Title	Testing and Developing a Protocol for Training and Assessment of Relational Precursors and Abilities
Author(s)	Moran, Laura
Publication Date	2014-02-07
Item record	http://hdl.handle.net/10379/4173

Testing and Developing a Protocol for Training and Assessment of Relational Precursors and Abilities

Laura Moran

Dissertation submitted in partial fulfilment of the requirements for the
Degree of Doctor of Philosophy in Applied Behaviour Analysis
Supervisor: Dr. Ian Stewart
School of Psychology
National University of Ireland, Galway
August, 2013

Table of Contents

Chapter No.	Description	Pages
	Declaration and Statements………………………….........	iv
	Abstract…………………………………………………	v
	Acknowledgements………………………………….......	vi
	Publications and conference presentations ………..........	vii
	List of figures and tables…………………………….......	viii
	Abbreviations. ……………………………………..........	xi
	Glossary of terms……………………………………........	xii
Chapter 1	Literature Review	
	Applied Behaviour Analysis and Language Skills……...	2
	Derived Relational Responding………………………....	12
	Theories of Derived Stimulus Relations………………...	16
	Relational Frame Theory………………………….........	17
	Derived Relational Responding skills in persons with Developmental Delay…………………………………..	23
	Teaching Derived Coordinate Relations………………...	24
	Assessing the Development of Derived Relational Responding……………………………………………..	28
	Training and Assessment of Relational Precursors and Abilities (TARPA)……………………………………	31
	Summary and Research Aims…………………………	38
Chapter 2	Comparing Derived Relations and Language Ability in Children with Autism Spectrum Disorders & Typically Developing Children: An Initial Test of the TARPA Protocol	
	Introduction……………………………………………..	41
	Experiment 1……………………………………………	45
	Method …………………………………………………	45
	Results and Discussion…………………………………	58
	Experiment 2……………………………………………	65
	Method …………………………………………………	66
	Results and Discussion…………………………………	67
	General Discussion………………………………..........	82
Chapter 3	Refining the TARPA: An Examination of Two Procedural Changes to the TARPA Protocol	
	Experiment 3……………………………………………	89
	Method …………………………………………………	93
	Results………………………………………………......	99
	Discussion………………………………………………	108
Chapter 4	Further Research: Investigating Correlations with Cognitive Ability and Track Order. Reliability	
	Introduction……………………………………………......	115

	Experiment 4a…………………………………………..	122
	Method…………………………………………………..	123
	Results and Discussion………………………………….	124
	Experiment 4b…………………………………………..	130
	Method …………………………………………………	130
	Results and Discussion………………………………....	131
	Experiment 4c…………………………………….........	143
	Method.…………………………………………………	144
	Results and Discussion………………………………....	145
	General Discussion………………………………….…...	153
Chapter 5	Testing A Brief Version of the Adjusted TARPA with Children with ASD	
	Experiment 5…………………………………………...	161
	Method …………………………………………………	166
	Results and Discussion………………………………….	169
	General Discussion……………………………………...	185
Chapter 6	General Discussion	
	Overview of Research Area and Aims……………….….	191
	Summary of Thesis Findings…………………………....	192
	Theoretical Issues and Implications……………..........	197
	Limitations and Directions for Future Research………...	214
	Summary and Conclusion……………………………….	221
References		223
Appendices	Appendix A……………………………………………..	249
	Appendix B……………………………………………..	256
	Appendix C……………………………………………..	257
	Appendix D……………………………………………..	259
	Appendix E……………………………………………..	260
	Appendix F……………………………………………..	262

Declarations and Statements

This work has not been previously been accepted in substance for any degree, and is not being concurrently submitted in candidature for any degree.

Signed………………………………. (candidate)

Date………………………………….

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed………………………………. (candidate)

Date………………………………….

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed………………………………. (candidate)

Date………………………………

Abstract

This thesis involved testing and further developing a multi-level computer-based protocol for the assessment of core language skills in typically developing children and children with Autism Spectrum Disorder (ASD). The Training and Assessment of Relational Precursors and Abilities (TARPA; Moran, Stewart, McElwee & Ming, 2010) is based on a Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001) conception of language and cognition. RFT is situated within behaviour analysis and is based on the premise that learned contextually controlled patterns of derived relational responding underpin all language and cognitive abilities. The aim of the TARPA is to enable assessment and training of the earliest stages of the emergence of this ability. The purpose of this work was to validate the basic TARPA protocol and to further improve aspects of it over the course of several experiments focused on the assessment of derived relational ability in children of different ages and including both typically developing children and children with ASD for whom generative language is a particular problem.

Experiments 1 (n = 10) and 2 (n = 13) validated the TARPA as an assessment protocol in children with ASD and typically developing children respectively. These studies involved (i) examining quantitative correlations between the TARPA and the Preschool Language Scale-4 (PLS-4; Zimmerman, Steiner & Pond, 2002) and (ii) testing the proposed hierarchical sequence of stages and sub-stages using order analysis (Krus, Bart & Airasian, 1975). The results showed strong correlations between the TARPA and the PLS-4 and the order analysis confirmed that, for the most part, the structure of the TARPA was robust.

The third experiment involving typically developing children (n = 10) was similar to Studies 1 and 2 in that it correlated the TARPA with the PLS-4 but it also adapted the protocol to deal with certain technical issues that arose in Study 1. Study 4 (with 26 typically developing children) advanced this work further by (i) correlating the TARPA with additional measures of cognitive functioning (i.e., Stanford-Binet 5; Roid, 2003) and (ii) investigating the effect of the order of presentation of TARPA tracks (sections of the protocol that are based on different stimulus modalities including auditory, visual and audiovisual) on participants overall score. Results showed strong and significant correlations between the TARPA and both the PLS-4 and SB5 which were comparable with correlations between the two latter. A strong track order effect was also evident in that presentation of the auditory track first seemed to negatively affect subsequent performance.

Experiments 1- 4 provided detailed insights into the TARPA and the protocol evolved considerably based on findings from this regime of testing. Using the more refined version of the protocol that was developed in this way, a final study with children with ASD (n = 20) was conducted. This advanced previous work with this population by (i) using additional measures of functioning including the SB5, Vineland Adaptive Behavior Scales and the Gilliam Autism Severity scales and (ii) examining test-retest reliability. Results showed a strong and significant correlation between the TARPA and alternative measures of functioning. Other interesting patterns were also observed in that children performed substantially better on primarily visual tasks.

In summary, the work presented in this thesis is promising for the future development of the TARPA and gives an early indication that the protocol could eventually provide a comprehensive tool for the assessment and training of core linguistic and cognitive abilities for use with a number of different populations, including typically developing children and children with ASD or other forms of developmental delay. This no doubt would constitute a key addition to existing tools for intensive behavioral intervention.

Acknowledgements

My sincere gratitude to:

My supervisor – Dr. Ian Stewart. You have been an enormous support from the very beginning to the very end of this research. Your generosity with your time, in person and in answering *hundreds* of emails, is very much appreciated. I feel privileged to have had the opportunity to learn from such an excellent researcher.

Dr. Olive Healy and Dr. Jennifer Holloway - whose enthusiasm for applied behaviour analysis, and expert knowledge have been a real inspiration to me.

All of my friends - for being the best friends a person could ask for. A special mention to the great friends I have made along the way, your friendships have made this PhD all the more enjoyable!

My parents - Tom and Helen, thank you for everything you have done for me. Your good sense, generosity and constant support saw me through to the end of this thesis.

My brother Thomas and my sister Evelyn – for your encouragement, practical help and friendship.

Pat – for your patience and understanding, I could not have finished this thesis without you.

All of my participants and their families - for giving your time, without you all, this thesis would certainly not have been possible. A special thanks also to the schools that facilitated this research.

Publications and Conference Presentations Resulting from this Thesis

Publications

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2013). Relational Ability and Language Performance in Children with Autism Spectrum Disorders & Typically Developing Children: A Further Test of the TARPA Protocol. *The Psychological Record (In press)*.

Manuscripts in Preparation

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2013). Relational Ability and Language Performance in Children with Autism Spectrum Disorders: A Further Test of the TARPA Protocol.

Conference Presentations

Moran, L. & Stewart, I. (2013). *The development of comprehensive, user-friendly, computer- based protocol (Training & Assessment of Relational Precursors & Abilities; TARPA) for the assessment and training of relational framing in young children with ASD.* Paper presented at the Psychological Society of Ireland, Division of Behaviour Analysis; Annual Conference (NUIG). April, 2013.

Moran, L. & Stewart, I. (2012). *The development of comprehensive, user-friendly, computer- based protocol (Training & Assessment of Relational Precursors & Abilities; TARPA) for the assessment and training of relational framing in young children with ASD.* Paper presented at the Annual Congress of Psychological Society of Ireland, Cork, Ireland, November, 2012.

Moran, L. & Stewart, I. (2012). *The development of comprehensive, user-friendly, computer- based protocol (Training & Assessment of Relational Precursors & Abilities; TARPA) for the assessment and training of relational framing in young children with ASD.* Paper presented at the Association of Behavior Analysis, International. Seattle, May, 2012.

Moran, L. & Stewart, I. (2010). *The development of comprehensive, user-friendly, computer- based protocol (Training & Assessment of Relational Precursors & Abilities; TARPA) for the assessment and training of relational framing in young children with ASD.* Paper presented at the Annual Congress of Psychological Society of Ireland, Athlone, Ireland, November, 2010.

Moran, L., & Stewart, I. (2010). *The development of a comprehensive, user-friendly, computer-based protocol (Training & Assessment of Relational Precursors & Abilities; TARPA) for the assessment and training of relational framing in young children with ASD*. Paper presented at the Psychological Society of Ireland, Division of Behaviour Analysis; Annual Conference (NUIG).

List of Figures

Page 13	*Figure 1:* Sidman's (1971) equivalence training procedure
Page 51	*Figure 2:* Screenshots of the TARPA. (i) Title screen; (ii) Stage selection; (iii) Two choice discrimination with visual stimuli (Stage 1); (iv) Test of generalised identity matching with visual stimuli (Stage 2); (v) Conditional discrimination with auditory stimuli (Stage 2); (vi) Arbitrary conditional discrimination with auditory sample and visual comparisons (Stage 3); (vii) Example of reinforcement screen; (viii) On screen data output.
Page 95	*Figure 3:* Token board used in Experiment 3.
Page 99	*Figure 4:* Termination Criteria for Stage 3 Levels used in Experiments 1/2 and Experiment 3
Page 103	*Figure 5:* A scatter plot showing the relationship between PLS-4 scores and TARPA performance. Participant 5 is marked as a possible outlier.

List of Tables

Page 58 *Table 1:* Demographic Information and TARPA Performance Parameters for Experiment 1.

Page 60 *Table 2:* TARPA and PLS4 Scores for Experiment 1.

Page 62 *Table 3:* Performances of Individual Participants for each TARPA Stage and Level in Experiment 1.

Page 68 *Table 4:* Demographic Information and TARPA Performance Parameters for Experiment 2.

Page 69 *Table 5:* TARPA and PLS4 Scores for Experiment 2.

Page 71 *Table 6:* Performances of Individual Participants for Each TARPA Stage and Level in Experiment 2.

Page 75 *Table 7:* Order analysis of the TARPA including all stages and levels based on the data provided by the 23 participants across Experiments 1 and 2.

Page 81 *Table 8:* Order analysis of TARPA Stage 3 for each of the four tracks based on the data provided by the 23 participants across Experiments 1 and 2.

Page 100 *Table 9:* Expt. 3. Demographic Information and TARPA Performance Parameters.

Page 101 *Table 10:* Expt. 3. TARPA and PLS4 scores.

Page 105 *Table 11:* Performances of Individual Participants for Each TARPA Stage and Level in Experiment 3.

Page 108 *Table 12:* Expt. 3. Order analysis of the TARPA including all levels within Stage 3 based on the data provided by the 10 participants.

Page 126 *Table 13:* Expt 4 A. Older children. Participant sex, age and performance on the TARPA, PLS4 and SB-5.

Page 128 *Table 14:* Performances of Individual Participants for Each TARPA Stage and Level in Experiment 4a.

Page 132 *Table 15:* Expt 4B. Demographic Information and TARPA Performance Parameters.

Page 133 *Table 16:* Expt 4B. TARPA, PLS4 and SB-5 scores.

Page 135 *Table 17:* Expt. 4A & Expt 4B. Correlations between the TARPA and the PLS4 and SB-5.

Page 139 *Table 18:*Expt. 4A & Expt 4B.Correlations between the TARPA and the PLS4 and SB-5 versus correlations between the PLS4 and SB-5.

Page 140 *Table 19:* Performances of Individual Participants for Each TARPA Stage and Level in Experiment 4b.

Page 145 *Table 20:* Expt 4C. Demographic Information and TARPA Performance Parameters.

Page 146 *Table 21:* Expt 4C. TARPA, PLS4 and SB-5 scores.

Page 151 *Table 22:* Performances of Individual Participants for Each TARPA Stage and Level in Experiment 4C.

Page 169 *Table 23:* Demographic Information and TARPA Performance Parameters.

Page 171 *Table 24:* GARS, VABS (composite), TARPA, PLS4 and SB5 scores.

Page 173 *Table 25:* Correlations between the TARPA and alternative measures (e.g., TARPA & PLS4, etc.) versus correlations between the alternative measures (e.g., PLS4 & SB5, etc.).

Page 175 *Table 26:* Correlations between Trials to Criterion and certain Visual and Auditory portions of the TARPA (Stages' 1 and 2).

Page 178 *Table 27:* Performances of Individual Participants for Each TARPA Stage and Level.

Page 182 *Table 28:* Order analysis of TARPA Stages 1 and 2 for each of the levels in the Visual track based on the data provided by the 20 participants.

Page 183 *Table 29:* Order analysis of TARPA Stages 1 and 2 for each of the levels in the Auditory track based on the data provided by the 20 participants.

Page 184 *Table 30:* Order analysis of TARPA Stage 3 for each of the levels in the Auditory Visual 1 track based on the data provided by the 20 participants.

Abbreviations

ABLLS	Assessment of Basic Language and Learning Skills
ABLA	Assessment of Basic Learning Abilities
ASD	Autism Spectrum Disorder
CABAS	Comprehensive Application of Behavior Analysis to Schooling
CD	Conditional Discrimination
CE	Combinatorial Entailment
DRR	Derived relational responding
GARS	Gilliam Autism Rating Scales
MET	Multiple exemplar training
ME	Mutual Entailment
PLS-4	Preschool language scales – 4th edition
RFT	Relational Frame Theory
SB5	Stanford Binet – 5th edition
TARPA	Training and Assessment of Relational Precursors and Abilities
TOF	Transformation of function
VABS	Vineland Adaptive Behavior Scales
VBMAPP	Behavior Milestones Assessment and Placement Program

Glossary of Terms

Arbitrarily applicable relational responding	Learned relational responding under the control of arbitrary contextual cues beyond the formal properties of the relata involved and not reliant on their physically properties nor on direct experience with them.
Auditory-Auditory (AA) track	AA is one of four tracks within Stage 3 of the TARPA. All stimuli in the AA track are auditory (i.e., spoken nonsense syllables).
Auditory-Visual (AV) track	AV is one of four tracks within Stage 3 of the TARPA. The AV tracks involve both visual (e.g., abstract pictures) and auditory (i.e., spoken nonsense syllables) stimuli.
Combinatorial Entailment	A defining feature of relational framing that refers to the derivation of a new relation based on the combination of previously acquired relations. For example, combinatorial entailment applies when in a given context A is related in a characteristic way to B, and to C, and as a result a relation between B and C is mutually entailed.
Visual-Visual (VV) track	VV is one of four tracks within Stage 3 of the TARPA. All stimuli in the VV track are visual (i.e., abstract pictures).
Multi-modality	Refers to the four different tracks (visual-visual, auditory-visual 1, auditory-visual 2, auditory-auditory) in Stage 3 used in the TARPA.
Derived relational responding	The derivation of new untaught relations from previously acquired (i.e., either directly taught or derived) relations.
Track	There are four tracks in Stage 3 of the TARPA. A track involves 2 choice conditional discriminations where the sample and comparison are formally dissimilar (i.e., an arbitrary conditional discrimination).
Track Order	The sequence in which TARPA Stage 3 tracks are presented during a TARPA assessment.
Transformation of function	A defining feature of relational framing that refers to the modification of the stimulus functions of relata based on their participation in relational frames.

Order Analysis	A statistical technique that analyses descriptive data (e.g., passing/failing of levels) in order to identify hierarchies that exist in a data set.
Mutual Entailment	A defining feature of relational framing that refers to its fundamental bi-directionality. Mutual entailment applies when in a given context A is related in a characteristic way to B, and as a result B is now related in another characteristic way to A.
Relational Responding	Responding to one stimulus in terms of another.
Arbitrary	By social whim or convention. It is arbitrary, for example, that English speakers train the stimulus "apple" as being equivalent to a particular type of fruit; speakers in other language communities train up completely different stimuli.
Non-Arbitrary	Based on physical properties (e.g., colour, height, etc.)

Chapter 1

Literature Review

Applied Behaviour Analysis and Language Skills

A large body of research over the past 20 years has suggested that children with Autistic Spectrum Disorders (ASD) can benefit greatly from intervention based on the principles of applied behavior analysis (Anderson, Avery, DiPietro, Edwards, & Christain, 1987; Birnbrauer & Leach, 1993; Eikeseth, Smith, Jahr, & Eldevik, 2007; Fenske, Aalenski, Krantz, & McClannahan, 1985; Harris & Handleman, 2000; Howard, Sparkman, Cohen, Green, & Stainslaw, 2005; Lovaas, 1987; Smith, Groen, & Wynn, 2000). The teaching of language skills has been identified as one of the critical components of effective intervention programmes for children with ASD, for whom impairments in communication are a core diagnostic feature (e.g., Lord & McGee, 2001), and behaviour analytic approaches typically place a special emphasis on the analysis and development of such skills (e.g., see Sundberg & Michael, 2001).

The Lovaas Model

Behavioural intervention for teaching language skills to children with ASD has evolved considerably since Lovaas's (1987) seminal early behavioural intervention study. The Lovaas model targeted language skills predominantly within a discrete trial training approach (DTT; Lovass, 1977). DTT is a teaching method that comprises units of instruction allowing for multiple learning opportunities within a short period of time. The Lovaas model organised language skill acquisition into a number of different levels including first, intermediate and advanced. First language skills included basic rule following, verbal imitation, and matching; intermediate skills included the ability to identify and label objects; while advanced skills included using prepositions, pronouns, concepts, etc. Typically, listener skills ('receptive language') were targeted before speaker skills ('expressive language'). Speaker skills were not functionally defined; instead, the form (i.e., specific speech sounds) was targeted, often using chaining and differential reinforcement of successive approximations (Lovass, 1981). While this DTT approach had success the limited generalisation of these language skills across settings and people (i.e., stimulus generalisation) was noted as a weakness (e.g., McEachin, Smith & Lovaas, 1993). There were also concerns about the implicit nature of the attainment of goals and the sequencing of the language training (e.g., Luciano, Rodriquez, Manas, Ruiz, Berens & Valdiva-Salas, 2009).

Skinner's Verbal Behavior

Other researchers turned to the operant account of language laid out by Skinner (1957) in his book, *Verbal Behavior*, to inform procedures for targeting language skills (e.g., Sundberg & Partington, 1998; Carbone, 2000; Sundberg & Michael, 2001; Greer & Ross, 2008). In *Verbal Behavior,* Skinner offered the first comprehensive behavioural account of language. Within this theory human language is conceptualised as an operant behaviour, emerging from the social contingencies operating in the verbal community and as such is known as verbal behaviour. This perspective on cognitive functioning represented a major departure from mainstream cognitive psychology and other conceptualisations of language. The conventional approach to the study of language has tended to focus on the symbolic nature of words. Language-able humans are said to be able to "manipulate symbols" (Clark, 1994), or to use words to "refer" to objects, events, or relations (Premack, 1976). For Skinner, however, concepts such as "refer" and "symbol manipulation" remained undefined and therefore were not deemed useful in explaining human language (Cullinan & Vitale, 2008). Instead, Skinner emphasised the effects of environmental antecedents and consequences on verbal behaviour.

Specifically, Skinner (1957) defined verbal behaviour as any behaviour on the part of a speaker reinforced through the mediation of a listener who is trained by a verbal community so as to mediate such reinforcement. For example, if a child asks for water and this behaviour has historically been reinforced by listeners providing a proper consequence for it (i.e., water), and if the behaviour of the listener has been reinforced precisely so as to deliver such consequences, then the behaviour of the speaker in asking for coffee is “verbal” (Sundberg & Michael, 2001). According to this account then, verbal behaviour is behaviour subject to the same controlling variables as any other operant except that reinforcement of verbal behaviour is socially mediated (i.e., through the trained listener). Skinner describes a number of specific classes of verbal behaviour (i.e., the verbal operants); including mands, tacts, echoics, textuals, intraverbals and autoclitics, each of which is functionally as opposed to formally defined.

The example described above of a child asking for water is considered to be a ‘mand’ response. This is defined as “a verbal operant in which the response is reinforced by a characteristic consequence and is therefore under the functional control of relevant conditions of deprivation or aversive stimulation” (Skinner, 1957, pp. 35-36). If a child is thirsty (deprived of water) and obtaining water from a parent (the listener) reinforces their saying ‘water’ then a ‘mand’ has occurred. Teaching methods based on this conceptualisation involve altering the environment in order to evoke verbal behaviour that is under its control (e.g., contriving situations where the item to be manded is under deprivation or enhancing the reinforcer effectiveness of the item to be manded for the child).

The child in the previous example requested water (i.e., a mand), however if the same child produced the same arbitrary vocal stimulus (‘water’) but under different sources of environment control it would be considered functionally different. For example, if a child said ‘water’ in the presence of actual water and the response was reinforced by a social praise (e.g., ‘Good, yes it is water’) this would be an instance of a tact. If the child responded to the question ‘Tell me something you drink?’ with the response ‘water’ and this was reinforced by a social praise this would be an example of an intraverbal responses. The examples provided here of the mand, tact and intraverbal response highlight the functional nature of Skinner’s analysis; in each case the child produced the same arbitrary vocal stimulus (‘water’) but each response was under different environment control (antecedent and consequential).

Applications of Skinner’s Verbal Behavior. Applications of both Skinner’s analysis and extensions of his theory have proven very useful in teaching children with ASD functional communicative repertoires and this approach has become an integral part of early behavioural intervention (e.g, Sundberg et al., 1998; Sundberg et al., 2001). Take for example the success of mand training (e.g., Plavnick & Ferreri, 2012; Jennett, Harris & Delmolino, 2008). This approach to teaching children with ASD the ability to request has been demonstrated to have positive outcomes in terms of stimulus generalisation and maintenance as well as positive collateral effects (e.g., reduction in problem behaviours, increased eye contact). Research has also shown teaching procedures based on the tact response (e.g., Partington, Sundberg, Newhouse, & Spengler, 1994; Sundberg, Endicott, & Eigenheer, 2000; Schauffler & Greer, 2006) and the intraverbal response (e.g., Goldsmith, LeBlanc, & Sautter, 2007; Ingvarsson, Tiger, Hanley, & Stephenson, 2007) to be effective in establishing language skills in children with ASD.

While the *Verbal Behavior* can be implemented using DTT, DTT is highly structured and as mentioned above has been criticised on the basis that it does not facilitate stimulus generalisation (i.e., children learn only to respond to specific instructions under specific contexts and these responses fail to transfer to the "real world" in which the response needs to occur) (Smith, 2001). For this reason *Verbal Behavior* is often delivered using naturalistic teaching which is less structured than DTT. There is a vast applied behaviour analytic literature on different naturalistic teaching approaches; examples of which include Incidental Teaching, Milieu Teaching, Enhanced Milieu Teaching, the Mand-Model Approach, Pivotal Response Treatment, Embedded Teaching, and the Natural Language Paradigm (Barnes, Grannan, Lovett, & Rehfeldt, 2012).

