, M., Barnes-Holmes, Y., & Barnes-Holmes, D. (2009). Derived comparative and transitive relations in young children with and without autism. *The Psychological Record, 59*, 221–246. <https://doi.org/10.1007/BF03395660>
- Gould, E., Tarbox, J., O'Hora, D., Noone, S., & Bergstrom, R. (2011). Teaching children with autism a basic component skill of perspective taking. *Behavioral Interventions, 26*, 50–66. <https://doi.org/10.1002/bin.320>
- Grannan, L., & Rehfeldt, R. A. (2012). Emergent intraverbal responses via tact and match-to sample instruction. *Journal of Applied Behavior Analysis, 45*, 601–605. <https://doi.org/10.1901/jaba.2012.45-601>
- Greer, R. D., & Yuan, L. (2008). How kids learn to say the darndest things: The effect of multiple exemplar instruction on the emergence of novel verb usage. *The Analysis of Verbal Behavior, 24*, 103–121. <https://doi.org/10.1007/BF03393060>
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post Skinnerian account of human language and cognition*. Plenum Press.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 185–241. <https://doi.org/10.1901/jeab.1996.65-185>
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179. <https://doi.org/10.1177/001440290507100203>
- Howarth, M., Dudek, J., & Greer, R. D. (2015). Establishing derived relations for stimulus equivalence in children with severe cognitive and language delays. *European Journal of Behavior Analysis, 16*, 49–81. <https://doi.org/10.1080/15021149.2015.1065635>
- Hume, K., Steinbrenner, J. R., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yücesoy-Özkan, S., & Savage, M. N. (2021). Evidence based practices for children, youth, and young adults with autism: Third generation review. *Journal of Autism and Developmental Disorders, 51*, 4013–4032. <https://doi.org/10.1007/s10803-020-04844-2>

- Jackson, M. L., Mendoza, D. R., & Adams, A. N. (2014). Teaching a deictic relational repertoire to children with autism. *The Psychological Record*, *64*, 791–802. <https://doi.org/10.1007/s40732-014-0078-z>
- Kanne, S. M., Gerber, A. J., Quirnbach, L. M., Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2011). The role of adaptive behavior in autism spectrum disorders: Implications for functional outcome. *Journal of Autism and Developmental Disorders*, *41*, 1007–1018. <https://doi.org/10.1007/s10803-010-1126-4>
- Kavanagh, D., Barnes-Holmes, Y., & Barnes-Holmes, D. (2020). The study of perspective taking: Contributions from mainstream psychology and behavior analysis. *The Psychological Record*, *70*, 581–604. <https://doi.org/10.1007/s40732-019-00356-3>
- Kent, G., Galvin, E., Murphy, C., Barnes-Holmes, Y., & Barnes-Holmes, D. (2017). Relational responding: Testing, training, and sequencing effects among children with autism and typically developing children. *Behavioral Development Bulletin*, *22*, 94–110. <https://doi.org/10.1037/bdb0000041>
- Kilroe, H., Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2014). Using the T-IRAP interactive computer program and applied behavior analysis to teach relational responding in children with autism. *Behavioral Development Bulletin*, *19*, 60–80. <https://doi.org/10.1037/h0100578>
- Klin, A., Saulnier, C. A., Sparrow, S. S., Cicchetti, D. V., Volkmar, F. R., & Lord, C. (2007). Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: The Vineland and the ADOS. *Journal of Autism and Developmental Disorders*, *37*, 748–759. <https://doi.org/10.1007/s10803-006-0229-4>
- Kobari-Wright, V. V., & Miguel, C. F. (2014). The effects of listener training on the emergence of categorization and speaker behavior in children with autism. *Journal of Applied Behavior Analysis*, *47*, 431–436. <https://doi.org/10.1002/jaba.115>
- LaFrance, D. L., & Tarbox, J. (2020). The importance of multiple exemplar instruction in the establishment of novel verbal behavior. *Journal of Applied Behavior Analysis*, *53*, 10–24. <https://doi.org/10.1002/jaba.611>
- Ledford, J. R., Chazin, K. T., Lane, J. D., Zimmerman, K. N., & Ayres, K. A. (2020, September). Single case analysis and review framework (SCARF). Retrieved from: <https://ebip.vkcsites.org/scarfv2/>
- Lee, G. P., Miguel, C. F., Darcey, E. K., & Jennings, A. M. (2015). A further evaluation of the effects of listener training on derived categorization and speaker behavior in children with autism. *Research in Autism Spectrum Disorders*, *19*, 72–81. <https://doi.org/10.1016/j.rasd.2015.04.007>
- Lee, G. T., Xu, S., Zou, H., Gilic, L., & Lee, M. W. (2019). Teaching children with autism to understand metaphors. *The Psychological Record*, *69*, 499–512. <https://doi.org/10.1007/s40732-019-00355-4>
- Lovett, S., & Rehfeldt, R. A. (2014). An evaluation of multiple exemplar instruction to teach perspective-taking skills to adolescents with Asperger Syndrome. *Behavioral Development Bulletin*, *19*, 22–36. <https://doi.org/10.1037/h0100575>