All of these approaches have a common theme of using a child's motivation to create learning opportunities and because teaching trials are conducted in the natural environment with common, everyday stimuli, these approaches are associated with improved outcomes in terms of stimulus generalisation. Take for example incidental teaching, McGee, Krantz, and McClannahan (1985) demonstrated its effectiveness in teaching three boys with ASD to use prepositions to describe the position of items. Employing a multiple baseline design across participants and sets of prepositions, acquisition of tacts using prepositions and generalisation of the use of tacts in a novel environment was compared in instruction via incidental teaching and a DTT approach. Stimuli were placed on shelves out of the children's reach. The child initiated a trial by tacting or manding for an item, and the teacher required language elaboration (i.e., responding "The car is on top of the box." when asked "Where is it?" by the teacher). Responses were followed with behaviour specific praise and access to the desired item. Results indicated incidental teaching was as effective as a DTT approach in teaching the children to tact the position of items using prepositions; however, the incidental teaching procedure produced better generalisation of use of the taught prepositions in a free-play setting in comparison to prepositions taught using DTT.

There have also been other variations and extensions of the verbal behaviour approach (e.g., Greer & Ross, 2008). For example, the work of Greer and colleagues which has been referred to as verbal developmental theory (e.g., Greer & Speckman, 2009) is rooted in Skinner's analysis. While this approach targets the verbal operants, it also acknowledges the role of the listener and has emphasised the development of teaching procedures that establish the ability to learn in the absence of direct instruction. This work has identified key stages of pre-verbal and verbal development that children progress through on the way to becoming fully verbal, interactive human beings. The ability to imitate actions ("see-do") is a stage; the ability to actually learn by observing others is another and these abilities are targeted in a systematic and sequential manner (e.g., Greer & Ross, 2008). The work of Greer and colleagues will be further addressed later in this review.

Criticisms of Skinner's *Verbal Behavior*. Despite the successful application particularly in the field of behavioural intervention for children with ASD, it has been argued that at a theoretical and at an applied level Skinner's *Verbal Behavior* falls short as an approach to human language and cognition (Hayes, Barnes-Holmes & Roche, 2001). Criticism from outside behaviour analysis appeared almost immediately after *Verbal Behavior* was published. Chomsky (1959) pointed to the fact that children acquire thousands of words and many 'rules' that combine them without direct instruction and claimed that it would be impossible for a child to acquire such an extensive repertoire by direct instruction alone. He referred to this as the “poverty of the stimulus” and asserted that children “[...] must be born with a mental component that helps them learn language”. Despite its acceptance by mainstream psychologists, Chomsky’s (1959) review is seen as inaccurate and ill-informed by many within the behaviour analytic community (e.g., MacCorquodale, 1970; Schlinger, 2010). However, particularly recently, criticism of *Verbal Behavior* has also come from within behaviour analysis (e.g., Hayes et al., 2001).

Hayes and colleagues critique Skinner’s theory on a number of points. First, they argue that at a theoretical or conceptual level Skinner’s definition of verbal behaviour is too broad. In an operant experiment, the behaviour of the organism under investigation is reinforced by an experimenter trained to do so; thus, by Skinner’s definition the behaviour of non-humans in such experiments is verbal. Indeed, Skinner himself recognised this: ‘Our definition of verbal behavior, incidentally, includes the behavior of experimental animals […]. The animal and the experimenter comprise a small but genuine verbal community” (Skinner, 1957, footnote 11, p. 108). While Skinner might be content to allow this as a consequence of his approach, however, it seems bizarre at best.

A second and ultimately more serious criticism is that Skinner’s definition is nonfunctional as it “[...] turns not on the history of the organism of interest [speaker], but on the history of another organism [listener]” (Hayes et al., 2001, p. 12). This is a particularly serious criticism as it suggests a departure from the radical behavioural roots of behaviour analysis according to which functional definitions are stated in terms of the history of the individual organism and the current contextual circumstances. To illustrate their point, they describe a hypothetical experiment: Imagine looking into two identical experimental chambers; in each chamber, a rat is pressing a lever, and, on average, every fifth lever-press produces a pellet of food.

Clearly, one could switch the rats from one chamber to another, and after a period of adjustment, they would carry on as before, for as far as the rats are concerned, nothing of importance has changed. Now suppose that we look outside the chambers and we notice that in one case the pellets are being delivered by an experimenter on a variable-ratio-5 schedule, but in the other, the back side of the lever is agitating an open bag of pellets so that it jars loose a pellet at the same rate, and it falls into food hopper. By Skinner's definition, the behaviour of one rat would be considered verbal, but the behaviour of the other would not (Hayes et al., 2001, pg. 13). This is a fundamentally problematic outcome for an ostensibly functional approach to behaviour.

It has been suggested that these conceptual problems have contributed to the failure of *Verbal Behavior* to engender a broader research programme that extends beyond the research indicating certain successes of *Verbal Behavior* in the applied domain (e.g., Dymond & Alonso-Alvarez, 2010; Hayes, 1994; Hayes et al., 2001). Meanwhile, it can be argued that Skinner's account has also had detrimental consequences in the applied arena. The most substantial problem is with respect to the phenomenon of language generativity.

Generativity

Language generativity might be described as the ability to produce sentences never before said, and to understand sentences never before heard—to "speak with meaning," and "listen with understanding" (Hayes et al., 2001, p.3). Generativity is fundamental to the development of fully functional communication; and social interaction requires an increasingly complex repertoire in this respect on the part of the child. Despite its importance, establishing generative language in populations for whom it is deficient has proven to be a major challenge for behaviour analysts. For example, in the case of some children with ASD, rote, inflexible responding is a persistent problem in spite of intensive behavioural intervention (Lord et al., 2001).

Relational frame theorists agree to a substantial degree with Chomsky's criticism of Skinner's account in terms of it inadequacy with respect to explaining the sheer generativity of human language (e.g., Barnes-Holmes, Barnes-Holmes, McHugh & Hayes, 2004; Hayes et al., 2001) and argue that this becomes an issue with respect to populations with substantive deficiencies in this regard such as those with ASD (e.g., Barnes-Holmes & Murphy, 2007; Stewart, McElwee & Ming, 2013).

Response Generalisation

While Skinner's *Verbal Behavior* and applications of this theory do not adequately account or programme for generative responding, its importance is noted in the behaviour analytic literature. Generative responding is recognised as an important indicator of progress and is often discussed in terms of response generalisation (e.g., Stewart et al., 2013). A quote from Lovaas (2003) illustrates this; "...we had hoped that once the children learned to talk, they would develop the kind of response generalisation that would 'push them over' into normalcy" (p. 16). As another example, Sundberg (2008) describes the failure to show response generalisation as a critical barrier to children's progress in language:

> "The second type of generalization is response generalization. Here, a child may learn one response under the control of one stimulus (e.g., saying "cat" when asked to name an animal), but fail to provide any other responses that would be considered appropriate under that same stimulus (e.g., the response "rabbit" would also be considered a correct response to the question). The failure to demonstrate response generalization is often part of what is often identified as "rote verbal responding." A child always gives the same answer

> to questions, despite the fact that there could be many variations to what would be considered a correct answer." (p.118)

However, response generalisation as a concept is problematic for two reasons. First, the parameters in which response generalisation has been described have been ill defined. Some theorists incorporate physical dimensions into their definitions (e.g., Mayer, Sulzer-Azaroff & Wallace, 2011) while others speak of functionally equivalent behaviours (e.g., Cooper, Heron & Heward, 2007). Second, a satisfactory history of learning responsible for response generalisation that can be readily adapted to educational programming has not been outlined. Stewart et al. (2013) outline the failure of behaviour analysts to agree on a definition of response generalisation and also suggest that the term is more of a description of a phenomenon (i.e., emergent verbal behaviour skills) as a opposed to an explanation of how the phenomenon occurs or how it can be trained (e.g., Goldsmith, LeBlanc & Sautter, 2007; Noell, Connell & Duhon, 2006; Wesolowski, Zencius, McCarthy-Lydon & Lydon, 2005).

In a previous section verbal developmental theory (e.g., Greer & Ross, 2008; Greer & Speckman, 2009) was introduced as an extension of Skinner's *Verbal Behavior*. At this point it should be noted that this approach has contributed to the development of procedures that target generative verbal behaviour. Greer and colleagues have targeted the acquisition of naming as a key generative language skill. This work will be examined in a later section when teaching procedures that target generative language skills are discussed.

Derived Relational Responding

Generative verbal behaviour appears critical to fully functional language (Greer & Ross, 2008; Malott, 2003). Despite the success of behaviour analysis in other areas of ASD treatment, the field has arguably failed to provide an effective understanding of the skills that underlie this ability. Neither Skinner's *Verbal Behavior* nor the concept of response generalisation have provided a conceptual framework for developing procedures for teaching these skills. The protocol which is the focus of this thesis is based on Relational Frame Theory (RFT; Hayes et al., 2001); RFT as a more recent theory offers an alternative behavioural analytic account of language. This approach has identified derived relational responding (DRR) as the key process underlying language generativity. In what follows, a theoretical background to DRR and RFT is provided.

Stimulus Equivalence

The most prominent empirical example of DRR is stimulus equivalence, which was first demonstrated in Sidman's now classic 1971 study on teaching a young man with a learning disability to read. Prior to the experiment, the participant was already able to select particular pictures (A) in the presence of corresponding spoken words (B) and could produce the appropriate spoken words (B) in the presence of the pictures (A). He was then taught to pick appropriate textual stimuli (C) in the presence of the corresponding spoken words (B). He subsequently demonstrated a number of derived or untaught performances including (i) producing the appropriate spoken words in the presence of the textual stimuli which was a reversal of the taught performance (i.e., C→B); and (ii) choosing the appropriate textual stimuli in the presence of pictures and vice versa (i.e., C→A and A→C).

In effect, the participant was responding as if particular sets of spoken words, pictures and printed words were the same as or equivalent to each other and thus Sidman termed this phenomenon stimulus equivalence. Based on these and later results, Sidman suggested that stimulus equivalence is defined by three novel relations that emerge following training in a finite number of conditional discriminations (e.g., A-B; B-C). Drawing on mathematical set theory Sidman (1971) called these relations reflexivity, symmetry, and transitivity. Reflexivity is said occurs when each stimulus is matched to itself (e.g., A-A, B-B, C-C); symmetry occurs when each of the trained relations is reversed (e.g., B-A, C-B); and transitivity occurs when the two (or more) separate trained relations combine to produce a novel relation (e.g., A-C, C-A). Figure 1 provides a schematic representation of these emergent relations where thick lines represent trained relations (or pre-established relations) and dashed lines represent emergent relations.

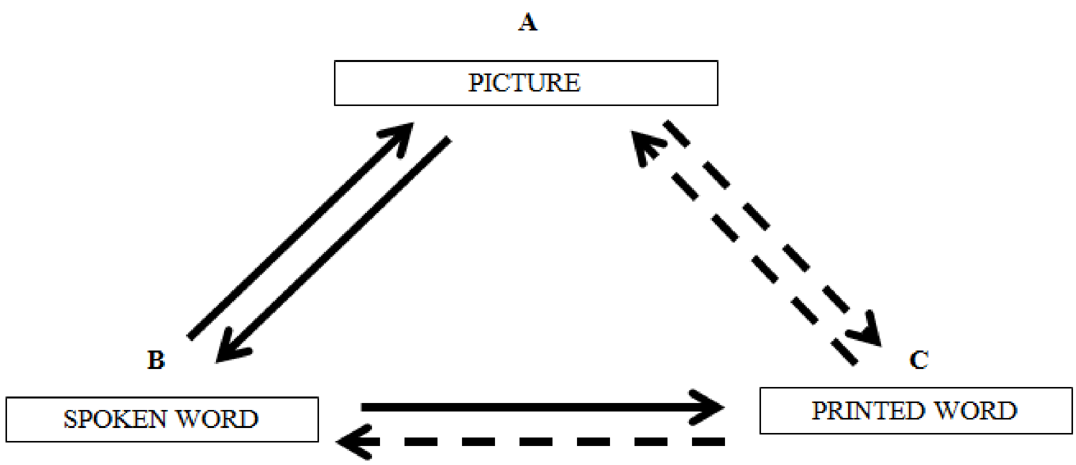


Figure 1. Sidman's (1971) equivalence training procedure -- thick lines represent trained relations or relations that were present pre-experimentally, and dashed lines represent emergent or "derived" relations.

The phenomenon of stimulus equivalence has generated much interest and debate within behaviour analysis for a number of reasons. First, it is not predicted by the traditional behaviour analytic principle of operant conditioning (Barnes, 1994). As described above (see Figure 1), in a typical stimulus equivalence experiment, participants are explicitly trained to select Stimulus A and Stimulus C in the presence of Stimulus B. Following this training and under test conditions, most participants will now consistently select Stimulus A in the presence of Stimulus C and vice versa. These untaught performances are not predicted because neither Stimulus A nor Stimulus C has a history of differential reinforcement as a discriminative stimulus with regard to the other, and as such neither stimulus should control selection of the other.

Second, stimulus equivalence has practical advantages since the fact that not all relations need be taught directly means efficiencies in terms of time and effort (Sidman, 1971). For example, Sidman (1971) reported that training 20 conditional discriminative relations resulted in the subsequent demonstration of approximately 40 additional derived performances.

Third, and perhaps most importantly, stimulus equivalence appears to be closely linked with human language. For example, in terms of its characteristics, it possesses several key features that are language-like including bi-directionality and generativity (Fields, Verhave & Fath, 1984). Furthermore, a range of empirical evidence supports the connection. One line of evidence has come from the contrast between verbal and non-verbal organisms in ability to show derived equivalence relations. In typically developing humans, derived equivalence relations develop in parallel with language ability (e.g., Lipkens, Hayes & Hayes, 1993) while humans with absent or delayed language repertoires tend to be unable to show equivalence (e.g., Devany, Hayes & Nelson, 1986) and the evidence for derived equivalence in non-humans is scant and at best disputable (e.g., Dugdale & Lowe, 2000; though see also Schusterman & Kastak, 1993). The link between equivalence and language is also supported by the results of neuroscientific research demonstrating that brain activity measured during derived equivalence tasks resembles that seen during language performance (e.g., Dickens, Singh, Roberts, Burns, Downes, Jimmieson & Bentall, 2001; Ogawa, Yamazaki, Ueno, Cheng & Iriki, 2010).

Transfer of Function

In addition a phenomenon that accompanies equivalence called transfer of functions can model how language can affect human behaviour (e.g., Dougher, Augustson, Markham, Greenway & Wulfert, 1994). Transfer of function describes a phenomenon in which the functions of a given stimulus (e.g., discriminative [e.g., Dymond & Barnes, 1994] or eliciting [e.g., Dougher, Augustson, Markham, Greenway & Wulfert, 1994]) transfer to other stimuli that participate in a relation of equivalence with the first stimulus. For example, Dougher et al. (1994) demonstrated the transfer of respondent functions through derived equivalence. Participants were taught a series of conditional discriminations which resulted in the formation of two four-member equivalence relations (i.e., A1; B1; C1; D1; A2; B2; C2; D2). Subsequently, one stimulus (i.e., A1) was given a respondent function through repeatedly pairing it with an electric shock. Physiological arousal was gauged through skin conductance. In a subsequent task, the same stimuli used in the conditional discrimination training randomly appeared on screen (i.e., A1; B1; C1; D1; A2; B2; C2; D2). As expected the 'A1' stimulus elicited increased skin conductance; however, unexpectedly the other stimuli in the relation (i.e., B1; C1; D1) also elicited an increased skin conductance This was not predicted as these stimuli were never directly paired with the electric shock, nor had they been related to ‘A1’ in terms of contiguity or formal similarity. In addition, the second stimulus set (i.e., A2; B2; C2 & D2) did not elicit increased physiological arousal. From an RFT perspective, this phenomenon allows a technical understanding of the influence of language over behaviour and the generativity that characterises this influence; this point will be elaborated upon later in this review.

Theories of Derived Stimulus Relations

The empirical link between derived equivalence and language is particularly intriguing and exciting for behaviour analysts. However, the question arises as to how it might be explained at a conceptual level. A number of different theories have been put forward in an attempt to provide an explanation (e.g., Sidman, 1994, 2000; Horne & Lowe, 1996; Lowenkron, 1998; Hayes, Barnes-Holmes & Roche, 2001). Sidman (1994; 2000) has argued that stimulus equivalence is a phenomenon that is explained through reinforcing contingencies, probably due to phylogenesis. This means that stimulus equivalence is a basic behavioural process just like reinforcement, generalisation and discrimination except that equivalence is a capacity that evolves in humans alone. This is considered non-parsimonious because it suggests a whole new basic behavioural phenomenon (e.g., Hayes & Barnes, 1997; Sidman, 2000; Hayes et al., 2001).

Naming (Horne & Lowe, 1996) and Joint Control (Lowenkron, 1998) are somewhat similar in that both explain DRR as being based on a covert process (either naming or joint control respectively) that is established through multiple exemplar training (MET) and which subsequently enables DRR to occur. MET involves training a pattern of behaviour in the context of multiple physically different sets of stimuli; in other words, training is such that while conditions for reinforcement remain constant, the stimulus topographies are deliberately varied. In fact, as we shall see, RFT also explains DRR as being based on MET. However, a key difference between these two accounts and that provided by RFT is that these accounts both argue that MET allows a covert mediational process which then allows DRR. The fact that these theories appeal to a covert mediational process should be a key criticism of them from a radical behavioural point of view. Mediational concepts are seen as misleading with regard to the radical behavioural goal of gaining prediction and influence over behavior. Reliance on such processes also draws criticism from a more general scientific perspective, since the requirement of any additional process, mediational or otherwise, makes them less parsimonious, all other things being equal, than theories such as RFT that do not appeal to such processes (Stewart et al., 2013). The criticisms just presented are theoretical but in fact there is also empirical evidence countering naming in particular. For example, it has been empirically demonstrated that derived relations can emerge in the absence of naming (e.g., Barnes-Holmes, Barnes-Holmes, Roche & Smeets, 2001; Luciano, Gómez & Rodríguez, 2007).

The one remaining major behaviour analytic theory of derived relations is RFT (Barnes-Holmes et al., 2001; Hayes et al., 2001; Hayes, 1991; 1992). The research in this thesis is based on this theory, so in the next section it will be described in detail.

Relational Frame Theory

RFT suggests that the link between derived equivalence and language comes about because they are essentially the same phenomenon, namely generalised contextually controlled arbitrarily applicable relational responding or more simply, relational framing.

Background

Many species, including humans can demonstrate generalised relational responding based on physical properties of the relata (e.g., picking an object that is physically the same as another object), referred to as non-arbitrary relational responding (e.g., Hayes, Fox, Gifford, Wilson, Barnes-Holmes & Healy, 2001; Reese, 1968; Stewart & McElwee, 2009). However, RFT posits a further type of generalised relational responding in which the relational response is determined by contextual cues independent of the properties of the related objects. For example, if I am told that X is the same as Y and Y is the same as Z, then I can derive that Y is the same as X, Z is the same as Y, X is the same as Z and Z is the same as X. In this case, the pattern of DRR is not based on the actual properties of the letters, but on the contextual cue 'same as', which was established to function as such in the course of my learning history.

RFT theorists argue that this type of sameness (or 'coordination') relational responding is the phenomenon that underlies stimulus equivalence. From the RFT perspective, however, equivalence/sameness is only one type of derived relational pattern. Over the last two decades, RFT researchers have provided empirical evidence for a variety of other patterns of derived relations in addition to equivalence including distinction (e.g., Roche & Barnes, 1997), comparison (e.g., Berens & Hayes, 2007), opposition (e.g., Barnes-Holmes, Barnes-Holmes & Smeets, 2004), analogy (e.g., Persicke, Tarbox, Ranick & St.Clair, 2012; Stewart, Barnes-Holmes &Roche, 2004), temporality (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004) and deixis (e.g., McHugh, Barnes-Holmes, & Barnes-Holmes, 2004) (see Dymond & Roche, 2013, for an overview of recent research) and RFT proponents argue that this variety of relational patterns or frames underlies the diversity, complexity and generativity of human language.

Two characteristics of derived arbitrarily applicable relational responding or relational framing that seem particularly important from the perspective of improving language skills in children with ASD are that it is (i) extremely generative and (ii) that it can be trained. Evidence for the generativity of this behaviour has been provided by many of the RFT studies that have appeared in the literature, though a few in particular deliberately highlight this characteristic (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004; Stewart, Barnes-Holmes & Roche, 2004; Wulfert & Hayes, 1988). For example, Stewart et al. (2004) used an RFT-based procedure known as the Relational Evaluation Procedure (REP) to establish abstract shapes as contextual cues for SAME and DIFFERENT relations and for TRUE and FALSE responses, respectively, and then employed these cues both to model analogical reasoning as the relating of derived relations between derived relations as well as to demonstrate that an in-principle infinite number of new analogical relations was possible based on this technique.

Relational framing as an operant behaviour

As an operant, relational framing can also be trained. RFT proponents have argued that DRR is learned naturally by typically developing children via everyday language interactions during which they are exposed to contingencies that establish these response patterns (e.g., Lipkens et al., 1993; Luciano et al., 2007). From this perspective, caregivers provide masses of learning opportunities for children to relate stimuli, which follow a consistent pattern across exemplars (i.e., MET). Consider, for example, the very early history of training responsible for establishing sameness (coordination) relations between a word and an object. Caregivers will often utter the name of an object in the presence of an infant and then reinforce any orienting response that occurs towards the particular object (‘hear name A → look at object B’). They will also often present an object to the infant and then model and reinforce an appropriate naming response (‘see object B → hear and say name A’). RFT suggests that after a sufficient number of name-to-object and object-to-name exemplars have been taught, the generalised operant response class of derived ‘naming’ is established. Effectively, the multiple-exemplar bi-directional training establishes particular contextual cues as discriminative for the derived naming response. For instance, imagine a child with such a history is told "This is a teddy." Contextual cues, including the word 'is' and other aspects of the naming context, will now be discriminative for symmetrical responding between the name and the object. Thus, without any additional training, the child will now point to the teddy when asked "Where is the teddy?" (name A → object B) and will answer "teddy" when presented with the teddy and asked "What is this?" (object B → name A). RFT argues that such MET also enables stimulus equivalence. In this case the key contextual cue that brings the pattern to bear is the matching to sample protocol itself, since children are also taught in educational exercises that when a stimulus is picked in the presence of another stimulus, then the two stimuli ‘go together’. Hence, relational framing is seen as a generalised or overarching response class generated by a history of reinforcement across multiple exemplars, and once established any stimulus or response event, irrespective of form, may participate in a relational frame.

Properties of Relational Framing

While there are many different forms or patterns of relational framing from the RFT perspective, they all share three core properties: mutual entailment, combinatorial entailment and transformation of stimulus function. These terms are analogous to symmetry, transitivity, and transfer of function in stimulus equivalence, but are broader terms that can be applied to relations other than sameness or coordination.

Mutual entailment describes that feature of relational framing whereby if a stimulus A is related to another stimulus B in a certain context, then a novel relation between B and A may be derived in that context. Symmetry (i.e., where ‘A goes with B’ is trained and ‘B goes with A’ is derived) seen in stimulus equivalence is a sub-type of mutual entailment where the relation is one of coordination. Critically however, mutual entailment describes the fundamental bi-directionality of relational responding, even when such bi-directionality is not symmetrical; mutual entailment can involve relations other than sameness or coordination, and as such the specific frame determines the nature of the derived relation, for example, if A is bigger than B, then B is smaller than A; if A is above B, then B is below A, and so on.