- Luke, N., Greer, R. D., Singer-Dudek, J., & Keohane, D. (2011). The emergence of autoclitic frames in atypically and typically developing children as a function of multiple exemplar instruction. *The Analysis of Verbal Behavior*, *27*, 141–156. <https://doi.org/10.1007/BF03393098>
- Masia, C. L., & Chase, P. N. (1997). Vicarious learning revisited: A contemporary behavior analytic interpretation. *Journal of Behavior Therapy and Experimental Psychiatry*, *28*(1), 41–51. [https://doi.org/10.1016/s0005-7916\(96\)00042-0](https://doi.org/10.1016/s0005-7916(96)00042-0)
- Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities*, *30*, 1107–1114. <https://doi.org/10.1016/j.ridd.2009.06.003>
- Mayes, S. D., Calhoun, S. L., Murray, M. J., & Molitoris, S. (2012). Autism and ADHD: Overlapping and discriminating symptoms. *Research in Autism Spectrum Disorders*, *6*, 277–285. <https://doi.org/10.1016/j.rasd.2011.05.009>
- McHugh, L., & Reed, P. (2008). Using relational frame theory to build grammar in children with autistic spectrum conditions. *The Journal of Speech and Language Pathology – Applied Behavior Analysis*, *3*, 60–77. <https://doi.org/10.1037/h0100233>
- McQuaid, G. A., Pelphrey, K. A., Bookheimer, S. Y., Dapretto, M., Webb, S. J., Bernier, R. A., McPartland, J. C., Van Horn, J. D., & Wallace, G. L. (2021). The gap between IQ and adaptive functioning in autism spectrum disorder: Disentangling diagnostic and sex differences. *Autism*. <https://doi.org/10.1177/1362361321995620>
- Miguel, C. F. (2016). Common and intraverbal bidirectional naming. *The Analysis of Verbal Behavior*, *32*, 125–138. <https://doi.org/10.1007/s40616-016-0066-2>
- Ming, S., Moran, L., & Stewart, I. (2014). Derived relational responding and generative language: Applications and future directions for teaching individuals with autism spectrum disorder. *European Journal of Behavior Analysis*, *15*, 199–224. <https://doi.org/10.1080/15021149.2014.11434722>

- Ming, S., Mulhern, T., Stewart, I., Moran, L., & Bynum, K. (2018). Training class inclusion responding in typically developing children and individuals with autism. *Journal of Applied Behavior Analysis, 51*, 53–60. <https://doi.org/10.1002/jaba.429>
- Ming, S., & Stewart, I. (2017). When things are not the same: A review of research into relations of difference. *Journal of Applied Behavior Analysis, 50*, 429–455. <https://doi.org/10.1002/jaba.367>
- Molina-Cobos, F. J., & Amador- Castro, M. C. (2010). Theory of mind in young people with Down's syndrome. *International Journal of Psychology and Psychological Therapy, 10*, 363–385.
- Montoya-Rodríguez, M. M., McHugh, L., & Molina-Cobos, F. J. (2017a). Teaching perspective taking skills to an adult with Down syndrome: A case study. *Journal of Contextual Behavioral Science, 6*, 293–297. <https://doi.org/10.1016/j.jcbs.2017.04.012>
- Montoya-Rodríguez, M. M., Molina-Cobos, F. J., & McHugh, L. (2017b). A review of relational frame theory research into deictic relational responding. *The Psychological Record, 67*, 569–579. <https://doi.org/10.1007/s40732-016-0216-x>
- Montoya-Rodríguez, M. M., & Molina-Cobos, F. J. (2019). Training perspective taking skills in individuals with intellectual disabilities: A functional approach. *Journal of Contextual Behavioral Science, 14*, 1–10. <https://doi.org/10.1016/j.jcbs.2019.08.003>
- Moran, L., Stewart, I., McElwee, J., & Ming, S. (2010). The training and assessment of relational precursors and abilities (TARPA): A preliminary analysis. *Journal of Autism and Developmental Disorders, 40*, 1149–1153. <https://doi.org/10.1007/s10803-010-0968-0>
- Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational ability and language performance in children with autism spectrum disorders and typically developing children: A further test of the TARPA protocol. *The Psychological Record, 64*, 233–251. <https://doi.org/10.1007/s40732-014-0032-0>
- Morgan, G. A., Greer, R. D., & Fienup, D. M. (2020). *Descriptive analyses of relations among bidirectional naming, arbitrary, and nonarbitrary relations*. Advance online publication. <https://doi.org/10.1007/s40732-020-00408-z>
- Murphy, C., & Barnes-Holmes, D. (2009). Derived more-less relational mands in children diagnosed with autism. *Journal of Applied Behavior Analysis, 42*, 253–268. <https://doi.org/10.1901/jaba.2009.42-253>
- Murphy, C., & Barnes-Holmes, D. (2010). Establishing complex derived manding with children with and without a diagnosis of autism. *The Psychological Record, 60*, 489–504. <https://doi.org/10.1007/BF03395723>
- Newsome, K. B., Berens, K. N., Ghezzi, P. M., Aninao, T., & Newsome, W. D. (2014). Training relational language to improve reading comprehension. *European Journal of Behavior Analysis, 15*, 165–197. <https://doi.org/10.1080/15021149.2014.11434512>
- O'Connor, M., Farrell, L., Munnally, A., & McHugh, L. (2017). Citation analysis of relational frame theory: 2009–2016. *Journal of Contextual Behavioral Science, 6*, 152–158. <https://doi.org/10.1016/j.jcbs.2017.04.009>
- O'Connor, J., Rafferty, A., Barnes-Holmes, D., & Barnes-Holmes, Y. (2009). The role of verbal behavior, stimulus nameability, and familiarity on the equivalence performances of autistic and normally developing children. *The Psychological Record, 59*, 53–74. <https://doi.org/10.1007/BF03395649>
- Pérez-Gonzalez, L. A., García-Asenjo, L., Williams, G., & Carnerero, J. J. (2007). Emergence of intra-verbal antonyms in children with pervasive developmental disorder. *Journal of Applied Behavior Analysis, 40*, 697–701. <https://doi.org/10.1901/jaba.2007.697-701>
- Persicke, A., Tarbox, J., Ranick, J., & St. Clair, M. (2012). Establishing metaphorical reasoning in children with autism. *Research in Autism Spectrum Disorders, 6*, 913–920. <https://doi.org/10.1016/j.rasd.2011.12.007>
- Persicke, A., Tarbox, J., Ranick, J., & St. Clair, M. (2013). Teaching children with autism to detect and respond to sarcasm. *Research in Autism Spectrum Disorders, 7*, 193–198. <https://doi.org/10.1016/j.rasd.2012.08.005>
- Pomorska, K., Ostaszewski, P., Barnes-Holmes, Y., & McEntegart, C. (2020). Protocols for assessing derived relations in typically developing children and children with autism spectrum disorder. *The Psychological Record, 70*. <https://doi.org/10.1007/s40732-020-00442-x>
- Raaymakers, C., García, Y., Cunningham, K., Krank, L., & Nemer-Kaiser, L. (2019). A systematic review of derived verbal behavior research. *Journal of Contextual Behavioral Science, 12*, 128–148. <https://doi.org/10.1016/j.jcbs.2019.02.006>
- Ranick, J., Persicke, A., Tarbox, J., & Kornack, J. A. (2013). Teaching children with autism to detect and respond to deceptive statements. *Research in Autism Spectrum Disorders, 7*, 503–508. <https://doi.org/10.1016/j.rasd.2012.12.001>

- Rehfeldt, R. A., & Barnes-Holmes, Y. (2009). *Derived relational responding : Applications for learners with autism and other developmental disabilities*. New Harbinger Publications, Inc.
- Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method of evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*, *38*, 1311–1319. <https://doi.org/10.1007/s10803-007-0517-7>
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5–13. <https://doi.org/10.1044/jshr.1401.05>
- Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*, 510–550. <https://doi.org/10.1037/a0029312>
- Speckman, J., Greer, R. D., & Rivera-Valdes, C. (2012). Multiple exemplar instruction and the emergence of generative production of suffixes as autoclitic frames. *The Analysis of Verbal Behavior*, *28*, 83–99. <https://doi.org/10.1007/2FBBF03393109>
- Stauch, T., LaLonde, K., Plavnick, J. B., Bak, M. Y. S., & Gatewood, K. (2017). Intraverbal training for individuals with autism: The current status of multiple control. *The Analysis of Verbal Behavior*, *33*, 98–116. <https://doi.org/10.1007/2Fs40616-017-0079-5>
- Steinbrenner, J. R., McIntyre, N., Rentschler, L. F., Pearson, J. N., Luelmo, P., Jaramillo, M. E., Boyd, B. A., Wong, C., Nowell, S. W., Odom, S. L., & Hume, K. A. (2022). Patterns in reporting and participant inclusion related to race and ethnicity in autism intervention literature: Data from a large-scale systematic review of evidence-based practices. *Autism*, *26*, 2026–2040. <https://doi.org/10.1177/13623613211072593>
- 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. 51–72). Context Press.
- Tate, R. L., McDonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the single case experimental design (SCED) scale. *Neuropsychological Rehabilitation*, *18*, 385–401. <https://doi.org/10.1080/09602010802009201>
- Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Horner, R., Kazdin, A., Kratchowill, T., McDonald, S., Sampson, M., Shamseer, L., Togher, L., Albin, R., Backman, C., Douglas, J., Evans, J. J., Gast, D., Manolov, R., & ... Wilson, B. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology*, *4*, 1–9. <https://doi.org/10.1037/arc0000026>
- What Works Clearinghouse. (2020). *Procedures and standards handbook, version 4.1*. Retrieved from: <https://ies.ed.gov/ncee/wwc/Protocols#procedures>
- Zagrabska-Swiatkowska, P., Mukhern, T., Ming, S., Stewart, I., & McElwee, J. (2020). Training class inclusion responding in individuals with autism: Further investigation. *Journal of Applied Behavior Analysis*, *53*, 2067–2080. <https://doi.org/10.1002/jaba.712>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.