Combinatorial entailment refers to a derived stimulus relation in which two or more stimulus relations (trained or derived) mutually combine (e.g., if A→B and C→B, then A→C and C→A). Transitivity seen in stimulus equivalence (i.e., ‘if A goes with B’ and ‘C goes with B’, then ‘A goes with C and C goes with A) is a sub-type of combinatorial entailment where the relation is one of coordination. However, as with mutual entailment, combinatorial entailment can involve relations other than sameness or coordination, and therefore the specific frame determines the nature of the derived relation. For example, in a frame of opposition, if A is opposite B, and C is opposite B, then A and C are the same; in a frame of comparison, if a euro is worth more than a dollar, and a dollar is worth more than a ruble, then a ruble is worth less than a euro; and so on.

The third property of relational framing, transformation of stimulus function, is the behavioural property that provides stimulus relations with psychological content. Transformation of stimulus function describes how when a given stimulus in a relational frame has certain psychological functions, the functions of other stimuli in that frame are modified in accordance with the underlying derived relation. In a previous section, it was outlined how Dougher et al. (1994) demonstrated the transfer of respondent functions through equivalence relations. From an RFT perspective transfer of function through equivalence relations is a sub-type of transformation of function. With transfer of function, the psychological function that appears in the related stimulus is the same as the function inherent in the original stimulus; however in the case of transformation of function, the function is transformed based on the underlying derived relation (i.e., the specific frame determines the change in behaviour). This has also been demonstrated experimentally. In a study by Dougher, Hamilton, Fink and Harrington (2007) participants were taught a relational network between three arbitrary stimuli; A is less than B and B is less than C. Subsequently, B was paired with an electric shock. Results showed that C elicited more physiological arousal than B, even though it was never paired directly with the shock.

Transformation of function is considered critically important, as it involves behaviour being changed via relational framing. For example, imagine that a child has already learned that she can purchase an item in the shop with a particular coin and is then told that another, previously unseen coin is worth more than the first one. If the child has a sufficient repertoire of comparative relational framing, then the reinforcing function of the second coin will be transformed so that it becomes more appetitive than the first, and if given a choice, the child will likely ask for the second, novel coin, in preference to the first, even though she has only had directly reinforcing experience with the latter.

Derived Relational Responding skills in persons with Developmental Delay

As discussed above, typically developing children learn DRR through exposure to the natural language environment. However, many children with ASD and/or developmental delay, do not easily learn this key form of responding and thus show deficits in comparison with normally developing children with respect to relational framing ability and hence linguistic generativity (e.g., Barnes, McCullagh & Keenan, 1990; Devany, Hayes & Nelson, 1986; Eikeseth & Smith, 1992; O'Donnell & Saunders, 2003; Rehfeldt, Dillen, Ziomek & Kowalchuk, 2007; Rehfeldt & Barnes-Holmes, 2009).

Devany et al. (1986) looked at performance on a stimulus equivalence task in three groups of children ranging in chronological age while matched for mental age. Participants included typically developing preschoolers, children with developmental delay who used speech or signs spontaneously and appropriately, and children with developmental delay who did not. All children were taught a series of four related conditional discriminations and were then tested to see whether they showed patterns consistent with stimulus equivalence. All of the language-able children (developmentally delayed and typically developing) showed derived novel untaught relations based on equivalence, whereas none of the language-disabled children did so.

Barnes, McCullagh, and Keenan (1990) tested the effects of training four conditional discriminations with unfamiliar stimuli on the acquisition of emergent relations, in participants with the following impairments/repertoires: (i) children who were diagnosed with learning disabilities but demonstrated verbal competency equivalent to 2+ years; (ii) children who were diagnosed with learning disabilities along with severe to profound hearing impairments, but demonstrated verbal competency equivalent to 2+ years; and, (iii) children who were diagnosed with learning disabilities along with severe to profound hearing impairments, and did not demonstrate verbal competence. While all of the participants learned the conditional discriminations, those without verbal competencies were far less likely to demonstrate emergent relations than their verbally competent peers.

Despite their disadvantage in this regard, evidence from RFT and other researchers indicate that children with ASD and/or developmental delay can benefit from training that specifically targets these relational repertoires (e.g., Greer & Ross., 2008; Murphy & Barnes-Holmes, 2009a; 2009b; Rehfeldt et al., 2007). Some of this research will now be discussed.

Teaching Derived Coordinate Relations

Over the last decade a number of studies have provided empirical demonstrations of training DRR skills in young children for whom they are deficit or absent. This research is primarily based on RFT; however, other approaches such as naming theory (Horne & Lowe, 1996) have also contributed to the literature on DRR albeit from a different theoretical position. The RFT conceptualisation of DRR as a higher order operant behaviour outlines a clear learning pathway for these skills. From this point of view, MET of relational responding is seen as the means by which typical children develop language and it is also seen as a very important potential means of remediation in many or most cases when language is deficient.

Typically, an intervention based on MET is delivered as follows: if a participant after acquiring the directly taught relations (e.g., A-B relations), does not demonstrate DRR (e.g., derive B→A relations) they receive training in bidirectional responding across multiple sets (e.g., A ↔ B) where the stimuli may be topographically different but the conditions for reinforcement remain the same. This process of training and testing continues until the participant demonstrates DRR (i.e., an overarching relational operant class is established). In what follows, some empirical examples of the training of these skills in a frame of coordination or equivalence with both typically developing children and children with ASD and/or developmental delay are outlined.

Targeting Mutually Entailed Responding

Barnes-Holmes et al (2001) provided one of the first empirical demonstrations of the use of MET to facilitate mutually entailed relations in twelve typically developing young children. Across three experiments children were taught a number of action-object conditional discriminations (e.g., when experimenter waves, choose toy car; when experimenter claps, choose doll). Children were then tested for the derived mutually entailed relation (e.g., given toy car, wave; given doll; clap). Children who failed to show these patterns were exposed to MET (i.e., bidirectional training in additional sets of action-object relations) and a multiple baseline design showed that this training led to emergence of derived relations.

A number of studies have also established mutually entailed relations (specifically ‘naming’ relations) in children with developmental delay including ASD under the rubric of verbal development theory (Greer & Speckman, 2009). From an RFT perspective ‘naming’ involves mutually entailed relations of coordination between arbitrary verbal topographies and environmental stimuli or events (e.g., word-object). People with naming repertoires can hear caretakers label a stimulus and can then respond to the stimulus as both a listener and a speaker without direct instruction. The acquisition of naming is considered to be a fundamental DRR skill as it allows the proliferation of vocabulary (Greer, Stolfi, Chavez-Brown, & Rivera-Valdes, 2005).

Greer et al. (2005) used MET to establish mutually entailed responding in the form of the listener to speaker component of naming (i.e., being able to tact an object as a result of previously being taught to respond as a listener to the same object) in three children with mild developmental delay. First, children were taught matching responses using discrete trial presentations where the teacher spoke the name of the picture as the child matched (i.e., name-object; A-B training). Once children acquired the matching responses, the untaught repertoires were probed—specifically, these probes tested for the emergence of derived object-name (i.e., tact) relations (or B-A relations). The children who did not demonstrate DRR (i.e., after learning A-B relations they did not mutually entail B-A relations) were exposed to MET. During MET, they were directly taught to respond as listeners (e.g., match and point-to) and speakers (e.g., pure and impure tacts) to two sets of five pictures (i.e., response topographies were rapidly rotated across the teaching session). As a function of this MET, untaught speaker responses (mutually entailed responding; A↔B) emerged after only matching responses were taught for a third novel set of stimuli.

Two further studies (Fiorile & Greer, 2007; Greer, Stolfi & Pistoljevic, 2007) replicated the effects of the previous study and also isolated MET as the variable that led to the emergence of naming by comparing singular exemplar instruction and MET on the emergence of naming in preschool children who were missing the repertoire.

Targeting Combinatorially Entailed Responding

Working from an exclusively RFT perspective a number of studies have established derived coordinate relations. Gomez, Lopez, Martin, Barnes-Holmes and Barnes-Holmes (2007) extended previous work with mutually entailed relations described above (e.g., Barnes-Holmes et al., 2001) by successfully using MET to establish derived combinatorially entailed relations in young typically developing children.

Murphy, Barnes-Holmes and Barnes-Holmes (2005) also demonstrated the efficacy of MET for establishing derived combinatorially entailed relations in children with developmental delay. Specifically, Murphy et al. (2005) examined the development of a more flexible manding repertoire for children with ASD as a result of transfer of function through equivalence relations. Transfer of function was examined through the use of a token board game that was used to contrive conditioned establishing operations for particular types of tokens in order to fill the board.

Similar to the use of a picture exchange system for manding, one stimulus (A1) was trained to have a particular discriminative function for manding (obtaining one colour of token) and a second stimulus (A2) was trained to have a different discriminative function for manding (obtaining another colour of token). Subsequently, participants were trained in A-B and B-C conditional discriminations, and then tested for their ability to mand using C stimuli (i.e. transfer of the discriminative function of A1 to C1 and A2 to C2, which in this context has been termed derived manding).

One participant in this study did not immediately demonstrate transfer of function following training in the relevant conditional discriminations, and for this participant, MET was conducted. After directly training transfer of function (i.e. training the participant to mand using both A1 and C1 stimuli and A2 and C2 stimuli), a novel set of stimuli were used to repeat the initial mand and conditional discrimination training and then test for transfer of function. Following MET with three such sets of stimuli, the participant was able to demonstrate transfer of function with a fourth, novel set of stimuli. Luciano, Gomez-Becerra and Rodriequez-Valverde (2007) also successfully used MET to establish mutually entailed and combinatorially entailed relations in a young typically developing child whose age ranged from 15-23 months during the study.

This review has considered empirical examples of the training of DRR skills as they pertained to a frame of coordination or equivalence. However, it is worth noting that a number of studies have targeted other pattern of DRR for remediation including frames of comparison, opposition and perspective taking (for a review, see Rehfeldt & Barnes-Holmes, 2009; Stewart, Roche, O'Hora &Tarbox, 2013). At this point, the RFT approach to language, as contextually controlled patterns of DRR has been outlined. The studies reviewed here have illustrated the generativity of DRR and how this phenomenon is relevant to the rapid production of new patterns of responding. In addition, they offer empirical evidence not only of the generative effects themselves but also how the procedure of MET can be used to train these effects when they are deficient. As such, this approach shows a lot of promise as an account of language and language generativity in particular.

Assessing the development of Derived Relational Responding

Given the suggested importance of DRR skills for language and cognition, at this point in the progression of DRR based work on language, one thing that may be helpful for both researchers and practitioners alike is to be able to have a way of assessing an individual's repertoire of DRR skills that seem to support generative language skills. The development of the TARPA (Training and Assessment of Relational Precursors and Abilities) as a protocol for this purpose is the focus of the current thesis.Standardised assessment tools have played a very important role within behaviour analysis and psychology more broadly by providing a normative baseline of age appropriate responding that serves both applied as well as basic science purposes. One of the first such tools within behaviour analysis was the Assessment of Basic Learning Abilities (ABLA; Kerr, Mayerson & Flora, 1977; Stewart, McElwee & Ming, 2010). The ABLA is a table top protocol employing manipulables that is used to measure whether individuals with developmental disabilities can learn (using standard prompting and reinforcement procedures) to perform a series of discrimination tasks of varying levels of difficulty. The test consists of six sequenced mini learning tasks. The authors describe the tasks as: a simple imitation task, a two-choice position discrimination, a two-choice visual discrimination, a two-choice visual match-to-sample discrimination, a two-choice auditory discrimination, and two-choice auditory-visual discrimination (Martin & Yu, 2000). While the ABLA has been successful in predicting a variety of outcomes, including the ability to demonstrate equivalence, and performance on standardised tests of linguistic and cognitive ability (e.g., Marion, Vause, Harapiak, Martin, Yu, Sakko, & Walters, 2003; Richards, Williams, & Follette, 2002; Vause, Martin, Yu, Marion & Sakko, 2005), it does not measure DRR skills.

Other more recently developed and more prominent examples of behaviour analytic assessment protocols are the *Verbal Behavior Milestones Assessment and Placement Program* (VBMAPP; Sundberg, 2008) and the *Assessment of Basic Language and Learning Skills* (ABLLS; Partington & Sundberg, 1998). Both these tools are theoretically situated within Skinner's *Verbal Behavior* (1957) and are designed to: measure a range of linguistic and functional skills; guide individualised intervention and instruction needed to address skills deficits; and evaluate progress over the course of a treatment programme. These assessment tools are widely employed in intervention programmes for children with ASD and/or developmental delay and arguably do implicitly measure generative verbal skills. The following items taken from the assessments illustrate this:

"Shows response generalization for 5 items (i.e., tacts the same stimulus with two different words *teacher* and *Katie*; *cat* and *Garfield*; *dog* and *Maggie*)." (Sundberg, 2008b, p. 46)

"Shows response generalization by describing the same 10 objects, events, pets, people, etc. in 3 different ways (e.g., in reference to a pet dog Toby, the child says at different times *a dog*, *an animal*, *Toby*)." (Sundberg, 2008b, p. 67)

"Generalized response forms: The student will be able to use other appropriate responses after learning a response to a given situation...Upon seeing a dog, the student may say "dog", "puppy", "K-9", "pooch", etc. When answering a question regarding "things to eat", the student may say "apple, banana, bread" OR "cake, pizza, apple." (Partington & Sundberg, 1998, p. 62)

However, by measuring a particular response to a particular statement it is difficult to identify the relevant source of stimulus control (i.e., it is not clear whether the learner has truly demonstrated generative language skills or not). While the ABLA, VB-MAPP and ABLLS facilitate the assessment and training of key behavioural repertoires, none of them explicitly measure DRR skills which as outlined previously are considered critical to fully functional generative language. The TARPA aims to assess this key behavioural repertoire.

Training and Assessment of Relational Precursors and Abilities (TARPA)

The TARPA is a multi-level computer based protocol that assesses a number of key forms of responding that are critical (from an RFT / DRR perspective) to the development of generative behaviour, including: (1) basic discrimination, (2) conditional discrimination (similarity), (3) conditional discrimination (non-similarity), (4) mutually entailed relational responding, (5) combinatorial entailed relational responding, (6) transformation of functions. Each stage is further subdivided into multiple levels based on modality of the stimuli (e.g., visual only, auditory only, etc.) involved and the more advanced stages are subdivided again into sections based on whether the participant is being assessed for their ability to learn particular repertoires necessary for the derivation of relations or their ability to derive those relations themselves.

Thus, the TARPA provides a comprehensive multi-modality assessment of the prerequisite skills to relational responding skills as well as assessment of relational responding abilities at the level of equivalence. The hierarchical ordering and content of the stages and levels of the TARPA is based on relevant theory and research. The initial stages were designed and arranged based on evidence from previous studies which investigated basic learning skills including simple discrimination skills and conditional discrimination skills with nonarbitrary and arbitrary stimuli (e.g., Jackson, Williams, & Biesbrouck, 2006; Kerr et al., 1977; McIlvane, Dube, Kledaras, Innaco & Stoddard,1990; Vause, Martin, Yu, Marion, & Sakko, 2005; Williams & Reinbold, 1990).

The ABLA (Kerr et al., 1977) described previously, is one behaviour analytic tool that has influenced early stages and levels within the TARPA. One of the earliest findings of ABLA research was that the six levels were hierarchically ordered in difficulty. Kerr et al (1977) found that participants who passed a certain level of discrimination ability also passed lower levels, and participants who failed at a certain level did not pass subsequent higher levels. This pattern was evident for 111 of the 117 participants in their study. This finding has also been replicated by other studies with typically developing children as well as persons with developmental disabilities (e.g., Casey & Kerr, 1977; Martin, Yu, Quinn & Patterson, 1983).

The categories of behaviours measured by the ABLA that move from a basic discrimination, to a conditional discriminations based on formal similarity to conditional discriminations with no formal similarity have been shown to be relevant to DRR and language. For example, a study by Vause et al. (2005) suggested that performance on the ABLA may be a prerequisite for demonstrating equivalence relations. Five participants with minimal verbal repertoires were studied; only the two participants who passed ABLA Level 6 (which assesses an individual's ability to learn an arbitrary auditory-visual conditional discrimination) demonstrated positive equivalence test outcomes. Other research has demonstrated a relationship between performance on the ABLA and functional language skills such as vocal imitation, manding and tacting (e.g., Marion et al., 2003) as well as adaptive and cognitive functioning skills more generally (e.g., Richards et al., 2002). Similar categories of behaviour are measured within the initials stages and levels of the TARPA and the TARPA also employs similar standard testing procedures to those used in the ABLA. However, the TARPA assesses these skills in a more systematic manner across multiple stimulus modalities (e.g., visual, auditory & audiovisual).

While early stages of the TARPA are modelled somewhat on the ABLA, the TARPA is theoretically situated in RFT and it also covers a much more extensive range of skills (e.g., relational responding and relational framing). As such, the arrangement and content of higher stages within the TARPA are derived from research on the emergence of relational framing abilities some of which has already been discussed (e.g., Lipkens et al., 1993; Luciano, et al, 2007; Murphy et al., 2005). Apart from the necessity of the evaluation of this assessment tool as part of the development of a comprehensive DRR training protocol, this work also constitutes an important step in its own right. Although there is by now a significant amount of RFT-based research into many basic and applied areas of language and complex human behaviour, there is as yet no standardised tool for the assessment and training of relational framing abilities themselves or of the precursor skills supporting those abilities.

TARPA Research and Development

Moran, Stewart, McElwee and Ming (2010) tested an earlier version of the TARPA using children with ASD. In this study, scores on the TARPA were found to correlate strongly and significantly with overall scores on the Vineland Adaptive Behavior Scales (VABS; Sparrow, Cicchetti & Balla, 2005). Further tests examining the relationship between TARPA and VABS sub-scale scores showed that the TARPA correlated highly, though just outside significance with the Communication sub-scale, and showed lower correlations with the Daily living and Socialisation sub-scales, providing some support for its potential as a measure of language skills specifically.

Moran et al. (2010) also provided some insight into the hierarchical ordering of the TARPA. Ideally, in any multi-stage hierarchically structured test, a participant should not be able to pass a stage of purported higher difficulty having failed a stage of purported lower difficulty. Inspection of the data of individual participants across stages of the TARPA in this study indicated that in general the TARPA conforms to this criterion. Lastly, Moran et al. (2010) also showed some differences in performance based on stimulus modality. For instance, a number of participants showed weaker responding on sections involving exclusively auditory stimuli than on sections involving at least some visual stimuli, irrespective of the strength of their overall performance.

Pilot Testing. Subsequent to Moran et al. (2010) a new, arguably more comprehensive version of the TARPA was developed. This version of the protocol involved both two- and three-choice comparisons and tested for the properties of derived relations (i.e., mutual and combinatorial entailment and transformation of function) across stimulus modalities. The Moran et al. protocol was limited with respect to the stimulus modalities, for example, there was no training or testing of these properties with exclusively visual stimuli and both mutual and combinatorial entailment were tested with auditory stimuli only. The new protocol tested all three properties in each of four different modalities (i.e., all visual, all auditory, and two separate auditory-visual modalities). In addition to this, the new protocol increased the range of tests of basic discrimination and added tests of generalised identity matching in both the visual and auditory modalities. This version of the protocol also included tests of maintenance of trained relations in both the mutual and combinatorial entailment levels (with several in the latter). Finally, the transformation of function test in this version was a simpler and logistically more practical assessment. In Moran et al., the transformation of function test was adapted from a task employed by Murphy, Barnes-Holmes & Barnes-Holmes (2005) in which participants were assessed for transformation of the discriminative functions of stimuli that could be employed to mand for tokens in the context of a game. This task was relatively complex and its implementation in the context of a computer-delivered protocol increased this complexity. In the current version of the protocol, a very different and considerably simpler test was employed in which participants were initially required to perform particular actions in the presence of particular stimuli and were then tested for transformation of discriminative functions in the presence of stimuli in derived relations with the initial set.

This more comprehensive version of the TARPA was subject to pilot testing (as part of this and other projects) prior to its use in the current thesis. This involved testing typically developing young children (between the ages of 2-10 years) on the protocol (i) in order to gain an insight into the practicalities of administering the protocol in its newer form and (ii) to identify any user interface and general protocol issues that may obscure identification of DRR performance. A number of specific issues emerged during testing that informed further adjustments to the protocol.

All participants demonstrated difficulty in acquiring baseline conditional discrimination that involved three choice comparisons. For example, a 10 year old child during pilot testing with three-choice comparisons took extensive training to acquire the requisite conditional discriminations and even then did not demonstrate equivalence. The initial inclusion of three-choice comparisons was based on an argument against the use of two choice procedures in studies of emergent matching to sample (e.g., Carrigan et al., 1992; Sidman, 1987) which has been refuted somewhat in more recent literature (e.g., Boelens, 2002; Saunders, Chaney &

Marquis, 2005). The decision to use just two choice tests throughout the protocol was taken therefore in order to minimise training times and thus make TARPA assessment as brief and convenient as possible.

The pass / fail criterion for individual levels was also examined during pilot testing. Initially, in order to pass a training level (e.g., acquire a conditional discrimination) or tested level (e.g., of derived relational ability) participants had to accumulate a fixed number of consecutively correct trials without any errors. For example, in the case of acquiring a conditional discrimination, participants had to accumulate six consecutively correct trials; similarly to pass a test of derived relational ability they had to respond correctly on each of the six tested trials. However, during pilot testing this appeared to be too strict in that it allowed very little leeway for error on the part of the participant. During training participants frequently got three, four and five responses consecutively correct before making an error; likewise, during testing participants often made one error and as such failed the overall test (e.g., mutual entailment, combinatorial entailment, etc.). Often these errors appeared to be a function of attention or the prolonged length of the task. Hence, it was decided to adjust the criterion to allow participants one error. In the case of training levels, the pass criterion was adjusted to 5/6 consecutively correct responses and in the case of tests of derived relational ability, the criterion was adjusted to 7/8 correct trials. In addition, new stimuli (different sounds and pictures such as clapping and balloons) were added to the reinforcement component of the TARPA in order to enhance its appeal to young children.

Another issue related to the TARPA stimuli; the stimuli used in the TARPA are abstract and arbitrarily configured to control for history effects and are similar to stimuli typically seen in basic laboratory investigations involving derived stimulus relations. Despite attempts to ensure that non arbitrary relations between stimuli do not come to influence responding in arbitrary match-to-sample procedures, participants behaviour can sometimes come under this form of control (Stewart & McElwee, 2009); participants often identify consistencies and may respond on the basis of these in training and/or testing. An example of this type of inappropriate stimulus control was identified during pilot testing; in the test for combinatorial entailment involving visual stimuli (i.e., visual-visual track), one participant commented that 'the big ones go together' (i.e., he was responding on the basis of non-arbitrary properties of the stimuli). This set of stimuli was subsequently changed and no further issues with stimuli were noted.

Inter Observer Agreement (IOA) data (i.e., calculating the percentage of trials on which two independent observers agree as to the outcome of individual TARPA trials) was used in Moran et al. (2010) and was also given consideration during pilot testing. It was decided however that because the TARPA is completely automated with the exception of one level (i.e., transformation of function that requires interaction between the experimenter and participant) it would be unfeasible to employ a second experimenter for this research programme. It was deemed necessary in Moran et al. (2010) because in this study a mouse was used as the input device and in the case where participants were not able to manipulate a mouse in order to select their responses the experimenter had to use the mouse for these participants by making the selection once the child had indicated their response by pointing to the stimulus. This raised the issue of inadvertent cueing. The current protocol however was delivered on a touch screen computer which eliminated these difficulties.

Summary and Research Aims

As outlined in this review, communication is a core deficit in ASD and as a result the development of interventions that target communication ability is a priority for autism research. One key deficit in communication ability of children with ASD is generative language; in other words, the ability to be able to understand as well as to create novel sentences (Greer & Ross, 2008). Arguably, at a conceptual and applied level behaviour analysis has struggled to adequately account for and teach generative language skills (e.g., Stewart et al., 2013). However, research into derived stimulus relations and in particular RFT (Hayes, et al., 2001) has provided important new theoretical and empirical insights into the core skills needed for generative language. RFT has begun to outline a promising set of procedures (e.g., Rehfeldt & Barnes-Holmes, 2009) for targeting these skills. However, there are no existing tools that measure these DRR skills and their prerequisites. The aim of this research is to develop the TARPA as a robust and efficient method for assessing the emergence of derived relational ability and thus generative verbal behaviour in a number of different populations including typically developing children and children with ASD.

This research aims to validate the TARPA by (i) testing the external validity of the TARPA through correlation with alternative measures of language and cognitive ability and (ii) formally examining the hierarchical structure of the TARPA. It is envisaged that this process will lead to possible refinements of the protocol, which in turn will require testing. The aim of this regime of testing and development is also to (i) identify patterns of responding that require further investigation; (ii) indicate possible avenues for improving the validity of the TARPA; and (iii) provide a basis for the development of a training dimension of the TARPA. In conclusion, the results from the research will provide guidance in respect to a number of important parameters and will thus be an important step in the process of the development of the TARPA as a key tool for assessing and facilitating the development of generative verbal behaviour.

Chapter 2

Experiments 1 & 2

Comparing Derived Relations and Language Ability in Children with Autism Spectrum Disorders & Typically Developing Children: An Initial Test of the TARPA Protocol

The aim of this first study was to extend previous work on the TARPA (Training and Assessment of Relational Precursors and Abilities; Moran, Stewart, McElwee & Ming, 2010), a novel protocol for assessing the emergence of DRR. As mentioned in Chapter 1, Moran et al. (2010) tested an earlier version of the TARPA using children with ASD. One aspect of this study was the correlation of performance on the TARPA with functioning as measured using the Vineland Adaptive Behaviour Scales (VABS; Sparrow, Cicchetti & Balla, 2005). Five children between the ages 6 and 13 participated. A score on the VABS was attained for each child and the child was then tested with the TARPA across a number of sessions. A Spearman's rank correlation test showed a strong and significant correlation ($\rho = 0.97$; $p = 0.05$) between performance on the TARPA and adaptive functioning as measured using the VABS. Further tests were conducted to examine the relationship between TARPA performance and scoring for each of the sub-scales of the VABS. These showed that the TARPA correlated highly, though just outside significance, with the Communication sub-scale ($\rho = 0.947$; $p = 0.056$), while showing lower, more clearly non-significant correlations with the Daily Living ($\rho = 0.56$; $p = 0.25$) and Socialisation ($\rho = 0.56$; $p = 0.25$) sub-scales. This was consistent with what might be expected from a protocol measuring DRR which RFT sees as a key process underlying language and thus it was considered a good preliminary result for the TARPA.

Moran et al (2010) also showed some differences in performance based on stimulus modality. A number of participants showed weaker responding on sections involving exclusively auditory stimuli than on sections involving at least some visual stimuli, irrespective of the strength of their overall performance. For example, P3 passed both Stage 1 levels that involved visual stimuli only but failed one of the two levels involving only auditory stimuli. In the second stage, he passed Level 1 involving visual stimuli only but failed Level 2 involving only auditory stimuli. In Stages 3 and 4, though he passed some levels involving visual stimuli, he consistently failed levels involving auditory stimuli only. Finally, he failed Stage 5, which, again, involved only auditory stimuli. As regards P4 and P5, both showed their first failure on the auditory only level of Stage 4 and both also failed the critical test for derived relations in Stage 5, which involved only auditory stimuli, having passed at least some of the training sub-components which involve visual as well as auditory stimuli. Furthermore, P5 then went on to pass Stage 6, which, though apparently more difficult in that it involved a transfer of function test as well as a test of combinatorial entailment, happened to involve a combinatorial training and testing procedure that involves only visual stimuli.

Lastly, Moran et al. (2010) also provided some insight into the hierarchical ordering of the TARPA. Ideally, in any multi-stage hierarchically structured test, a participant should not be able to pass a stage of purported higher difficulty having failed a stage of purported lower difficulty. Analysis of the data by visual inspection from individual participants across stages of the TARPA provided in Moran et al. (2010) indicated that in general the TARPA conforms to this criterion.

The current study, which comprised two experiments, extended Moran et al. in several respects. First, the version of the TARPA employed in this study was a more comprehensive and systematic measure. Whereas the Moran et al. protocol was limited with respect to the stimulus modalities involved in the testing of the properties of derived relations (i.e., mutual and combinatorial entailment and transformation of function; for example, there was no training or testing of these properties with exclusively visual stimuli and both the latter properties were tested with auditory stimuli only), the current protocol tested all three properties in each of four different modalities (i.e., all visual, all auditory, and two separate auditory-visual modalities). In addition to this, the current protocol increased the range of tests of simple discrimination and added tests of generalised identity matching in both the visual and auditory modalities.

This protocol also constituted a more systematic measure of DRR. In contrast with the Moran et al protocol, it included tests of maintenance of trained relations in both the mutual and combinatorial entailment levels (with several in the latter) and in addition, the transformation of function test in this protocol was a simpler and logistically more practical assessment. In regard to the latter, in Moran et al., the transformation of function test was adapted from a task employed by Murphy, Barnes-Holmes and Barnes-Holmes (2005) in which participants were assessed for transformation of the discriminative functions of stimuli that could be employed to mand for tokens in the context of a game. This task was relatively complex and its implementation in the context of a computer-delivered protocol increased this complexity. In the current version of the protocol, a very different and considerably simpler test was employed in which participants were initially required to perform particular actions in the presence of particular stimuli and were then tested for transformation of discriminative functions in the presence of stimuli in derived relations with the initial set.

A second way in which this study constituted an extension of Moran et al. is that whereas the latter correlated the TARPA with the VABS, one subscale of which affords an indirect measure of communication ability, the current study correlated the TARPA with a widely accepted performance-based measure of language ability, namely the Preschool Language Scale (4^{th} edition) (PLS-4; Zimmermann, Steiner &

Pond, 2002). The PLS-4 consists of two subscales that assess auditory comprehension/receptive and expressive skills and it also provides standard scores, percentiles, and developmental age scores for the overall scores. The PLS-4 has been used in other behavioural analytic research to assess the language skills of children with ASD (e.g., Delgado & Oblak, 2007) and as such provided a more robust assessment of language skills than the Moran et al. (2010) study. Third, the current study employed the TARPA not just with children with ASD (Experiment 1) but also with typically developing children (Experiment 2). Testing was conducted on the latter to acquire data from an alternative population than children with ASD whose performance might be expected to allow more detailed analysis of the advanced levels of the TARPA, and whose performance level could also be compared with that of the children with ASD.

Fourth, the current study employed a considerably larger number of participants (i.e., n = 10 in Experiment 1 and n = 13 in Experiment 2, giving a total of 23 for the study as a whole) than in the Moran et al study, which included only five participants. This allowed a more extensive analysis of the relationship between performance on the TARPA and language ability because while the correlations reported in Moran et al. (2010) were strong, the small sample size limited the generalisability of these findings. Fifth, in order to investigate the appropriateness of the sequence of TARPA sub-stages, the current study conducted a formal hierarchical analysis of the protocol. As mentioned above, Moran et al. (2010) conducted a more rudimentary analysis of the structure of the protocol based on a visual inspection of the data, the formal analyses conducted in this study involved a method known as order analysis (e.g., Krus, Bart & Airasian, 1975). Order analysis is suitable for analysing descriptive data (e.g., the passing/failing of levels within the

TARPA) in order to identify different hierarchies that exist in a data set. Other behaviour analytic assessments have been subject to similar analyses as part of their development (e.g., Kerr, Meyerson & Flora, 1977) and as such the use of this method in this study represented a critical next step in the development of the TARPA. Finally, in Moran et al (2010), a mouse was used as the input device, which resulted in some procedural difficulties. For example, some participants in Moran et al. (2010) were not able to manipulate a mouse in order to select their responses and the experimenter used the mouse for these participants by making the selection once the child had indicated their response by pointing to the stimulus. This however raised the issue of inadvertent cueing. In the current study, a touch screen computer was employed, which eliminated these difficulties.

Experiment 1

Method

Participants

Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland Galway Research Ethics Committee.

A prior power analysis indicated that a sample of at least six participants was needed to have 80% power for detecting a large sized effect at the traditional alpha value of .05. A large effect size was predicted based on Moran et al. (2010).

Ten children (7 male, 3 female; age range 3 years 4 months - 12 years 5 months; see also Table 1), each of whom had an independent diagnosis of ASD provided by a licensed psychologist, were recruited through a service provider for children with ASD and/or intellectual disability at which the experimenter was employed. Two were attending a specialised school for children with an ID, one was being taught in the context of a home program and the remaining seven were attending a mainstream school at which their learning was facilitated by a special needs assistant. None of the children had previously taken part in research. Consent for conducting the study was obtained from the director of the service provider. Parental consent was obtained for each child who participated and verbal consent was also obtained from each of the participants.

Materials

The two measures employed were the Preschool Language Scales - 4th Edition (PLS-4) and the Training and Assessment of Relational Precursors and Abilities (TARPA).

PLS-4. This is a 130-item standardised test of language that consists of two subscales that assess auditory comprehension (AC) and expressive communication (EC) skills. The AC subscale assesses skills such as attention to speakers, object play, comprehension of basic vocabulary, response to grammatical markers, identification of rhyming words and ability to make comparisons. The EC subscale assesses skills including object naming, object description, expression of quantity and the use of grammatical markers. There are a total of 62 AC and 68 EC items assessed. The PLS-4 yields norm-referenced test scores (standard scores, percentile ranks, and an age equivalent) for both subscales, as well as for total language (TL) score. These scores are available at 3 month intervals from birth to 11 months, and at 6 month intervals for ages 1 year to 6 years, 5 months. Test-retest reliability for the PLS-4 has been reported as ranging between .82 and .95 for the subscale (AC and EC) scores and between .90 and .97 for the total (TL) scale score, while internal reliability has been reported as ranging from .66 to .96 (Zimmermann et al., 2002). In addition, the test has been validated with typically developing children as well as children previously identified as having a variety of disorders including language deficits, developmental delay and ASD.

TARPA. The Training and Assessment of Relational Precursors and Abilities (TARPA) protocol (see Figure 2 for screenshots of various elements) was programmed in Visual Basic ™ and was presented on a touch screen Hewlett Packard ('Pavilion') laptop computer. The TARPA software presented all relevant stimuli and recorded all response data. The protocol consists of 3 stages as follows (1) basic discrimination; (2) conditional discrimination (formally similar stimuli); and (3) conditional discrimination (formally dissimilar stimuli). The first two stages assess the participant's ability to learn skills that are precursors to relational framing, while the third stage, which is the most substantial, tests properties of emergent relational framing itself.

Stage (1). This initial stage tests the participant's ability to learn simple discriminations and has two tracks (visual and auditory) with three levels in each track. Level 1 of the visual track requires a visual discrimination between a blank box and an abstract picture (see Figure 2[iii]). In this, as in all other visual discrimination tasks on all three stages, a correct response involves touching the appropriate on-screen visual stimulus. Level 2 requires a discrimination between two abstract pictures; and Level 3 requires a discrimination between three abstract pictures. The other track features auditory stimuli. In the case of each of the auditory stimuli that is presented throughout the TARPA protocol, there is a corresponding on-screen 'button'. The auditory stimulus is produced when its corresponding button first appears on the screen as well as when its button is touched as a means of selecting that auditory stimulus. In Stage 1, Level 1 of the auditory track requires a discrimination between a spoken nonsense syllable and no sound. Level 2 requires a discrimination between two spoken nonsense syllables and Level 3 requires a discrimination between three spoken nonsense syllables.

Stage (2). This stage tests the participant's ability to learn conditional discriminations in which the sample and correct comparison are formally or physically similar and has two tracks (visual and auditory) with three levels in each track. Level 1 of the visual track involves a 2 choice conditional discrimination involving abstract pictures in which the designated correct comparison on each of the two task trials is identical to the sample. Level 2 is a 3 choice conditional discrimination involving abstract pictures in which the designated correct comparison on each of the three task trials is identical to the sample. Level 3 involves the same task type as Level 2 but tests generalised identity matching by presenting a different set of abstract stimuli on each trial (see Figure 2 [iv]). Level 1 of the auditory track (see Figure 2[v]) involves a 2 choice conditional discrimination involving nonsense syllables in which the designated correct comparison on each of the two task trials is identical to the sample. Level 2 is a 3 choice conditional discrimination involving nonsense syllables in which the designated correct comparison on each of the three task trials is identical to the sample. Level 3 involves the same task type as Level 2 but tests generalised identity matching by presenting a different set of nonsense syllables on each trial.

Stage(3). This stage involves conditional discriminations in which the sample and comparison stimuli are formally dissimilar (see Figure 2[vi] for an example) referred to as arbitrary conditional discriminations) and includes four different stimulus modality tracks: Visual-Visual, Auditory-Visual 1, Auditory-Visual 2 and Auditory-Auditory. In addition each track is further subdivided into levels as follows:

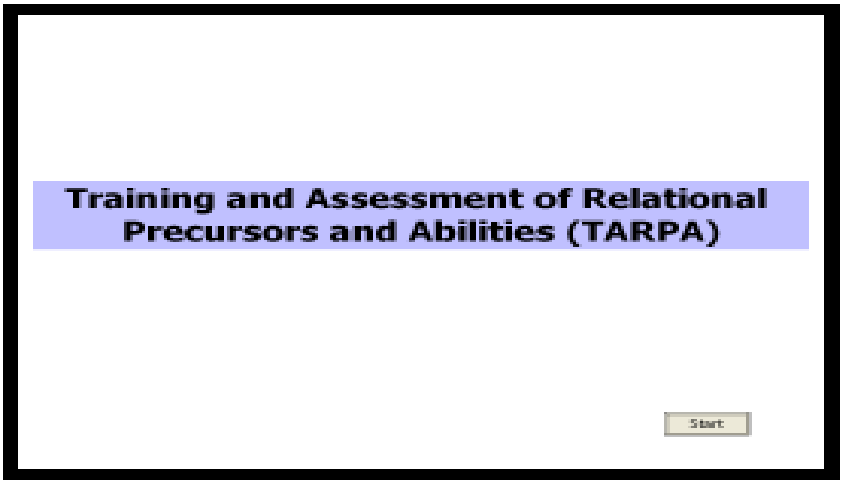
(1) Conditional discrimination (2) Mutual entailment, (3) Combinatorial entailment and (4) Transformation of function.

Stage 3 Level 1 involves a two choice arbitrary conditional discrimination in which the designated correct comparison is physically dissimilar from the sample. In the Visual-Visual track, all stimuli are abstract pictures; in the Auditory-Auditory track, all stimuli are spoken nonsense syllables; in the Auditory-Visual 1 track, the samples are spoken nonsense syllables while the comparisons are abstract pictures; while in the Auditory-Visual 2 track, the samples are abstract pictures while the comparisons are spoken nonsense syllables.

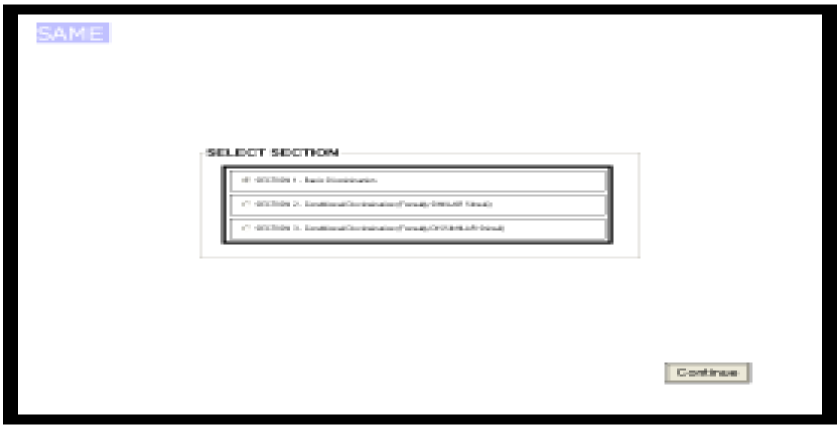
Stage 3 Level 2 tests for mutually entailed arbitrary relational responding and involves three sections. Section 1 (A → B) involves the same conditional discrimination as employed in Level 1. Section 2 (A → B Maintenance) involves the same conditional discrimination as in Section 1 but in the absence of reinforcement. Section 3 (B → A) tests for the mutually entailed counterpart relation to that probed in Sections 1 and 2. In the Visual-Visual and Auditory-Auditory tracks of Section 3, all stimuli are abstract pictures and spoken nonsense syllables respectively; in the Auditory-Visual 1 track, the samples are abstract pictures while the comparisons are spoken nonsense syllables; while in the Auditory-Visual 2 track, the samples are spoken nonsense syllables while the comparisons are abstract pictures.

Stage 3 Level 3 tests for combinatorially entailed relational responding and involves six sections. Section 1 (A → B) involves the same conditional discrimination as employed in Level 1. Section 2 (A → B Maintenance) involves the same conditional discrimination as in Section 1 but in the absence of contingent reinforcement. Section 3 (C → B) involves a new conditional discrimination in which the comparison stimulus is the same as in Section 1 but the sample is a novel stimulus. Section 4 (C → B Maintenance) involves the same conditional discrimination as in Section 3 but in the absence of reinforcement. Section 5 (A → B; C → B) involves an assessment in the absence of reinforcement of both the previously trained and assessed conditional discriminations. Section 6 (A → C; C → A) tests for both the combinatorially entailed relations that can be derived from the previously trained unidirectional relations (i.e., A → B and C → B). In the Visual-Visual and Auditory-Auditory tracks of Section 6, all stimuli are abstract pictures and spoken nonsense syllables respectively; in the Auditory-Visual 1 track, the A stimuli are spoken nonsense syllables while the B and C stimuli are abstract pictures; while in the Auditory-Visual 2 track, the A and C stimuli are spoken nonsense syllables while the B stimuli are abstract pictures.

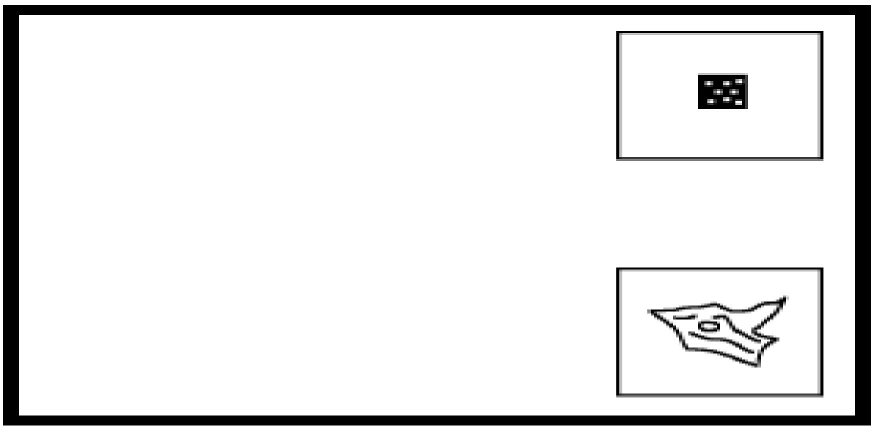
Stage 3 Level 4 tests for transformation of functions via the combinatorially entailed relations (i.e., A → C and C → A) tested for in the previous level and involves three sections. Section 1 tests for maintenance of both the trained (i.e., A → B and C → B) and tested combinatorially entailed (i.e., A → C and C → A) relations from the previous stage. Section 2 trains a discriminative function for each of the A stimuli; namely, waving in the presence of A1 and clapping in the presence of A2. Finally, Section 3 tests whether the C stimuli have acquired novel discriminative functions via derived relations (i.e., waving in the presence of C1 and clapping in the presence of C2).



(i) Title screen

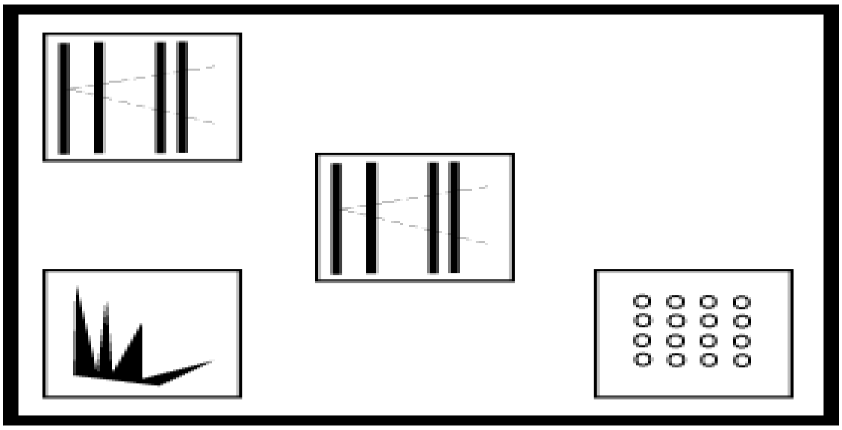


(ii) Stage selection

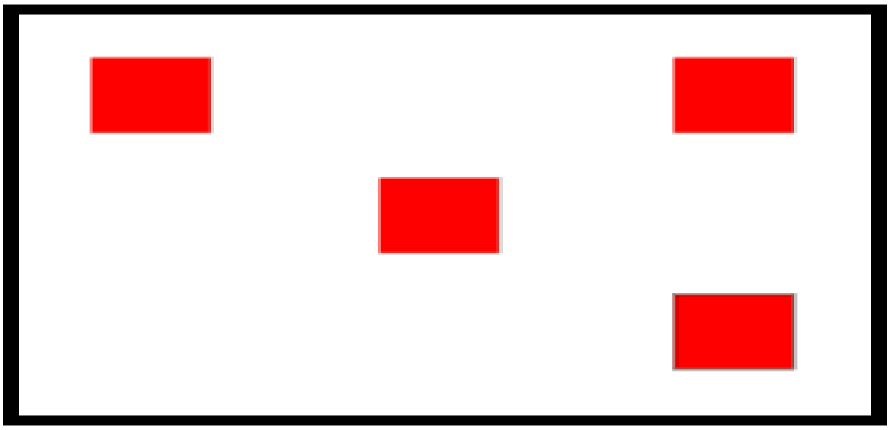


(iii) Two choice discrimination with visual stimuli (Stage 1)

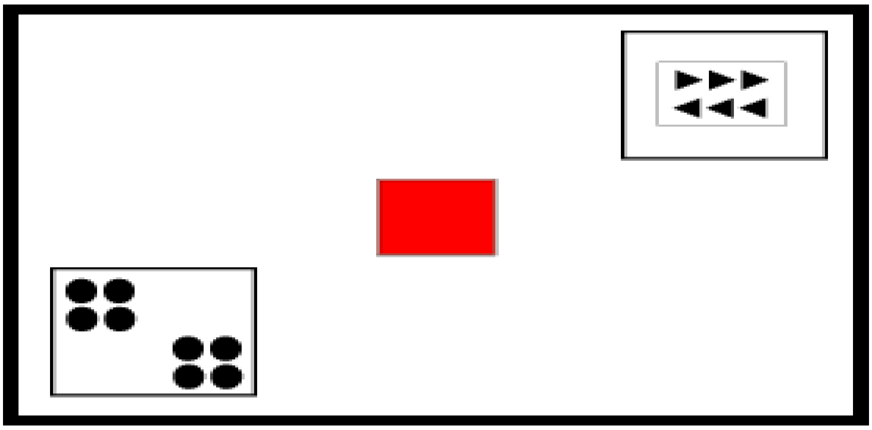
Figure 2. Screenshots of the TARPA.



(iv) Test of generalised identity matching with visual stimuli (Stage 2)

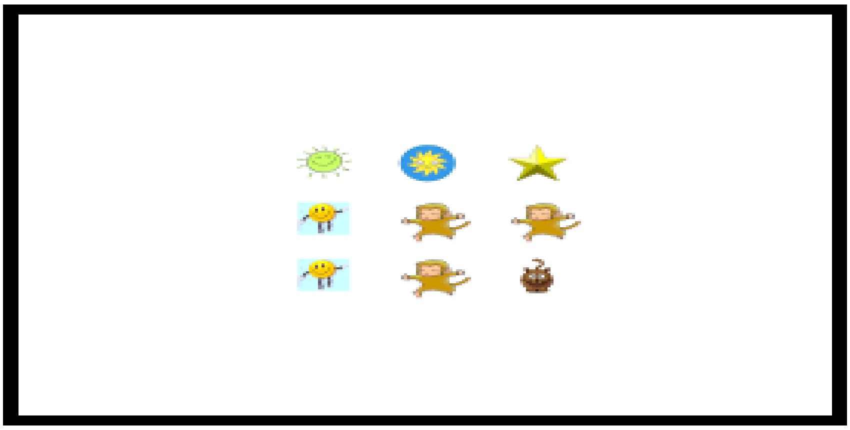


(v) Conditional discrimination with auditory stimuli (Stage 2)



(vi) Arbitrary conditional discrimination with auditory sample and visual comparisons (Stage 3)

Figure 2. (continued).



(vii) Example of reinforcement screen

(viii) On screen data output.

Figure 2. (continued).

Procedure

The first session involved PLS-4 assessment. TARPA assessment began in the next session and continued in subsequent sessions until complete.

PLS-4. This was administered to the child by the experimenter with the assistance of the child's caregiver or teacher (where required). The PLS-4 assessment was conducted in one sitting that lasted on average about 35 minutes per child.

TARPA. Assessment using the TARPA protocol began in the next session. For this the child and the experimenter sat in front of a laptop computer on which TARPA tasks were presented. Each child required a number of assessment sessions. Session lengths were dependent on duration of on-task behaviour, the length of time for which each child was available in that particular session and progress on the TARPA.

Training, Maintenance & Testing. Some parts of the TARPA (e.g., Stage 1, Visual, Level 1) assessed the child's capacity to learn certain skills (i.e., to make a simple discrimination). On these training parts, a particular protocol was followed in which first the correct response was demonstrated by the experimenter, then the correct response on the part of the child was physically guided or prompted and finally the child was assessed for an independent performance (this will hereafter be referred to as the DGI [Demonstration Guidance Independent] protocol). Training trials included a continuous schedule of reinforcement for correct responding involving computer-delivered reinforcement (see Figure 2[vii]) and verbal praise.

Also, all participants were on individualised schedules of reinforcement. More specifically, in the case of each child, the schedule of reinforcement typically used in their normal educational setting was employed; potential reinforcers were chosen based on consultation with the child's caregiver/teacher; and participants were also given the opportunity to choose what they worked for throughout assessment (several participants [P4, P5, P6, P7 & P9] were asked while in the case of others [P1, P2, P3, P8 & P10] choice boards were used). In addition to training parts there were maintenance sections and derived relational testing sections. In the case of both of these, there were no demonstration or guided trials (i.e., the participant received independent trials only) and there was no contingent reinforcement; however, during these sections, children continued to be provided with non-contingent reinforcement including praise.

Response Definitions. Responding during each part (level/section) was scored as either “correct,” “incorrect” or “no response”. A "correct" response was defined as touching the correct on-screen stimulus while refraining from touching an incorrect stimulus. An "incorrect" response was defined as touching an incorrect stimulus or touching both correct and incorrect stimuli. Any other behaviour was scored as "no response" and was counted as incorrect. If the participant performed correctly then they were exposed to further similar trials in that level/section. If a participant failed to perform correctly on any trial then they were re-exposed to the demonstration and guided trials before continued exposure to independent trials. On a demonstration trial the experimenter gave the vocal antecedent ‘look’ before modeling touching the correct stimulus and saying ‘This is the right one’; on a guided trial the experimenter lightly touched the participant’s hand or elbow to guide them to the correct response; while on an independent trial the participant was given a vocal antecedent (e.g., ‘Find the right one’) and had to respond independently within five seconds.

Each level/section had a criterion for number of correct responses needed to pass as well as for cumulative errors designating a fail (see Appendix B). If a participant succeeded in demonstrating a pre-determined number of consecutive correct responses then they were deemed to have passed that level/section and were subsequently exposed to the next section.

Criteria for Passing / Failing. If a participant made a pre-determined cumulative number of incorrect responses then they were deemed to have failed that level/section. Failure at any particular point of the TARPA had consequences that depended on the part of the TARPA concerned (see Materials).

In Stage 1, participants were exposed to all levels of both the visual and auditory tracks. If a participant failed all levels in both tracks, then the assessment was finished at that point. However, if they passed even one level in either track then they were also exposed to all levels of both tracks in Stage 2. With respect to the latter, there were similar guidelines. If a participant failed all levels then the assessment was finished; however, if they passed even just one level in either track, then they were allowed to proceed to Stage 3.

As described earlier, Stage 3 involved four tracks each with a different stimulus modality and the participant was exposed to each of the four tracks in a quasi random order. Each track involved 4 levels, each of which included a certain number of sections. Level 1 (A-B conditional discrimination training) involved only one section. A participant was given two chances to pass. If they failed twice then the assessment of that track came to a finish, while if they passed then they were exposed to Level 2 (mutual entailment training and testing).

Level 2 involved three sections. The participant was given two chances to pass the first section (a review of A-B conditional discrimination training). If they failed then the assessment finished for that track. If they passed then they were exposed to the second section (unreinforced A-B conditional discrimination – 'maintenance'). If they failed the latter then they were re-exposed to training in Section 1 (i.e., with reinforcement) and then given one further exposure to Section 2. If they failed again then the assessment finished for that track. If they passed Section 2 then they were exposed to Section 3 (derived B-A mutual entailment testing). If they passed the test of mutual entailment then they were moved on to Level 3 (combinatorial entailment training and testing), while if they failed then the assessment for that track was finished. Level 3 involved 6 sections. If a participant passed any section then they moved on to the next one. If they failed Section 1 (A – B conditional discrimination training) twice then the assessment for that track finished there. If they failed Section 2 (A-B maintenance) then they received one further exposure to both Sections 1 and 2 and if they failed the latter again then the assessment of that track was finished. If they passed Section 2, then they were exposed to Section 3 (C-B conditional discrimination training). If they failed Section 3 twice then the assessment for that track finished. If they passed Section 3, then they were exposed to Section 4 (C-B maintenance). If they failed the latter, then they received one further exposure to both Sections 3 and 4 and if they failed the latter again then the assessment of that track was finished. If they passed Section 4, then they were exposed to Section 5 (A-B and C-B conditional discrimination training mixed maintenance). If they failed the latter then they were exposed to the previous sections (i.e., 1 to 4) one further time. If they failed again then the assessment for that track was finished. If they passed then they were exposed to Section 6 (derived A-C and C-A combinatorial entailment testing). If they failed the latter then their assessment in that track was finished whereas if they passed then they were exposed to Level 4 (transformation of function).

In Level 4, the final level of assessment for each track, participants were first re-exposed to sections that tested maintenance of the A-B and C-B conditional discriminations as well as of the derived A-C and C-A relations. They were then exposed to discriminative function training. If they failed discriminative function training then the assessment for that track was finished while if they passed then they received transformation of function testing.

It should be noted that all data from participants in both Experiments 1 and 2 were included in a hierarchical analysis. In order to provide additional data for this purpose, a subset of participants from both experiments was tested on all levels of the TARPA, meaning that they continued to be tested even after they had officially failed at a particular level. The participants exposed to this testing included P6 from Experiment 1. However, the results of the hierarchical analysis will not be discussed until the Results section for Experiment 2.

Results & Discussion

Demographic Information and performance parameters for the TARPA (including information on number and approximate length of TARPA sessions in minutes) are shown in Table 1.

Table 1

Demographic Information and TARPA Performance Parameters for Experiment 1

Pt. No.	Age (Y: M)	Sex	No. Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested
1	07:04	m	2	15	30	N/A	S1L3 (A&V)
2	04:07	f	6	15	90	S1L2 (V)	S2L3 (A&V)
3	06:01	m	8	12	96	S2L1 (V)	S3L1 (CD)
4	04:09	m	8	20	160	S3L1 (CD)	S3L2 (ME)
5	05:10	m	5	40	200	S3L2 (ME)	S3L3 (CE)
6	07:11	m	3	20-25	70	S3L4 (TOF)	S3L4 (TOF)
7	12:05	f	4	30	120	S3L4 (TOF)	S3L4 (TOF)
8	03:04	m	10	20	200	S2L2 (V)	S3L1
9	08:07	f	3	50	150	S3L4 (TOF)	S3L4 (TOF)
10	08:08	m	6	35	210	S3L1 (CD)	S3L2 (ME)

Note. A = Auditory; V = Visual; CD = Conditional Discrimination. ME = Mutual Entailment; TOF = Transformation of Functions.

Scores for the TARPA and the PLS-4 are shown in Table 2. Participants had to pass a level in order to receive all the points corresponding to that level. Scores for the TARPA were calculated as per the scoring table provided in Appendix B. Children who passed a level received all the points corresponding to that level. Scores (raw, standard and age equivalent) for the PLS-4 were calculated based on criteria outlined in the official PLS-4 manual. All correlations are based on raw (as opposed to age normative) PLS-4 scores. A Spearman's rank order correlation showed a highly significant correlation between the TARPA scores and the PLS-4 total scores ρ (8) = 0.99 (p < .01). Tests were also conducted to examine the relationship between TARPA performance and that of each of the two PLS-4 subscales. These revealed highly significant correlations with both Auditory Comprehension ρ (8) = 0.94 (p < .01) and Expressive Communication ρ (8) = 0.999 (p < .01).

Table 2

TARPA and PLS4 Scores for Experiment 1

Pt. Number	TARPA Score	PLS4 Raw AC	PLS4 Raw EC	PLS4 Total	PLS4 SS	PLS4 LAE
1	0	19	11	30	N/A	00:09
2	0.66	19	19	38	55	01:01
3	2.16	25	23	48	55	01:05
4	6.32	32	34	66	55	02:01
5	23	52	54	106	55	04:00
6	45	62	66	128	N/A	06:04 – 06:05
7	59	62	68	130	N/A	> 06:05
8	2.65	32	25	57	55	01:09
9	44	62	65	127	N/A	06:01- 06:03
10	12.15	31	35	66	N/A	02:01

Note. AC = Auditory Comprehension; EC = Expressive Communication; SS = Standardised Score; LAE = Language Age Equivalent.

Individual TARPA performance

A detailed breakdown of individual performances on the TARPA is shown (in Table 3). There was wide variation in performance across participants. P1 failed all levels in both tracks of Stage 1 (simple discrimination) and did not proceed any further. P2 passed the first two visual levels of Stage 1 and failed all subsequent levels to which he was exposed including all levels of the auditory track in Stage 1 and all levels of both tracks in Stage 2. P3 and P8 each passed levels in both Stages 1 and 2 but failed all initial levels of Stage 3. P3 passed the first two visual levels of

Stage 1 but failed the third level. The latter failure appeared to be an attentional issue, and he was therefore retested using a new set of stimuli in the context of which he passed. He engaged in off-task responding while being tested for the first of the auditory levels of Stage 1 and was therefore again retested with a new set of stimuli and passed. He also passed the second of the auditory levels but then failed the third. He passed Level 1 of the visual track of Stage 2 but failed all subsequent levels of the visual track and all levels of the auditory track in this stage. He also failed the first level of each of the four tracks of Stage 3. P8 passed all levels in both tracks of Stage 1 with the exception of the final auditory level. In Stage 2, he passed the first two visual track levels and failed everything else including all three auditory levels.

All the remaining participants passed at least some levels of Stage 3 but there was also significant variation within this stage. P4 and P10 passed one or more tracks at Level 1 (arbitrary conditional discrimination) but failed to move beyond this level. P5 passed mutual entailment in two tracks (visual-visual and auditory visual 2) but failed to show combinatorial entailment in any track. P6 and P9 both demonstrated at least combinatorial entailment in two tracks. P6 showed combinatorial entailment and transformation of function in the visual-visual track and combinatorial entailment alone in the auditory visual 1 track while P9 showed combinatorial entailment and transformation of function in the visual-visual and auditory visual 2 tracks. Finally, P7 showed transformation of function for all tracks except auditory visual 2.

Table 3

Performances of Individual Participants for each TARPA Stage and Level in Experiment 1

Stage 1: Simple discrimination

Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Vis. L1	F	P	P	P	P	P	P	P	P	P
Vis. L2	F	P	P	P	F**P**	P	P	P	P	P
Vis. L3	F	F	F**P**	P	P	P	P	P	P	P
Aud. L1	F	F	F**P**	P	P	P	P	P	P	P
Aud. L2	F	F	P	F	P	P	P	P	P	P
Aud. L3	F	F	F	F	P	P	P	F	P	F

Stage 2: Non Arbitrary conditional discrimination (i.e., formally SIMILAR stimuli)

Vis. L1	n/a	F	P	P	P	P	P	F**P**	P	P
Vis. L2	n/a	F	F	P	P	P	P	P	P	P
Vis. L3	n/a	F	F	P	P	P	P	F	P	F
Aud. L1	n/a	F	F	F	P	P	P	F	P	P
Aud. L2	n/a	F	F	F	P	P	P	F	P	F
Aud. L3	n/a	F	F	F	P	P	P	F	P	F

Note. F followed by bold P indicates that the pt. showed serious attention deficits during the first test and passed after retesting with new stimuli.

Table 3. Continued.

Performances of Individual Participants for Each TARPA Stage and Level in Experiment 1

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Visual-Visual										
L1 CD	n/a	n/a	F	P	P	P	P	F	P	P
L2 ME	n/a	n/a	n/a	F	P	P	P	n/a	P	F
L3 CE	n/a	n/a	n/a	n/a	F	P	P	n/a	P	n/a
L4 TOF	n/a	n/a	n/a	n/a	n/a	P	P	n/a	P	n/a
Auditory-Visual 1										
L1 CD	n/a	n/a	F	F	P	P	P	F	P	F
L2 ME	n/a	n/a	n/a	n/a	F	P	P	n/a	F	n/a
L3 CE	n/a	n/a	n/a	n/a	n/a	P	P	n/a	n/a	n/a
L4 TOF	n/a	n/a	n/a	n/a	n/a	F	P	n/a	n/a	n/a

Note. CD = Conditional discrimination; ME = Mutual entailment; CE = Combinatorial entailment; TOF = Transfer of function.

Table 3. Continued.

Performances of Individual Participants for Each TARPA Stage and Level in Experiment 1

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Auditory-Visual 2										
L1 CD	n/a	n/a	F	F	P	P	P	F	P	P
L2 ME	n/a	n/a	n/a	n/a	P	P	P	n/a	P	F
L3 CE	n/a	n/a	n/a	n/a	F	F	P	n/a	P	n/a
L4 TOF	n/a	n/a	n/a	n/a	n/a	n/a	F	n/a	P	n/a
Auditory-Auditory										
L1 CD	n/a	n/a	F	F	F	P	P	F	P	P
L2 ME	n/a	n/a	n/a	n/a	n/a	P	P	n/a	P	F
L3 CE	n/a	n/a	n/a	n/a	n/a	F	P	n/a	F	n/a
L4 TOF	n/a	n/a	n/a	n/a	n/a	n/a	P	n/a	n/a	n/a

Note. CD = Conditional discrimination; ME = Mutual entailment; CE = Combinatorial entailment; TOF = Transfer of function.

An additional analysis was performed with respect to the hierarchical ordering of TARPA sub-stages. In a multi-stage hierarchically structured test, a participant should not be able to succeed at an ostensibly more difficult level after failing on an ostensibly less difficult one. The data reported by Moran et al. suggested that the initial TARPA protocol on which they conducted their analysis broadly conformed to this criterion. In order to perform a similar analysis on the data from the current TARPA, the scores obtained by individual participants on each of the three stages were converted into percentages of correct responding; for example, if a child was correct on all levels of a stage then they scored 100% correct, while if they were correct on only half then they scored 50% correct. Comparison of these percentages showed that none of the ten participants achieved a greater percentage on a more advanced than on a less advanced stage, thus supporting the hierarchical structuring of the TARPA.

Summary

The main pattern of results for Experiment 1 was the finding of a strong and significant correlation between the TARPA and the PLS-4 as well as between the TARPA and each of the two PLS-4 subscales (i.e., Auditory Comprehension and Expressive Communication) in children with ASD. Furthermore, as just indicated, this pattern of results was found with a more comprehensive protocol (involving most importantly a greater range of modalities in the tests for each of the derived relational properties) and with a larger sample of children. Finally, there was also evidence of the hierarchical structuring of the TARPA.

Experiment 2

The aim of Experiment 2 was to extend the findings of prior TARPA research including Experiment 1 in two potentially important respects. First, it employed typically developing children. Testing was conducted on the latter to acquire data from an alternative population than children with ASD. The performance of the former might be expected to allow more detailed analysis of the upper portion of the TARPA, and would also allow for comparison with the ASD group. As suggested in Chapter 1, typically developing children readily learn DRR through exposure to the natural language environment (e.g., Lipkens, Hayes & Hayes, 1993; Luciano, Gomez-Becerra, & Rodriguez-Valverde, 2007) while children with ASD, amongst other categories of developmental delay, do not easily learn these patterns (e.g., Rehfeldt, Dillen, Ziomek, & Kowalchuk, 2007). Hence, it might be expected that typically developing children would outperform children with ASD on the TARPA. The findings of Experiment 2 might also allow for comparison with respect to other dimensions such as performance across different stimulus modalities, for example.

This second experiment also added an additional 12 participants to the overall sample tested, giving a total of 22 for the study as a whole, thus increasing the power of a potential correlational analysis. This larger number also facilitated a more formal order analysis of the protocol (see e.g., Krus, Bart & Airasian, 1975) than had previously been possible.

Order analyses have been used in the case of previous behaviour analytic assessments. For example, Kerr, Meyerson and Flora (1977) used an order analysis to demonstrate that levels employed in the Assessment of Basic Learning Abilities (ABLA) were hierarchically ordered. More specifically, they showed a general trend (111 out of a total of 117 participants) in which failure at any ABLA level predicted failure at higher levels, while passing a level only occurred if all lower levels had been passed. This finding has since been replicated in other studies (see Martin and Yu, 2000 for a review) and was also found to hold for typically developing children (Casey & Kerr, 1977). The order analysis conducted in the current study, which included a focus on the levels of Stage 3 in particular, was made possible by combining participants from both experiments. This afforded a sufficiently large sample size to allow a meaningful comparison between levels.

Method

Participants

Thirteen typically developing children (6 male, 7 female; age range 2 years 11 months - 6 years 4 months; see also Table 4) were recruited through a primary school (St. John's National School, Breaffy, Castlebar, Co. Mayo, Ireland) and a playschool attached to the school. Consent for conducting the study was obtained from the school principal for the primary school. Parental consent was obtained for each participant. Verbal consent was also obtained from each of the children themselves. None of the children had taken part in research before. Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland Galway Research Ethics Committee. One participant (P22) dropped out of the study before completing TARPA assessment. Though the limited data provided by this participant could not be included in any of the correlational analyses, they were used in the order analysis.

Materials

See Experiment 1.

Procedure

The PLS-4 and the TARPA were administered in the same manner as in Experiment 1.

Results & Discussion

Table 4 provides demographic information and details on TARPA performance including information on number and approximate length of sessions (in minutes).

Table 4

Demographic Information and TARPA Performance Parameters for Experiment 2

Pt. No.	Age (Y: M)	Sex	No. Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested
11	04:00	m	8	25	200	S3L2 (ME)	S3L3 (CE)
12	03:02	m	10	20	200	S3L1 (CD)	S3L2 (ME)
13	02:11	m	8	20	160	S1L3 (A&V)	S2 (A&V)
14	04:11	f	4	25	100	S3L2 (ME)	S3L3 (CE)
15	04:10	f	3	35	105	S3L3 (CE)	S3L4 (TOF)
16	06:01	f	4	35	140	S3L4 (TOF)	S3L4 (TOF)
17	05:10	m	4	15-20	60	S3L4 (TOF)	S3L4 (TOF)
18	05:01	f	2	30-45	60	S3L4 (TOF)	S3L4 (TOF)
19	06:04	m	2	25-30	50	S3L4 (TOF)	S3L4 (TOF)
20	05:08	m	3	30	90	S3L4 (TOF)	S3L4 (TOF)
21	05:04	f	4	20-25	80	S3L2 (ME)	S3L3 (CE)
22	04:10	f	n/a	n/a	n/a	n/a	n/a
23	05:02	f	5	20	100	S3L2 (ME)	S3L3 (CE)

Note. A = Auditory; V = Visual; CD = Conditional Discrimination; ME = Mutual Entailment; TOF = Transformation of Functions.

Table 5 shows each participant's score on the PLS-4 (raw auditory comprehension, raw expressive communication, total raw score, standard and age equivalent scores) and score on the TARPA. All correlations are based on raw (as opposed to age normative) PLS-4 scores. A Spearman's rank order correlation test was used to examine the relationship between performance on the TARPA and language. A highly statistically significant correlation was found ($\rho = 0.94$, $p < .01$) between total TARPA score and total PLS-4 score. Further tests were conducted to examine the relationship between TARPA performance and scoring on the subsections of the PLS 4. Findings showed significant correlations between the TARPA and auditory comprehension ($\rho = 0.98$, $p < .01$) and between the TARPA and expressive communication ($\rho = 0.89$, $p < .01$). The auditory subsection correlated more strongly with the TARPA than the Expressive subsection.

Table 5

TARPA and PLS4 Scores for Experiment 2

Pt. Number	TARPA Score	PLS4 Raw AC	PLS4 Raw EC	PLS4 Total	PLS4 SS	PLS4 LAE
11	18	49	52	101	90	03:08
12	11.65	39	42	81	87	02:08
13	2	35	40	75	91	02:11
14	22	55	57	112	96	04:05
15	30	59	63	122	116	05:04 - 05:05
16	45	61	66	127	99	06:01- 06:03
17	44	60	66	126	107	05:11 - 06:00
18	44	58	62	120	94	05:02
19	48	61	66	127	99	06:01 - 06:03
20	48	61	65	126	106	05:11 - 6:00
21	22	55	61	116	86	04:09
22	*	55	57	112	96	04:05
23	26	57	65	122	100	05:04 - 05:05

Note. AC = Auditory Comprehension; EC = Expressive Communication; SS = Standardized Score; LAE = Language Age Equivalent.

Individual TARPA performance

A detailed breakdown of individual performances on the TARPA is provided (see Table 6). All participants were typically developing children, but there was some variation in age and the effect of this variation is visible in the table. The youngest participant (P13), who was just under 3 years of age, passed all levels (both visual and auditory) of Stage 1 but then failed all subsequent levels to which he was exposed, including all those in Stage 2 and the first (arbitrary conditional discrimination) level of each of the tracks on Stage 3. P12 was three years and 2 months old. Though failing the third level (3 comparison discrimination of the auditory track of Stage 1 (and in fact this was the only failure in Stage 1 of any child in Experiment 2) he passed Stage 2 Levels 1 and 2 (2 and 3 comparison conditional discrimination with all visual stimuli) and also passed Level 1 (2 comparison arbitrary conditional discrimination) of the auditory-visual and auditory-auditory tracks in Stage 3. P11, who was exactly 4 years old, passed all levels in both Stages 1 and 2, except for Level 3 (non-arbitrary relational responding or identity matching) of the visual track of the latter. He passed the first level on each of the tracks of Stage 3 except for the visual track and also passed the test for mutual entailment in the auditory visual track.

All of the remaining participants passed all levels of Stages 1 and 2. Three of the participants (P15, P18 and P23) were aged approximately 5. These children all passed mutual entailment in at least one track but none passed it in all four tracks, while with regard to combinatorial entailment there was some variation in that P23 failed to show this pattern of derivation at all, P15 showed it in just one track, and P18 showed it in two. The three oldest participants in the group were aged around 6 years. These children all passed mutual entailment in all cases except one (i.e., P17 on AV1) and each of them also passed at least two tests of combinatorial entailment. The last two children, P20 and P21 were between ages 5 and 6. P20 was 5 years and 8 months while P21 was 5 years and 4 months. Despite the fact that these two children were only four months apart in age, however, their performances on both the TARPA and PLS-4 showed a much wider divergence. P20 showed a TARPA performance similar to that of the 6 year olds (showing combinatorial entailment on all stages except auditory-auditory) while also achieving an age equivalent score of 6 on the PLS-4 , while P21 showed a TARPA score similar to the lower scoring five year olds, and achieved an age equivalence PLS-4 score of 4 years and nine months.

Table 6

Performances of Individual Participants for Each TARPA Stage and Level in Experiment 2

Stage 1: Simple discrimination

Part	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
Vis. L1	P	P	P	P	P	P	P	P	P	P	P	P	P
Vis. L2	P	P	P	P	P	P	P	P	P	P	P	P	P
Vis. L3	P	P	P	P	P	P	P	P	P	P	P	P	P
Aud. L1	P	P	P	P	P	P	P	P	P	P	P	P	P
Aud. L2	P	P	P	P	P	P	P	P	P	P	P	P	P
Aud. L3	P	F	P	P	P	P	P	P	P	P	P	P	P

Stage 2: Non Arbitrary conditional discrimination (i.e., formally SIMILAR stimuli)

Vis. L1	P	P	F	P	P	P	P	P	P	P	P	P	P
Vis. L2	P	P	F	P	P	P	P	P	P	P	P	P	P
Vis. L3	F	F	F	P	P	P	P	P	P	P	P	P	P
Aud. L1	P	F	F	P	P	P	P	P	P	P	P	P	P
Aud. L2	P	F	F	P	P	P	P	P	P	P	P	P	P
Aud. L3	P	F	F	P	P	P	P	P	P	P	P	P	P

Table 6.Continued

Performances of Individual Participants for Each TARPA Stage and Level in Experiment 2

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)													
Part	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
Visual-Visual													
L1 CD	F	P	F*	P	P	P	P	P	P	P	P	P	P
L2 ME	F*	n/a	n/a	F	P	P	P	P	P	P	F	P	P
L3 CE	n/a	n/a	n/a	F*	F	P	P	F	P	P	n/a	F	F
L4 TOF	n/a	n/a	n/a	F*	F*	P	P	P*	P	F	n/a	F*	F*
Auditory-Visual 1													
L1 CD	P	F	F*	P	P	P	P	P	P	P	P	DO	P
L2 ME	F	n/a	n/a	F	P	P	F	P	P	P	F	n/a	F
L3 CE	n/a	n/a	n/a	F*	P	F	F*	P	P	P	F*	n/a	F
L4 TOF	n/a	n/a	n/a	n/a	F	P*	F*	P	P	P	F*	n/a	P*

Note. *indicates additional testing for the hierarchy. DO denotes dropout from testing

Table 6.Continued

Performances of Individual Participants for Each TARPA Stage and Level in Experiment 2

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)													
Part	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
Auditory-Visual 2													
L1 CD	P	P	F*	P	P	P	P	P	P	P	P	DO	P
L2 ME	P	F	n/a	P	F	P	P	P	P	P	F	n/a	P
L3 CE	F	n/a	n/a	F	n/a	F	P	P	F	P	F*	n/a	F
L4 TOF	n/a	n/a	n/a	F*	n/a	n/a	P	P	F*	P	F*	n/a	P*
Auditory-Auditory													
L1 CD	P	P	F*	P	P	P	P	P	P	P	P	DO	P
L2 ME	F	F	n/a	F	F	P	P	F	P	F	P	n/a	F
L3 CE	n/a	n/a	n/a	F*	n/a	P	F	n/a	F	F*	F	n/a	F*
L4 TOF	n/a	n/a	n/a	F*	n/a	F	F*	n/a	F*	F*	P*	n/a	F*

Note. *indicates additional testing for the hierarchy. DO denotes dropout from testing.

As in Experiment 1, an additional analysis was performed with respect to the hierarchical ordering of TARPA sub-stages. Once again, the scores obtained by individual participants on each of the three stages were converted into percentages of correct responding. Comparison of these percentages showed that none of the twelve participants achieved a greater percentage on a more advanced than on a less advanced stage, thus once again supporting the hierarchical structuring of the TARPA.

Combined Correlational Analysis

Data from Experiments 1 and 2 were combined for purposes of further correlational analysis. A Spearman's rank order test was used to examine the correlation between TARPA and PLS-4 performance for 22 participants across both experiments. A statistically significant correlation was found ($\rho = 0.96$, $p < .001$) between total TARPA and total raw PLS-4 score. Further tests were conducted to examine the relationship between TARPA and PLS-4 subsections (raw score) performance. Findings showed significant correlations between the TARPA and both auditory comprehension ($\rho = 0.95$, $p < .001$) and expressive communication ($\rho = 0.95$, $p < .001$).

Order Analysis

Order analysis (Kerr et al., 1977; Krus, Bart, & Airasian, 1977, Marion et al., 2003) was used to evaluate the hierarchical structure of the TARPA. Table 7 compares each of the ten TARPA levels with each of the other levels and in the case of each comparison shows (i) number of confirmations [C] and disconfirmations [D] for those two levels, where a confirmation is a pattern in which the lower level was passed and the higher level was failed while a disconfirmation is a pattern in which the higher level was passed and the lower level was failed; (ii) a z-score measurement for that pattern; and (iii) a p-value corresponding to that z-score (see Table7).

Table 7

Order analysis of the TARPA including all stages and levels based on the data provided by the 23 participants across Experiments 1 and 2

	Combination across Tracks																	
Stage/Level	Stage 1 Vis & Aud L2		Stage 1 Vis & Aud L3		Stage 2 Vis & Aud L1		Stage 2 Vis & Aud L2		Stage 2 Vis & Aud L3		Stage 3. L1. Cond. Dis		Stage 3. L2. Mut. Entail.		Stage 3. L3. Comb. Entail		Stage 3. L4. TOF	
	C	D	C	D	C	D	C	D	C	D	C	D	C	D	C	D	C	D
Stage 1 L1	2	0	7	0	7	0	8	0	11	0	13	0	11	0	12	0	12	0
Vis & Aud	Z = 1.41		Z = 2.64**		Z = 2.65**		Z = 2.82**		Z = 3.31***		Z = 3.60***		Z = 3.32***		Z = 3.4***		Z = 3.4***	
Stage 1 L2	-		6	0	6	0	7	0	10	0	13	0	9	0	12	0	12	0
Vis & Aud			Z = 2.45**		Z = 2.45**		Z = 2.64**		Z = 3.2***		Z = 3.60***		Z = 3**		Z = 3.4***		Z = 3.4***	
Stage 1 L3	-		-		2	1	3	0	8	0	7	1	8	0	12	0	12	0
Vis & Aud					Z = 0.58		Z = 1.73*		Z = 2.82**		Z = 2.12*		Z = 2.82***		Z = 3.4***		Z = 3.4***	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * p < .05. ** p < .01. *** p < .001.

Table 7. Continued.

	Combination across Tracks																	
Stage/Level	Stage 1 Vis & Aud L2		Stage 1 Vis & Aud L3		Stage 2 Vis & Aud L1		Stage 2 Vis & Aud L2		Stage 2 Vis & Aud L3		Stage 3. L1. Cond. Dis		Stage 3. L2. Mut. Entail.		Stage 3. L3. Comb. Entail C		Stage 3. L4. TOF	
	C	D	C	D	C	D	C	D	C	D	C	D	C	D		D	C	D
Stage 2 L1	-		-		-		2	0	6	0	8	1	8	0	12	0	12	0
Vis & Aud							Z = 1.41		Z = 2.449**		Z = 2.33***		Z = 2.82**		Z = 3.4***		Z = 3.4***	
Stage 2 L2	-		-		-		-		4	0	6	0	8	0	12	0	12	0
Vis & Aud									Z = 2*		Z = 2.45**		Z = 2.82**		Z = 3.4***		Z = 3.4***	
Stage 2 L3	-		-		-		-		-		1	0	6	0	12	0	12	0
Vis & Aud											Z = 1		Z = 2.45**		Z = 3.4***		Z = 3.4***	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value.
* p < .05. ** p < .01. *** p < .001.

Table 7. Continued.

	Combination across Tracks																	
Stage/Level	Stage 1 Vis & Aud L2		Stage 1 Vis & Aud L3		Stage 2 Vis & Aud L1		Stage 2 Vis & Aud L2		Stage 2 Vis & Aud L3		Stage 3. L1. Cond. Dis		Stage 3. L2. Mut. Entail.		Stage 3. L3. Comb. Entail		Stage 3. L4. TOF	
	C	D	C	D	C	D	C	D	C	D	C	D	C	D	C	D	C	D
Stage 3. L1. Cond. Dis (A-B)	-		-		-		-		-		-		18	0	26	0	21	0
													Z = 4.24***		Z = 5.09***		Z = 4.58***	
Stage 3. L2. Mutual Entailment	-		-		-		-		-		-		-		17	0	14	2
															Z = 4.12***		Z = 3.00**	
Stage 3. L3. Comb. Entailment	-		-		-		-		-		-		-		-		6	7
																	Z = -0.27	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value.
* p < .05. ** p < .01. *** p < .001.

As can be seen, in the majority of the cells of the table (i.e., 44/45) there were more confirmations than disconfirmations of the predicted order. In addition, in most cells, there is a statistically significant effect in favour of the predicted hierarchy for the comparison between number of confirmations and number of disconfirmations. In the case of all but one of the cells where there is not a significant effect, the trend is in the predicted direction but the number of recorded comparisons is low and this is likely the factor that is mainly responsible for the lack of significance. There are four cells in this category and these include the comparisons between S1L1 and S1L2; S1L3 and S2L1; S2L1 and S2L2; and S2L3 and S3L1. For each of these, the question arises as to whether further data might indicate a reversal of the predicted hierarchical order. In the case of the first of these (S1L1 versus S1L2), it seems unlikely, because it would seem that two stimuli, even readily discriminable ones, are unlikely to be more discriminable than one stimulus versus the absence of a stimulus. Similarly in the case of the third comparison (S2L1 versus S2L2), it would also seem unlikely, since the only difference in this case is that the higher level includes more comparisons. In the case of the second comparison (S1L3 versus S2L1), it is arguable that further empirical work investigating the level of relative difficulty is warranted. The current analysis is based on the amalgamation in the cases of Stages 1 and 2 of visual and auditory level data. This particular disconfirmation involves one participant with ASD who failed the most difficult auditory level in Stage 1 and passed the least difficult visual level in Stage 2.

A more in-depth empirical investigation might thus focus on the difference between auditory and visual stimuli, which may perhaps be most pronounced for those with ASD. This is a theme to which we will return in the general discussion. In the case of the fourth comparison (S2L3 versus S3L1), it is arguable that further empirical work is also warranted in this case. Despite the fact that S2L3 involves responding in accordance with non-arbitrary relations and S3L1 involves learning an arbitrary relation, the task in S2L3 involved three comparisons and required choosing correctly on 5 novel trial types whereas the task in S3L1 involved two comparisons and the same trial type throughout.

In the case of one of the 45 data cells in Table 7, there are more disconfirmations than confirmations and thus in this case the suggested hierarchy seems at odds with the found pattern. This comparison is S3L3 (combinatorial entailment) versus S3L4 (transformation of function). Of course, it should not be the case that participants can pass the latter after failing the former. Perhaps one reason that this test (which required the performance of particular actions by the participant dependent on the on-screen stimulus) might have been susceptible to a 'false positive' (i.e., a participant passing for reasons other than showing genuine transformation of function) is that it involved more direct interaction with the experimenter than other tasks and might therefore have been more susceptible to inadvertent cuing by the latter. This is certainly an issue of which those testing at this level of the protocol should be aware.

Another possibility is that some participants simply responded consistently across the task, whether correct or incorrect and that, in the event that their initial answer was correct then this consistency resulted in an ostensible pass. In order to minimise the possibility of this occurring in the future, further testing might expose participants to more than one transformation of function test. It should be noted that, even though the current analysis does not show a similar weakness with respect to order in other sections of the TARPA, the possibility that participants might have passed other levels in a similar way cannot be discounted. This possibility is made more probable by the fact that in almost all sections of the TARPA, the testing tasks involve just two comparisons. The decision to utilise two choice tests throughout the protocol was taken in order to minimise training times and thus make the TARPA assessment as brief and efficient as possible; nevertheless, this can increase the possibility of false positive results achieved on the basis of ‘consistently maintained correct guessing’. Of course, given multiple stages and levels across the protocol, it is unlikely that the latter might allow a participant to attain a much higher score than they would otherwise attain, but the possibility of ‘consistently maintained correct guessing’ as the basis for even one false positive is certainly something that needs to be taken into account with respect to the future development of the protocol. For example, one way to make such a possibility less likely particularly on more important levels such as those testing derivation might be to require correct performance with more than one stimulus set.

A hierarchical analysis for each of the 4 tracks in Stage 3 is shown (see Table 8). Once again, it shows number of confirmations [C] and disconfirmations [D]; (ii) a z-score measurement for that pattern; and (iii) a p-value corresponding to that z-score. A similar pattern for the TARPA as a whole is apparent. More specifically, in most cells there are no disconfirmations and/or there is a statistically significant difference between the number of confirmations and disconfirmations, which supports the hierarchy. The pattern with respect to significance is weaker than for Stage 3 in Table 7 but this is because there are less exemplars per track than across tracks. Once again, the predicted hierarchical pattern appears to hold for most cells of the table with the main deviation in respect of this pattern appearing for the comparison between S3L3 (combinatorial entailment) and S3L4 (transformation of function).

Table 8

Order analysis of TARPA Stage 3 for each of the four tracks based on the data provided by the 23 participants across Experiments 1 and 2

Visual-Visual

Levels	Stage 3. L2. Mutual Entailment		Stage 3. L3. Comb. Entailment		Stage 3. L4. Transf. Stim. Function	
	C	D	C	D	C	D
Stage 3. L1. Cond. Dis (A-B)	3	0	6	0	6	0
	Z = 1.73*		Z = 2.4**		Z = 2.44**	
Stage 3. L2. Mutual Entailment	-		5	0	4	0
			Z = 2.23*		Z = 2*	
Stage 3. L3. Comb. Entailment	-		-		2	1
					Z = 0.58	

Auditory-Visual 1

Levels	Stage 3. L2. Mutual Entailment		Stage 3. L3. Comb. Entailment		Stage 3. L4. Transf. Stim. Function	
	C	D	C	D	C	D
Stage 3. L1. Cond. Dis (A-B)	7	0	5	0	4	0
	Z = 2.65**		Z = 2.23**		Z = 2*	
Stage 3. L2. Mutual Entailment	-		1	0	2	1
			Z = 1		Z = 0.58	
Stage 3. L3. Comb. Entailment	-		-		2	2
					Z = 0	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * p < .05. ** p < .01. *** p < .001.

Table 8. Continued.

Auditory-Visual 2

Levels	Stage 3. L2. Mutual Entailment C	D	Stage 3. L3. Comb. Entailment C	D	Stage 3. L4. Transf. Stim. Function C	D
Stage 3. L1. Cond. Dis (A-B)	2	0	8	0	5	0
	Z = 1.41		Z = 2.82**		Z = 2.24*	
Stage 3. L2. Mutual Entailment	-		7	0	4	0
			Z = 2.65**		Z = 2*	
Stage 3. L3. Comb. Entailment	-		-		1	1
					Z = 0	

Auditory-Auditory

Levels	Stage 3. L2. Mutual Entailment C	D	Stage 3. L3. Comb. Entailment C	D	Stage 3. L4. Transf. Stim. Function C	D
Stage 3. L1. Cond. Dis (A-B)	6	0	7	0	6	0
	Z = 2.45**		Z = 2.65**		Z = 2.45**	
Stage 3. L2. Mutual Entailment	-		4	0	4	1
			Z = 2*		Z = 1.34	
Stage 3. L3. Comb. Entailment	-		-		1	2
					Z = - 0.57	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * $p < .05$. ** $p < .01$. *** $p < .001$.

General Discussion

The TARPA protocol was developed as a means of assessing the emergence of DRR or relational framing, which has been argued by RFT theorists to be the key process underlying language. Moran et al. (2010) showed a correlation between performance on an initial version of the TARPA and adaptive behaviour as measured by the Vineland Adaptive Behavior Scales (Sparrow, et al., 2005) in several children with ASD. Importantly the TARPA also showed a borderline correlation specifically with the communication subscale of the VABS. This result supported the RFT approach to language as DRR and the potential utility of this approach for predicting and remediating language for children with ASD, and, more specifically, suggested the potential of the TARPA as an RFT-based protocol for assessing DRR.

The current study was similar to Moran et al (2010) in that it correlated a version of the TARPA with a potentially important alternative measure of behaviour. However, it extended Moran et al. in several key respects. First, this study used a more comprehensive and systematic TARPA protocol; second, it compared this protocol with a mainstream language assessment (i.e., the PLS-4); third, it employed the TARPA not just with children with ASD (Experiment 1) but also with typically developing children (Experiment 2); fourth, it used a substantially larger number of participants (i.e., n = 10 in Experiment 1 and n = 13 in Experiment 2, for a total of 23 overall); fifth, it conducted a statistical hierarchical analysis of the protocol; sixth and finally, it used a touch screen rather than a mouse-based computer.

Correlation between PLS 4 and TARPA

The main result of this study was the finding of a set of strong and significant correlations between the TARPA and the PLS-4 as well as between the TARPA and each of the two PLS-4 subscales (i.e., Auditory Comprehension and Expressive Communication). Furthermore, as just indicated, this pattern of results was found with a more comprehensive protocol (involving most importantly a greater range of modalities in the tests for each of the derived relational properties) and with a larger sample of children, that included a subset of children with ASD (Experiment 1) as well as a subset of typically developing children (Experiment 2). This result thus constitutes yet further support for the TARPA as a methodology for assessment of DRR and for RFT as a functional analytic approach to language.

Order Analysis

The current study also advanced on Moran et al. (2010) in that it conducted a set of formal hierarchical analyses of the TARPA. In a multi-stage hierarchically structured test, a participant should not be able to succeed at an ostensibly more difficult level after failing on an ostensibly less difficult one. The data reported by Moran et al. suggested that the TARPA broadly conformed to this criterion but the analysis used in that study was a relatively rudimentary one. The current study involved a set of more formal hierarchical analyses (see e.g., Kerr et al., 1977; Krus, Bart, & Airasian, 1977, Marion et al., 2003) that assessed the TARPA as a whole as well as focusing in particular on the levels of Stage 3, which examined DRR. Findings from both analyses indicated that the hierarchical structure of the TARPA is generally robust. The one exception to this pattern was in the comparison between combinatorial entailment and transformation of function. It was suggested that this potential weakness might be remedied in future versions of the protocol by exposing participants to more than one transformation of function test. It was also acknowledged that, even though the current analysis did not show a similar weakness with respect to order in other sections of the TARPA, the possibility that such weaknesses might exist cannot be discounted. It was argued that given multiple stages and levels across the protocol, it is unlikely that this type of pattern of performance might allow a participant to attain a much higher score than they would otherwise attain. Nevertheless, it can be argued that for important levels such as those testing derivation, the TARPA might be improved by requiring correct performance with more than one stimulus set.

Auditory versus Visual performance

One pattern that Moran et al. (2010) found was that participants tended to show better performance with visual stimuli than with auditory stimuli. In the current study, a similar pattern is evident. In order to compare visual performance with auditory performance, participants' total scores on portions of the TARPA involving only visual stimuli (i.e., the total for scores on the visual tracks of Stages 1 and 2 plus the score for the visual-visual track of Stage 3) was compared with their total scores on portions involving only auditory stimuli (i.e., the total for scores on the auditory tracks of Stages 1 and 2 plus the score for the auditory-auditory track of Stage 3). Eight out of the ten participants with ASD showed higher visual than auditory scores while the remaining two participants registered equal scores (P1 scored zero on both, while P7 achieved the maximum on both). In the case of the typically developing children, nine out of the twelve showed higher visual than auditory scores, two achieved equal scores (P13 scored one on both, while P14 scored six on both) and one (P11) showed a higher score on auditory than on visual.

The Effect of Stimulus Modality in Stage 3 Tracks. This study also compared visual and auditory performance for Stage 3 alone by comparing aggregate performance on each of the four tracks in terms of the number of levels passed. For the children with ASD, this revealed that the visual-visual track showed the highest number of passes (16), while the auditory-auditory and auditory-visual 1 track showed the lowest number (9). For the typically developing children, the visual-visual track also showed the highest number of passes (26), the auditory-auditory track showed the lowest number (16), while the two auditory-visual tracks showed intermediate numbers of passes (AV1 = 23; AV2 = 25).

Finally, Stage 3 tracks were also compared in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function). In the case of the children with ASD, it was found that on the visual-visual track three participants showed all three properties whereas in the case of each of the other tracks which involved at least some auditory stimuli only one did so. Out of the typically developing children, there were three participants who showed all three properties on the visual-visual track, three who showed all three on the auditory-visual 1 track, three who showed all three on the auditory-visual 2 track and none who showed all three on the auditory-auditory track.

Hence, as suggested, the pattern revealed in the case of both the children with ASD in Experiment 1 and the typically developing children in Experiment 2 is one in which participants tended to show better performance with visual than with auditory stimuli. Moran et al. (2010) suggested that this pattern of results might be explained on the basis of the difference in behavioural control obtained by simultaneous and successive discriminations. While this remains a very plausible explanation, another possible factor that might explain this pattern is that situations involving the types of auditory discrimination required in the TARPA protocol are not very common in the everyday socio-verbal environment, at least not in the particular way in which they are presented in this context. It is possible therefore that at least some of the trials involving auditory stimuli might require more training than other types of trials because discrimination of auditory stimuli (i.e., discrimination of different sound patterns, identification of the context in which each is used etc.) is essential in language learning and yet it is not easily learned, requiring longer exposure in the everyday language environment.

This study is the first involving the use of a more comprehensive TARPA as a methodology for assessing DRR. As has been discussed, this study improves on Moran et al. (2010) in a number of important respects. While the results in terms of the (i) correlational analyses and (ii) hierarchical analysis are encouraging this work has also identified the transformation of function test in its current form as a potential weakness in the TARPA protocol. This issue was addressed by the next experiment in this series, reported in the next chapter.

Chapter 3

Refining the TARPA: An Examination of Two Procedural Changes to the TARPA Protocol

Experiments 1 and 2 presented in Chapter 2, advanced the development of the TARPA by correlating scores on the protocol with the Preschool Language Scale - 4th Edition (PLS-4; Zimmerman, Steiner, & Pond, 2002), a mainstream performance-based language assessment, in ten children with ASD (Experiment 1) and 13 typically developing children (Experiment 2). This study also involved a formal hierarchical analysis of the TARPA (e.g., Krus, Bart & Airasian, 1979). Results showed (i) a strong and significant correlation between the TARPA and PLS-4 full-scale and subscales and (ii) evidence to support the hierarchical structure of the protocol. The results also indicated specific issues that needed to be addressed as well as further avenues of investigation.

One such avenue is related to the Pass/Fail criteria employed on certain levels of the TARPA. In the experiments presented in the previous chapter, tests of derivation (i.e., mutual entailment, combinatorial entailment & transformation of function) only allowed one exposure to the test. More specifically, after learning and showing maintenance of the requisite baseline conditional discriminations, participants were then exposed to the appropriate derivation test only once; if they passed they then proceeded to the next level; however if they failed, testing on that track was finished. This Pass/Fail criteria was adopted in order to minimise the time taken to administer the protocol.

Administration time is an important variable to consider when testing young and/or developmentally delayed children. As mentioned in the previous chapter, testing young and/or developmentally delayed children is associated with a variety of issues (e.g., Peláez, 2009) including (1) difficulties in maintaining the child's attention to sustain appropriate responding and (2) difficulties in preserving the child's participation in a single-learner training environment. Shorter test administration times have been suggested as a means of minimising these issues, and thus producing more valid estimates of the child's ability (Coolican, Bryson & Zwaigenbaum, 2008). However, while the strong and significant correlations presented in Chapter 2 provided support for validity of the Pass/Fail criteria used in these experiments, multiple exposures to derivation tests are a relatively common feature in the equivalence literature (e.g., Cassidy, 2008; O'Connor, Rafferty, Barnes-Holmes & Barnes-Holmes, 2009).

For example, in a recent study which was similar in certain respects to the research being carried out for this thesis, O'Connor et al. (2009) investigated equivalence responding across three experiments in both children with ASD (n = 15) and typically developing children (n = 3). Specifically, they looked at the effect of (i) verbal ability (ii) stimulus nameability and (iii) stimulus familiarity on equivalence performance. The relationship between verbal ability and equivalence was assessed by including children with ASD categorised with different levels of verbal behaviour based on the Comprehensive Application of Behavior Analysis to Schooling (CABAS®; Greer & Ross, 2008; Greer, 2002) system. Using this model, children were assigned to one of four categories of verbal ability; Listener/Prespeaker, Speaker/Prereader, Speaker/Reader, and Reader/Writer, in order of competence.

Participants were then exposed to four conditions (called 'stages') and all stages involved training in conditional discriminations and tests for the emergence of equivalence relations. The stages differed only in terms of the stimuli involved and the presence or absence of a verbal antecedent. The four stages were; (i) nameable and familiar stimuli with an antecedent; (ii) unnameable and familiar with no antecedent; (iii) unnameable and unfamiliar with an antecedent and (iv) unnameable and unfamiliar with no antecedent. Procedurally, O'Connor et al. (2009) adopted a recursive approach to training and testing for equivalence relations. In each stage, if a participant failed to pass an equivalence test they were re exposed to the conditional discrimination training and then retested. This cycle of training and testing continued until all participants showed patterns consistent with equivalence responding. They found that equivalence performances were significantly influenced by levels of verbal behaviour and to some extent by stimulus nameability. Children with higher levels of verbal ability (i.e., typically developing children and Reader/Writers in the ASD group) produced more rapid equivalence performances than those with lower verbal skills (i.e., Listener/Prespeakers).

Standardised assessment tools typically have strict rules for the discontinuation of testing; this allows a test to be administered in an efficient manner that meets the needs of the applied setting. Recursive training and testing procedures, often seen in tests of derived relational ability, differ in that administration time can vary considerably from one participant to another, as one participant might require extensive training in order to demonstrate DRR while another might require very little training to show the same pattern. Demonstrating DRR within one or two exposures means that the repertoire is relatively robust whereas taking further exposures suggests it is less so. Ideally a test of derived relations would allow multiple exposures to training and testing and the number of exposures required would go into the calculation of the score obtained. However this would also make testing much longer in some cases, especially if, as in the case of the TARPA, multiple tests of derived relations have to be administered. This prolonged testing could then affect other variables such as attention, tiredness etc. It was decided initially that the TARPA would aim to provide a more efficient and less sensitive measure of DRR. As such DRR testing was shortened to only one exposure per separate DRR test type. However, on further review, it was decided that this criterion might be argued to be too strict in that it allows very little leeway for error on the part of the participant. Hence, in the context of on-going testing and development it was decided to attempt to compromise a bit more between efficiency of testing on the one hand and sensitivity to DRR repertoire on the other by adjusting the protocol to allow a second exposure to derivation tests in Stage 3 of the TARPA. In the current study if a participant failed a derivation test, they were retrained on the requisite conditional discriminations and were then exposed to the appropriate derivation test a second time.

Another avenue of investigation concerned the transformation of functions (TOF) test. A possible weakness of the protocol in this respect became apparent based on the results of the order analysis (Krus et al., 1977) used to evaluate the hierarchical structure of the TARPA. More specifically, there was a pattern whereby participants passed a higher level (the TOF, S3L4) after failing a lower level (combinatorial entailment, S3L3) which indicated a disconfirmation of the suggested hierarchical structure. This pattern suggested that S3L4 may be particularly susceptible to a 'false positive' (i.e., a participant passing for reasons other than showing genuine TOF).

It was suggested in the previous chapter that one reason that this test (which required the performance of particular actions by the participant dependent on the on-screen stimulus) might have been susceptible to a 'false positive' was that it involved more interaction with the experimenter than other tasks and might therefore have been more susceptible to inadvertent cuing by the latter. Another possible explanation was that some participants simply responded consistently across the task, whether correct or incorrect and that, in the event that their initial answer was inadvertently correct then this consistency resulted in an ostensible pass even in the absence of actual TOF. As a result of this finding, the TOF test was changed in this study. In order to minimise the possibility of 'false positives', participants were exposed to two TOF tests.

Another change made in this study was in relation to schedules of reinforcement. As mentioned previously, when testing young and/or developmentally delayed children testing conditions have to minimise disruptive behaviours and maximise attention. In order to achieve this, in addition to the computer delivered schedules of reinforcement built into the TARPA protocol, individualised schedules of reinforcement (that target on-task behaviour) were adapted across Experiments 1 and 2. For children with ASD (Experiment 1) typically this involved adapting existing schedules of reinforcement from their educational placement. Schedules were also adapted for typically developing children in Chapter 2 (Experiment 2), however, existing or systematic schedules were usually not in place for typically developing children prior to their participation in the experiment, and as such it took time to organise this for each child. In order to improve the testing process, a generic token board system was developed for use in this study (see Materials for further details and Appendix C for a sample token board).

Method

Participants

A prior power analysis indicated that a sample of six or more participants was needed to have 80% power for detecting a large sized effect at the traditional alpha value of .05. A large effect size was predicted based on Moran et al. (2010) and the results of the experiments reported in the previous chapter.

10 typically developing children (5 male, 5 female; age range 4 yrs 10 mths - 6 yrs 5 mths; see also Table 9) were recruited through two primary schools (Derrywash National School, Castlebar, Co. Mayo and Scoil Chroí Íosa, Galway City). Consent for conducting the study was obtained from each of the school principals in the respective schools and parental consent was also obtained in the case of each participant. Verbal consent was also obtained from each of the children themselves. None of the children had taken part in research before.

Materials

Materials included the TARPA, the PLS-4 and a token board system.

TARPA/PLS-4. The TARPA and the PLS-4 have been described in detail elsewhere (see Chapter 2).

Token Board System. The token board consisted of two laminated A4 boards that had seven grey stars on which tokens could be placed. There was a piece of Velcro on each star and the stars were arranged diagonally across the board and below the last star was an icon that said ‘prize’. Above each star was a picture of a screen shot from the TARPA that appeared when a section was complete. This shot contained the words ‘All done’. Colour and pictures of ‘smiley faces’ were used to make the board attractive to the children. There were seven laminated yellow stars that had a piece of Velcro on the back so they could be attached to the token board. There was also a prize bag that contained a variety of back up reinforcers including edibles (e.g., lollypops), children’s magazines and small toys (e.g., spinning top, toy cars, stickers, etc.). A sample token board can be found in Figure 3.

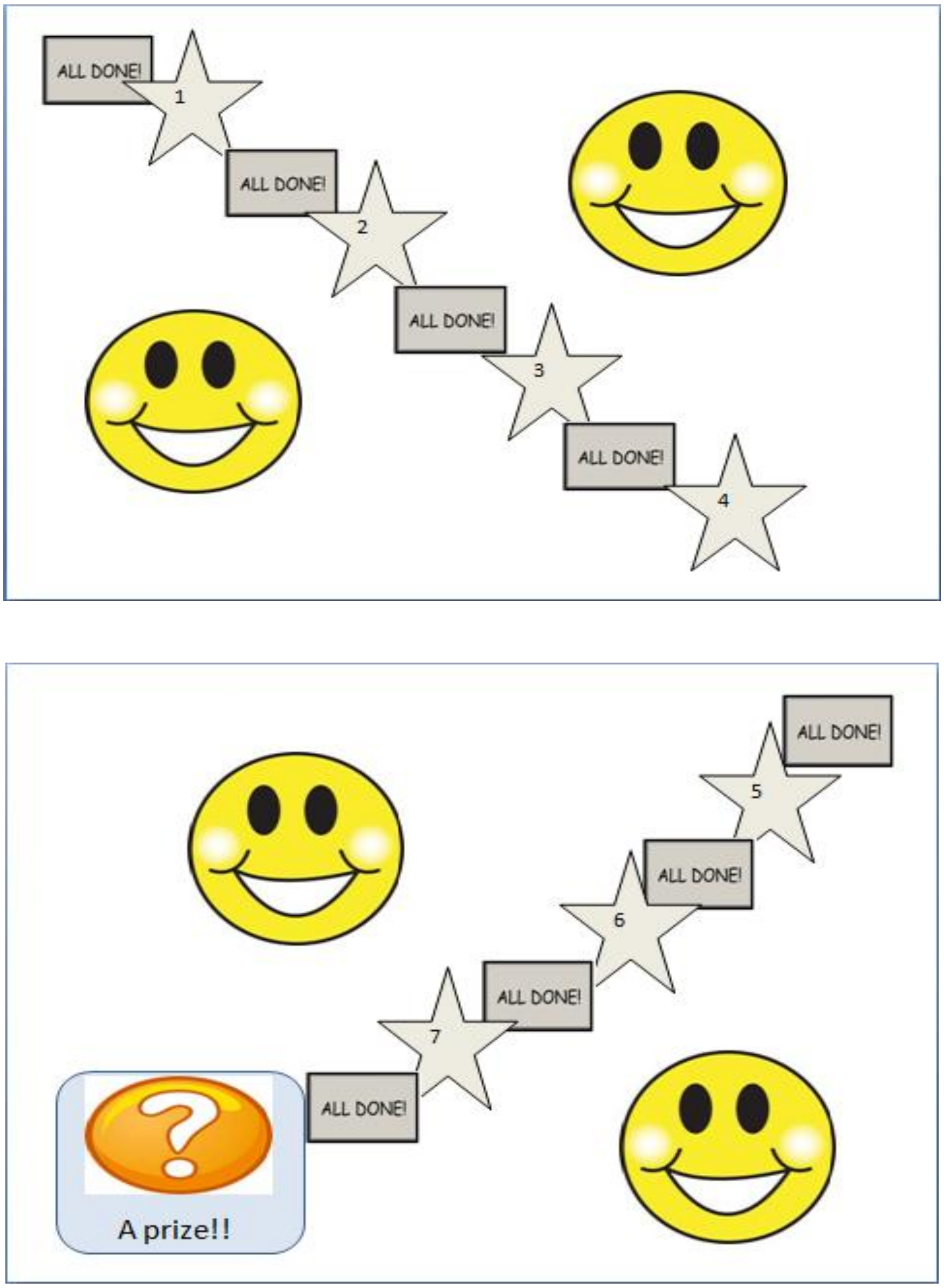


Figure 3. Token board used in Experiment 3.

Procedure

PLS-4. The PLS-4 was administered in the same manner as described in Chapter 2.

TARPA. The administration of the TARPA protocol has been described in detail elsewhere (i.e., Chapter 2 and Appendix A). In this section, only specific changes made to the TARPA procedure for the purpose of this experiment are described. These changes relate to the use of a token board system, the test of transformation of function (Level 4) and the Pass / Fail criteria.

Training, Maintenance & Testing. Participants where possible were tested on further levels after they had failed a particular level or stage of the TARPA in order to provide additional data for the hierarchy across the protocol. This additional testing for the hierarchy was conducted immediately after failure on a particular level.

Training trials included a continuous schedule of reinforcement for correct responding involving computer-delivered reinforcement and verbal praise (see Chapter 2 for a detailed description the TARPA procedure). A token system was also used. Participants received tokens contingent on various forms of on-task behaviour throughout the study. That is, participants received a token and verbal praise in the form of “nice working hard” or “nice being quiet/sitting” for the maintenance of on-task behaviours during the sessions. All feedback or reinforcement for on-task behaviour was presented only after a section or level had been completed (i.e., the ‘All done’ message had been displayed on the computer screen and at least 30 s had elapsed since the end of the last training or test trial). When participants filled the token board they could choose a back up reinforcer (reward) from the ‘prize’ bag. Breaks were also provided throughout the TARPA administration; breaks were several minutes and were always provided at an appropriate point during the session (e.g., after a level or track had been completed). However, if a child requested a break before this, the experimental trials were terminated for that session, and the same trials were repeated in the following session.

The procedure for maintenance sections and derived relational testing sections was the same as described in Chapter 2, in the case of both of these, participant received independent trials only and there was no contingent reinforcement for performance on these trials; however, during these sections, children continued to be provided with reinforcement for on task behaviour as described above.

Level 4 (Transformation of Function). Level (4) tests for transformation of functions and involves five sections ([0] CE maintenance; [1] task training and [2] transformation of function testing; [3] task training and [4] transformation of function testing). An example of a performance seen in this stage is as follows. Section (0) repeats the training and derivation test sections from Level 3 (combinatorial entailment). Participants get a refresher training in the A→B and C→B relations, and are then *tested* for maintenance of combinatorial entailment (A→C and C→A). Section (0) is optional and is at the experimenter's discretion.

In Section (1) the participant is trained to perform a particular action in the presence of a stimulus (abstract picture or sound) that appears in the centre of the screen (i.e., each stimulus [A1, A2] has a discriminative function). For example, (taken from the visual-visual track) "clapping" is reinforced in the presence of one abstract picture (A1) and "waving" is reinforced in the presence of the second abstract picture (A2).

Section (2) is a tested level, assessing for the transformation of function through the previously derived relations; C→A, and A→C (i.e., in order to pass this level the participant needs to "clap" in the presence of the abstract picture (C1) and "wave" in the presence of the other abstract picture (C2) due to relation of sameness or equivalence between these two sets of stimuli).

In Section (3) the participant is trained to perform a new set of actions in the presence of stimuli A (i.e., each stimulus (A1, A2) has a new discriminative function). For example, (taken from the visual-visual track) “touch nose” is now reinforced in the presence of one abstract picture (A1) and “tap table” is now reinforced in the presence of the second abstract picture (A2). Section (4) then assesses for the transformation of function (of the new actions) through the previously derived relations; C→A, and A→C. In order to pass transformation of function the participant has to pass both transformation of function tests (i.e., Sections 2 and 4 of Level 4).

Criteria for Passing/Failing. The criteria for passing/failing levels in Stage 3 of the TARPA were changed for this experiment, previously testing within a level proceeded until either a directly trained section had been failed *twice*, or a maintenance test section had been failed *twice* (given retraining between tests). If the derivation test section was passed, the level was designated a pass, if the derivation test section was failed, the level was designated as failed.

In this experiment, participants were given a second chance on the derivation tests; if a participant failed a derivation test on their first attempt they were put through the training cycle again. This meant that they again proceeded through training until a directly trained section had been failed *twice*, or a maintenance test had been failed *twice* or until they failed a derivation test for the *second* time. Figure 4 provides a schematic representation of the criteria for passing and failing used in Chapter 2 (upper panel) and used in this study (lower panel).

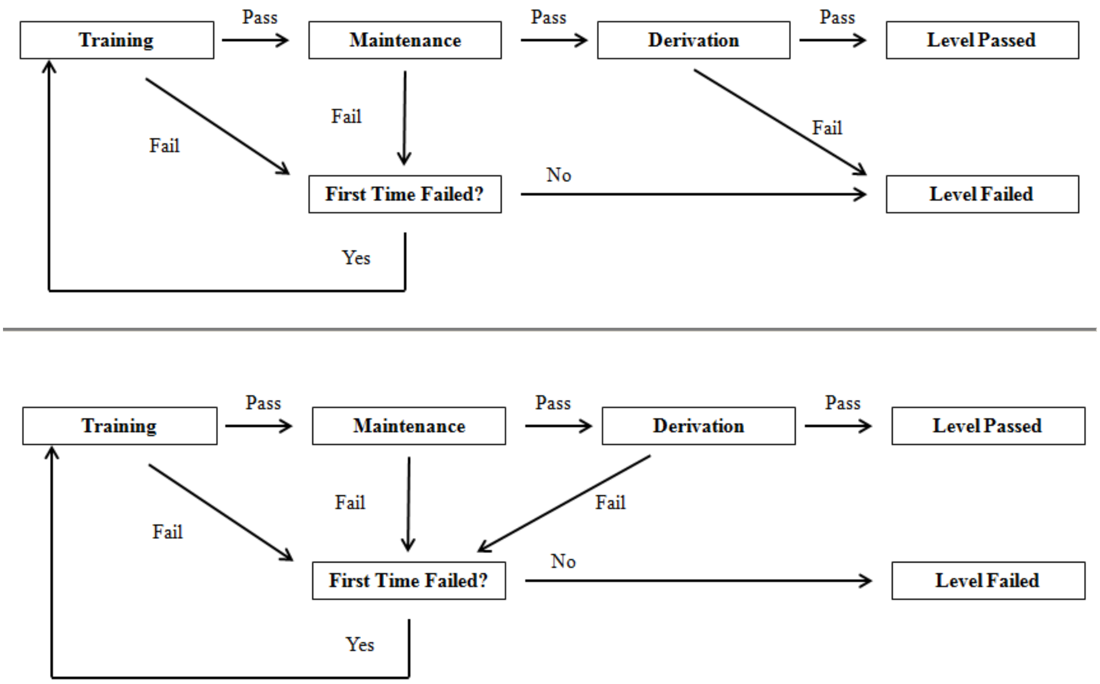


Figure 4. Termination criteria for Stage 3 Levels used previously and used in the current experiment.

Note. Upper panel shows the Termination Criteria for Stage 3 Levels used in Experiments 1 and 2 (Chapter 2).The lower panel shows the termination Criteria for Stage 3 Levels used in the current Experiment (Experiment 3).

Results

Scores and performance parameters (e.g., number and approximate lengths of sessions, etc.) for the TARPA are shown in Table 9. The column labeled *TARPA Score 1* show each participant's score based on the Pass/Fail criteria used in Chapter 2 while the column labeled *TARPA Score 2* shows each participant's score based on the Pass/Fail criteria used in this experiment. As can be seen in Table 9, for over half the participants (i.e., P2, P4, P5, P6, P9 & P10) the additional exposure to derivation tests that the Pass/Fail criteria used in this study allowed, did not change overall TARPA scores for these participants. For P1, P3, P7 and P8 the additional exposure to derivation tests resulted in them scoring between 8 and 12 points more than they would have based on the Pass/Fail criteria employed in Chapter 2.

Table 9

Demographic information and TARPA performance parameters for Experiment 3

Pt. No.	Age (Y: M)	Sex	Order of Tracks (S3)	TARPA Score 1	TARPA Score 2	No. Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested
1	04:11	m	1 2 3 4	26	38	MD	MD	MD	S3L3	S3L4
2	05:08	m	2 1 4 3	19	19	MD	MD	MD	S3L2	S3L3
3	06:00	f	2 4 1 3	22	33	4	25	100	S3L4	S3L4
4	06:00	m	1 2 4 3	12	12	4	22	88	S3L1	S3L2
5	04:11	f	4 3 2 1	15	19	4	25	100	S3L1	S3L2
			1 2 3 4	38	38	4	28	112	S3L3	S3L4
6	04:10	m	1 4 2 3	33	33	4	28	112	S3L4	S3L4
7	06:00	f	2 3 4 1	33	44	4	24.5	98	S3L4	S3L4
8	06:05	f	2 1 4 3	33	41	4	20	80	S3L4	S3L4
9	06:04	m	1 2 3 4	52	52	4	26	104	S3L4	S3L4
10	05:08	f	3 2 4 1	51	51	4	19.5	78	S3L4	S3L4

Y = Year; M = Month * P5 was tested on the TARPA twice (using different stimuli).

Correlation between PLS 4 and TARPA

Scores for both the TARPA and the PLS-4 are shown in Table 10. Spearman's rank correlation tests were used to examine the relationship between performance on the TARPA and language ability. All correlations are based on raw (as opposed to age normative) PLS-4 scores.

Table 10

TARPA and PLS-4 scores

Pt. Number	TARPA Score 1	TARPA Score 2	PLS-4 Raw AC	PLS-4 Raw EC	PLS-4 Total	PLS-4 SS	PLS-4 LAE
1	26	38	58	65	123	120	05:06
2	19	19	53	62	115	77	04:08
3	22	33	55	63	118	65	04:11
4	12	12	51	57	108	55	04:02
5	15/38	19/38	62	67	129	137	>06:05
6	33	33	54	66	120	115	05:02
7	33	44	62	65	127	105	06:01-06:03
8	33	41	60	68	128	103	06:04-06:05
9	52	52	61	65	126	95	05:11- 06:00
10	51	51	59	63	122	94	05:04- 05:05

Note. * P5 was tested on the TARPA twice (using different stimuli), both scores are shown.

A Spearman's rank order correlation revealed a non significant correlation between the TARPA scores and the PLS-4 total scores ($\rho = 0.46$, $p > 0.05$). Tests were also conducted to examine the relationship between TARPA performance and that of each of the two PLS-4 subscales. These were also non significant. Spearman's rank order tests were then conducted using scores on the TARPA based on the Pass/Fail criteria employed in Chapter 2 and scores on the PLS-4. These correlations were also non-significant.

Individual TARPA performance

Table 11 provides a more detailed breakdown of participants' performances on the TARPA within Stage 3 of the protocol. All participants passed Stages 1 and 2 and at least some levels within Stage 3; however, there was variation between participants within this stage. P1 passed Level 3 (combinatorial entailment) in one track (auditory-visual 1) but failed to move beyond Level 2 (mutual entailment) on the other tracks. P2 and P4 passed two or more tracks at Level 1 (arbitrary conditional discrimination). P4 failed to move beyond this level while P2 passed Level 2 (mutual entailment) in one track (auditory-visual 1). P3, P6, P7, P8, P9 and P10 passed all levels in the auditory-visual 1 track, across other tracks the performances of these participants differed significantly. In the visual-visual track, P3 and P6 failed to move beyond Level 1 (arbitrary conditional discrimination), P8 passed Level 3 (combinatorial entailment) while P7, P9 and P10 passed all levels within this track. In auditory-visual 2, P3, P6, P7, P8 passed Level 2 (mutual entailment) but failed to move beyond this, while P9 passed Level 3 (combinatorial entailment) and P10 passed all levels within this track. In auditory-auditory, only P9 of these participants moved beyond Level 1 (arbitrary conditional discrimination) and he passed Level 2 (mutual entailment) in this track.

P5 was first tested on the auditory-auditory track where she passed Level 1 (acquired the initial conditional discrimination) but failed Level 2 (mutual entailment). She was then tested on auditory-visual 2 where she failed Level 1 (arbitrary conditional discrimination). In auditory-visual 1 she passed Level 2 (mutual entailment) but failed Level 3 (combinatorial entailment). Then she was tested on the visual-visual track and she failed Level 2 (mutual entailment). This gave P5 an overall TARPA score of 19 (see Table 9 and Appendix B for a breakdown of scoring).

P5's score on the TARPA was considered very unusual as she scored very low on the TARPA (19 points) while scoring the highest of all participants on the PLS-4 (129, where 130 was the maximum score). Participants with scores in this

PLS-4 range tend to score well on the TARPA and this is evident both in the current experiment (e.g., P7 & P8) and in previous experiments (see Chapter 2; Participants 16, 17, 19 & 20). Figure 5 shows a scatter plot of the data. Examining this plot it can be seen that the pattern of performance of P5 seems to constitute an outlier from the general pattern. Further analysis indicated that her score was over 2 SDs from the mean, which given the small sample size of the current study was considered an unusual deviation from the trend (Field, 2009; Howell, 1998).

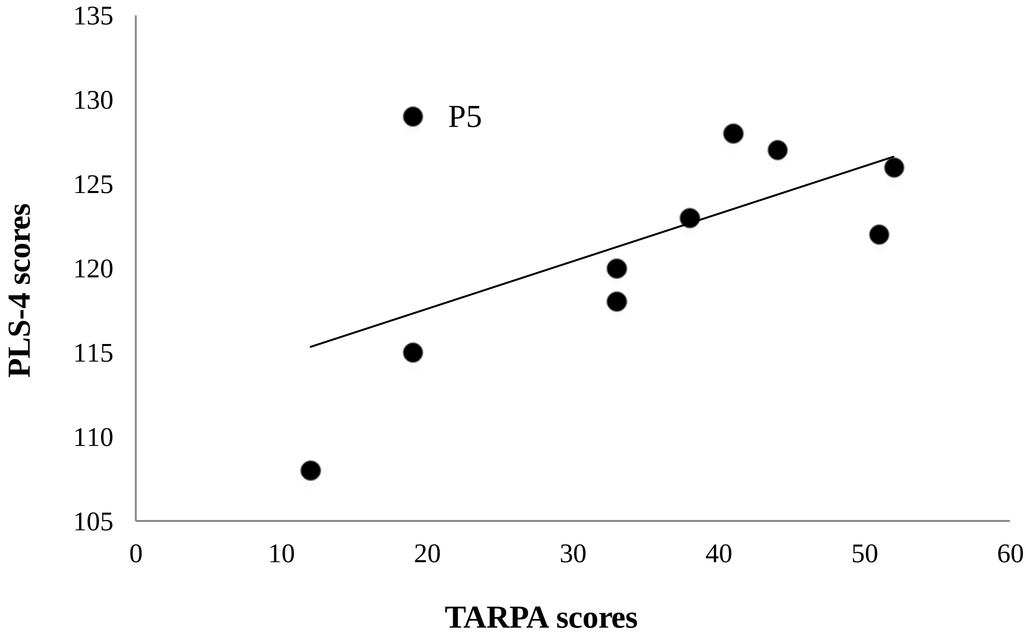


Figure 5. A scatterplot showing the relationship between PLS-4 scores and TARPA performance. P5 is marked as a possible outlier.

A possible explanation for P5's unusual patterns may be track order. Chapter 2 identified the auditory-auditory track as the track participants were least likely to show derivation patterns on; participants consistently showed the lowest number of levels passed in this track while the visual-visual track showed the highest number of passes. In the current study, P5 was the only participant to receive this track first and the visual-visual track last. It was therefore hypothesised that this order of track presentation may have affected her overall performance. For this reason, P5 was retested with a new set of stimuli four weeks later. This time she was given the reverse sequence of tracks; moving from tracks that participants consistently show derivation patterns on to tracks that participants are less likely to show derivation patterns on. Specifically, the track order was as follows; visual-visual, auditory-visual 1, auditory-visual 2 and auditory-auditory.

P5 showed considerably different patterns of responding when tested for the second time (see Table 11). In visual-visual and auditory-visual 1, she passed Level 3 (combinatorial entailment). In auditory-visual 2 she passed Level 2 (mutual entailment) and in auditory-auditory she passed Level 1 (initial conditional discrimination) but failed Level 2 (mutual entailment). This gave P5 an overall TARPA score of 38 (see Tables 9 and 10).

Table 11

Experiment 3. Performances of Individual Participants for Each Level within Stage 3 of the TARPA

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Visual-Visual										
L1 CD	P	F	P	P	P(P)	P	P	P	P	P
L2 ME	P	F*	F	F	F(P)	F	P	P	P	P
L3 CE	F		F*		P(P)	F*	P	P	P	P
L4 TOF	F*		F*		F(F)	F*	P	F	P	P
Auditory-Visual 1										
L1 CD	P	P	P	F	P(P)	P	P	P	P	P
L2 ME	P	P	P	F	P(P)	P	P	P	P	P
L3 CE	P	F	P		F(P)	P	P	P	P	P
L4 TOF	F		P		F(F)	P	P	P	P	P

Note. Levels marked with an asterisk (*) indicate testing past failure for the purpose of the order analysis.

Table 11. Continued.
Performances of Individual Participants for Each TARPA Stage and Level in Experiment 3

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Auditory-Visual 2										
L1 CD	P	P	P	P	F(P)	P	P	P	p	p
L2 ME	P	F	P	F	F(P)	P	P	P	p	p
L3 CE	F	n/a	F	n/a	F(F)	F	F	F	p	p
L4 TOF	F*	n/a	F*	n/a	F	F*	n/a	n/a	F	p
Auditory-Auditory										
L1 CD	P	P	P	F	P(P)	P	P	P	P	P
L2 ME	P	F	F	F*	F(F)	F	F	F	P	F
L3 CE	F	n/a	n/a	n/a	F	n/a	n/a	n/a	F	n/a
L4 TOF	F*	n/a	n/a	n/a	F	n/a	n/a	n/a	F*	n/a

Note. Levels marked with an asterisk (*) indicate testing past failure for the purpose of the order analysis.

Further Correlational Analyses

Further correlational analyses were conducted using P5's new score on the TARPA. A statistically significant correlation was found between total scores on the TARPA and total scores on the PLS-4 ($\rho = 0.69$, $p < 0.05$). Further tests were conducted to examine the relationship between TARPA performance and scoring on the subsections of the PLS-4. Findings showed statistically significant correlations between the TARPA and Auditory comprehension ($\rho = 0.80$, $p < 0.05$) while the correlation between the TARPA and the expressive communication subscale was non significant ($\rho = 0.44$, $p > 0.05$).

A statistically significant correlation was also found ($\rho = 0.69$, $p < 0.05$) between total scores on the TARPA based on the Pass/Fail criteria employed in Chapter 2 and total scores on the PLS-4. Further tests showed significant correlations between these TARPA scores and Auditory comprehension ($\rho = 0.76$, $p < 0.05$) while again the correlation between the TARPA and expressive communication was non significant ($\rho = 0.44$, $p > 0.05$).

Order analysis

Order analysis (Kerr et al., 1977; Krus, Bart, & Airasian, 1977, Marion et al., 2003) was used to evaluate the hierarchical structure of Stage 3 of the TARPA (all participants passed all levels within Stages 1 and 2). Table 12 compares each of the four Stage 3 levels with each of the other levels and in the case of each comparison shows (i) number of confirmations [C] and disconfirmations [D] for those two levels, where a confirmation is a pattern in which the lower level was passed and the higher level was failed while a disconfirmation is a pattern in which the higher level was passed and the lower level was failed; (ii) a z-score measurement for that pattern; and (iii) a p-value corresponding to that z-score (see Table 12).

Table 12

Experiment 3. Order analysis for TARPA Stage 3

	Combination across Tracks					
Levels	Stage 3. L2. Mutual Entailment		Stage 3. L3. Comb. Entailment		Stage 3. L4. Transf. Stim. Function	
	C	D	C	D	C	D
Stage 3. L1. Cond. Dis (A-B)	12	0	13	0	14	0
	Z = 3.4***		Z = 3.605***		Z = 3.741***	
Stage 3. L2. Mutual Entailment	-		11	0	12	0
			Z = 3.316***		Z = 3.4***	
Stage 3. L3. Comb. Entailment	-		-		4	0
					Z = 2**	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * $p < .05$. ** $p < .01$. *** $p < .001$.

As can be seen, in all cells of the table (i.e., 6/6) there were no disconfirmations of the predicted order. In addition, in all cells, there is a statistically significant ($p < 0.001$) effect in favour of the predicted hierarchy for the comparison between number of confirmations and number of disconfirmations.

Discussion

The current experiment was similar to Experiment 2 (Chapter 2) in that it correlated the TARPA with the PLS-4 in typically developing children. However, the current experiment used a version of the TARPA that featured a number of adjustments that were incorporated to improve the protocol. First, the Pass/Fail criteria was adjusted for Stage 3 (i.e., a second exposure on derivation test levels was allowed). Second, the transformation of function test was based on training and testing of two stimulus sets rather than just one.

The Adjusted TARPA protocol and Correlational Analyses

As described above P5's original score on the TARPA was considered unusual and correlations were non-significant when this score was used. However, when P5's second score was used there was a significant correlation between scores on the TARPA and the PLS-4. There was also a significant correlation between scores on the TARPA and the PLS-4 based on the previous Pass/Fail criteria employed in Chapter 2.

The correlational analyses involving the TARPA and the PLS-4 using scores based on the new and old Pass/Fail criteria for the TARPA did not suggest that the new Pass/Fail criteria changed outcomes on the TARPA considerably. These analyses revealed similar correlations between the (i) PLS-4 and the longer TARPA protocol (two exposures to derivation tests) and (ii) between the PLS-4 and the shorter protocol (one exposure to derivation tests). This perhaps suggests that the TARPA protocol when fully developed could consist of a long strictly correct version while also having a shorter more 'convenient' version. It was thought that this change to the protocol might increase testing time to unmanageable levels. However, while the additional testing did add time to the administration of the overall protocol it was not felt that this additional time made the test unfeasible.

While the correlational analysis did not detect discernible differences between TARPA outcomes with the new Pass/Fail criteria, the data from individual performances suggest otherwise. Specifically, a number of participants in this study passed higher levels and additional derivation tests in certain tracks given the second exposure (e.g., P1, P3, P7 & P8). For example, in auditory-visual 1, P3 failed Level

2 (mutual entailment) on her first attempt before passing with retraining. She then went on to pass Level 4 (transformation of function) within this track. Similarly, in the visual-visual track, P7 and P8 failed Level 2 (mutual entailment) on their first attempt before passing with retraining. They went on to pass Level 4 (transformation of function) and Level 3 (combinatorial entailment) respectively within this track. This single subject data clearly suggests that the second exposure to derivation tests may make the protocol a more robust test of DRR.

The Effect of Stimulus Modality on Stage 3 Tracks

Visual and auditory performance for Stage 3 alone was also analysed by comparing aggregate performance on each of the four tracks in terms of the number of levels passed. For these children, the auditory-visual 1 track showed the highest number of passes (AV1 = 28) followed by the visual-visual track (19) and the auditory-visual 2 track (17). The auditory-auditory track had the lowest number of passes (10). This is inconsistent with the previous experiments (Chapter 2) where the visual-visual track showed the highest number of passes. Similar to Chapter 2, the auditory-auditory track again showed the lowest number (10) of passes.

Stage 3 tracks were also compared in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function). There were five participants who showed all three properties on the auditory-visual 1 track and two who showed all three properties in the visual-visual track, while no participant showed these patterns on the auditory-visual 2 or auditory-auditory tracks. What seems clear from the data across Experiments 1-3 is a tendency for participants to perform better on DRR tasks that involve at least some visual stimuli relative to tasks that involve exclusively auditory stimuli.

Individual TARPA performance

P5's initial score on the TARPA was very low given her PLS-4 score and the performance of other participants with similar language scores across this study and previous studies. The general pattern of scores on individual tracks from this and previous experiments suggested that participants tend to pass more levels on tracks that involve visual stimuli and are less likely to pass tracks that involve only auditory stimuli (i.e., the modality of stimuli is important).

When initially tested P5 received the tracks that participants tended to perform worse on first and the tracks the participants tended to perform well on last (i.e., auditory-auditory; auditory-visual 1; auditory-visual 2 and visual-visual). It was hypothesised that getting this sequence of tracks affected her overall performance and score. Based on this P5 was retested four weeks later using new stimuli. At retest P5 was given the reverse order of tracks where she started on the track that participants tended to do well before progressing to the track where participants tended to perform less well on (i.e., visual-visual; auditory-visual 1; auditory-visual 2 and auditory-auditory). When re tested P5 scored higher and in line with her language assessment score. P5's performance across testing highlighted a number of important issues that need to be addressed.

While P5's performance tentatively suggests that getting a certain sequence of tracks could reduce overall TARPA score this needs to be examined with further research. For example, this might involve some participants being exposed to an order that starts with the track that has been found to be lowest scoring and proceeding through progressively more high scoring tracks, while others might be exposed to the reverse sequence of tracks. There is also the possibility that an unknown variable affected P5's performance (e.g., a reinforcement issue or some other artefact of testing conditions). However, given that (i) P5 was the only participant to get this exact sequence of tracks; (ii) her performance represented the only distortion of the trend and (iii) her performance when retested; it would seem likely that order of presentation of tracks was an important factor that affected her first performance. There is some support for this effect in the literature. For example, O'Connor et al. (2009) had a comparable finding in their recent study. They too had four stages (analogous to tracks in the TARPA) that had different stimulus sets that varied along dimensions of nameability and familiarity. Mean trials to criterion data indicated that these stages were sequentially more difficult. Participants in Experiment 1 of this study were given (what transpired to be) the easier sequence of stages, and matched participants in Experiment 2 were exposed to the reversed sequence. Participants in Experiment 2 took a much greater number of training trials to reach criterion (i.e., level of performance was affected). O'Connor et al. (2009) concluded that the reversed sequence rendered Experiment 2 more difficult that Experiment 1.

While the stimulus sets in the TARPA differ along modality (e.g., auditory, visual, audiovisual) and not nameability and familiarity, across this research definite differences between tracks in terms of measured levels of performance have emerged. Consistently, participants show fewer patterns of DRR on the auditory-auditory track in comparison to other Stage 3 tracks, however further research is needed to determine whether getting a certain sequence of TARPA tracks reduces overall score on the protocol.

Order Analysis

A formal analysis of the hierarchical structure of the TARPA protocol was conducted in Chapter 2. Results of the order analysis (Krus et al., 1975) indicated a weakness in the test for transformation of function. There were patterns whereby participants passed the transformation of function test (S3L4) after failing combinatorial entailment (S3L3). No participant in the current study passed the transformation of function test having failed the test for combinatorial entailment which suggests that the transformation of function test used herein is a more valid and reliable one than that used previously.

Summary

This study is the third in this series involving the use of the TARPA as a methodology for assessing DRR. The main result of this study was the finding of a strong and significant correlation between the extended TARPA (i.e., with two exposures to derivation tests) and the PLS-4. In addition, the changes made to Level 4 appear to have made the test of transformation of function a more valid and reliable test of this ability. This work, particularly the performance of P5 also identified a number of important directions for this research. Specifically, future work might focus on (i) investigating the effect of the order of presentation of the Stage 3 tracks on participants overall score and (ii) using additional measures of functioning such as IQ (e.g., Stanford-Binet 5; Roid, 2003).

Chapter 4

Further Research: Investigating Correlations with Cognitive Ability and Track Order.

The research presented in this thesis has focused on the development of the TARPA as a protocol for the assessment of DRR and its precursor skills. This has primarily involved correlating performance on the TARPA with standardised measures of language (e.g., Chapters 2 and 3). Correlations between the TARPA protocol and the Preschool Language Scales (PLS-4; Zimmerman, Steiner & Pond, 2002) have provided further support for the key premise of RFT; the idea that DRR is integral to language performance and has provided support for the TARPA as a measure of these important skills. This work has also helped identify additional research questions. Some of the issues identified are taken up in the current study. This study comprised three experiments and extended the work presented in Chapter 3 in several respects which will now be outlined.

As mentioned above previous work has involved correlations between the TARPA and measures of language ability, most prominently the PLS-4. The use of the PLS-4 was prompted by the results of the first study involving the TARPA. Moran et al. (2010) correlated performance on a preliminary version of the TARPA with scores on the Vineland Adaptive Behavior Scales (VABS; Sparrow, Cicchetti & Balla, 2005). Data from this study showed that the communication subscale of the VABS correlated highly and at borderline significance with the TARPA. This finding was coherent with the theoretical link between DRR and language and represented an encouraging early result for this work; the next step was to improve upon this study and one aspect of this was using a comprehensive performance based measure of language (i.e., the PLS-4) as opposed to an indirect questionnaire based measure (i.e., VABS). The findings presented in this thesis indicate that this particular line of inquiry has been fruitful in terms of developing the TARPA protocol. However, a key aspect of RFT is the idea that DRR not only constitutes verbal behaviour/language but that it underlies a whole gamut of cognitive abilities including those measured by traditional tests of intellectual functioning (e.g., Cassidy, Roche, & O'Hora, 2010; Cassidy, Roche, & Hayes, 2011). A number of recent studies have provided evidence for this proposition and some of this research will now be outlined.

These studies typically comprise DRR tasks involving multiple stimulus relations. For example, O'Hora, Pelaez, and Barnes-Holmes (2005) found that participants who successfully completed a complex relational task performed significantly better on the Vocabulary and Arithmetic subtests of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997) as compared to participants who failed to do so. O'Hora, Pelaez, Barnes-Holmes, Rae, Robinson and Chaudhary (2008) found that accuracy on temporal (before/after) relational responding correlated well with performance on the Block Design subtest of the WAIS-III. Similarly, O'Toole and Barnes-Holmes (2009) found that performance on an Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008) designed to assess participants fluency in before/after and similar/different relational responding correlated with intelligence quotients (IQ) as measured by the Kaufman Brief Intelligence Test.

Only one study was identified that examined derived equivalence responding and its relationship with scores on intelligence tests. In this study by Cassidy (2008) twelve typically developing children (between 8 and 12 years) were assessed on the Wechsler Intelligence Scale for Children (WISC-IIIUK) before being trained in four conditional discriminations; A1-B1, A2-B2, A1-C1 and A2-C2. Participants were required to reach a pre-set criterion of 100% correct responding across a block of 16 trials (i.e., four exposures to each of the four tasks) to finish training. The programme continued to deliver training blocks until criterion was reached (or up to a maximum of 1000 cycles). Testing for both symmetry and transitivity relations was similar in that participants cycled through testing blocks of 16 trials (with no feedback) until they met the criterion of 100% accuracy (up to a maximum of 24 cycles).

Correlational analyses examined the relationships between the total number of conditional discrimination training blocks and the total number of symmetry and transitivity testing blocks required to meet criterion. A strong and significant negative correlation was found between the symmetry testing scores and the vocabulary subtest scores. There was also a strong and significant negative correlation between performance on the symmetry test and the symbol search subtest from the performance (rather than verbal) domain of the WISC-IIIUK.

This research suggests that DRR as measured by the TARPA should correlate not only with language narrowly defined but also with intellectual ability more broadly. One of the aims of the current study was to compare the TARPA with a test of intellectual / cognitive functioning, and the Stanford-Binet 5 (SB5; Roid, 2003) was used for this purpose. This enabled a more wide-ranging and detailed investigation of the correlation between DRR as measured by the TARPA and intellectual ability.

As mentioned above Cassidy (2008) correlated IQ with training blocks to criterion on a stimulus equivalence task. Other studies (e.g., O’Connor et al., 2009; Peláez, Gewirtz, Sanchez & Mahabir, 2000) have also used training blocks to criterion or trials to criterion as a performance measure on a stimulus equivalence task. For example, Peláez et al. (2000) demonstrated that the number of training trials required for children aged 21 to 25 months to demonstrate equivalence correlated with performance on the Bzoch-League Receptive- Expressive Emergent Language Scale (Bzoch & League, 1991). Similarly, O'Connor et al. (2009) in a study involving children with ASD and typically developing children, showed a relationship between trials to criterion on four stimulus equivalence tasks and verbal ability. In the previously reported studies in this thesis, participants received points on the TARPA based on each level they passed; and their accumulated points across the protocol represented their score on the TARPA. The current study incorporated trials to criterion (i.e., training and testing trials taken to pass certain levels) as an additional measure of TARPA performance.

Another issue which arises from the findings of work reported in previous chapters relates to the effect of stimulus modality on performance on Stage 3 of the TARPA. A theme across previous chapters has been differences in children's performance across the four TARPA tracks in Stage 3. Across these experiments, performance on the four Stage 3 tracks has been compared in terms of (i) aggregate number of levels passed in each track and (ii) the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function) in a track.

A demonstrable difference between tracks in terms of level of performance has been observed. For both children with ASD (i.e., Experiment 1) as well as for typically developing children (i.e., Experiments 2 & 3) the auditory-auditory track has been repeatedly found to be the lowest scoring. Differences between the other tracks have been more marginal and less consistent. In Experiments 1 and 2 children were most likely to show patterns consistent with DRR in the visual-visual track; in Experiment 3 children tended to pass more levels in the auditory-visual 1 track. These patterns are interesting and warrant further research. One of the aims of the current study was to provide further data concerning the effect of stimulus modality on DRR.

While the effect of stimulus modality on DRR is interesting in itself, another related effect, namely the possibility that the order in which tracks are received might influence overall TARPA performance, is also interesting. Findings in Experiment 3 seemed to suggest that this might be the case. More specifically, one participant in Experiment 3 scored very low on the TARPA relative to (i) her PLS-4 scores and (ii) the scores of other children of similar ability across previous experiments. It transpired that she had received a particular sequence of Stage 3 tracks that no other participant in this experiment had received. Specifically, she started with the track that had been found to be the lowest scoring (i.e., auditory-auditory) and proceeded progressively through more high scoring tracks (i.e., visual-visual). She was subsequently retested and the order in which she received tracks was manipulated so that she was tested first on what tended to be the highest scoring track (i.e., visual-visual). The score she received at this point was higher and appeared to correspond more closely with her PLS-4 scores and patterns found across this thesis; thus suggesting that the order she had received Stage 3 tracks in had negatively affected her performance.

A recent study by O'Connor et al. (2009) has also provided evidence pertaining to this issue. O'Connor et al. (2009) found an 'order-of-difficulty' effect when examining derived equivalence performance with both typically developing children and children with ASD. In this case, the difficulty level was based on the variables of nameability and familiarity of stimuli. Children took a greater number of training trials to demonstrate equivalence when stimuli were either abstract or unnameable and they took the greatest number when they were both. In addition, a group of children exposed first to the most ‘difficult’ (i.e., abstract and unnameable stimuli) track before being exposed to progressively ‘easier’ tracks required more blocks of training and testing than an alternative group who received the ‘easiest’ to ‘most difficult’ sequence. O’Connor et al. concluded that the order in which these tasks were presented rendered the experiment more or less ‘difficult’ for the participants. As indicated above, it was speculated that the performance of one participant in a previous experiment may have been affected by order of difficulty in a similar way. The current study aimed to investigate this possibility more systematically. It was decided based on the evidence from previous studies that the likeliest order of difficulty was visual-visual, auditory-visual 1; auditory-visual 2 and auditory-auditory. As such, participants were assigned to one of two presentations of tracks; one group (designed ‘difficult to easy’) received the auditory-auditory track first and subsequently received exposure to progressively ‘easier’ tracks ending in the one designated ‘easiest’ (i.e., the visual-visual) while the other group received the reverse sequence.

Another issue which was addressed in this study concerns the transformation of function (TOF) test. In the current study, the test for TOF was presented before the test for combinatorial entailment, which has been argued to be more methodologically sound (see Barnes & Keenan, 1993; Dymond & Rehfeldt, 2000; Rehfeldt & Hayes, 1998). Dymond et al. (2000), for example, argued that if TOF is tested only after equivalence testing, then TOF-consistent responding may not be reflective of genuine TOF because the stimuli in derived relations are directly paired in the equivalence tests, and functions could thus transfer through a second-order respondent type process. Hence, they recommended that ideally TOF should be tested before the test for derived relations.

One further avenue for investigation in this study is related the development of a briefer one track version of the TARPA. The four track version of the TARPA takes a considerable amount of time to administer; a briefer one track version of the TARPA may be a desirable alternative when time doesn't allow the administration of the full protocol. In investigations of DRR as a key process underlying language, the auditory-visual tracks are particularly useful because obviously normal language development involves both the visual and auditory modalities in combination. Of the two auditory-visual tracks in this protocol the auditory-visual 1 may be the most preferable to use in a brief version of the TARPA for two reasons First, a correlational analysis using the data from children with ASD in Experiment 1 indicated that the AV1 track showed the highest levels of correlation with language ability of the Stage 3 tracks involving auditory stimuli. Second, this track (together with the visual-visual track) tends to be the highest scoring track. These reasons suggest that this track may be the most appropriate for use with children with ASD and/or developmental delay. While all participants in this study received the full TARPA protocol, a 'brief TARPA score' was also taken from the data for the purposes of a correlational analysis. The 'brief TARPA score was based on scores from Stages 1 and 2 (all levels) as well as performance on AV1.This study also looked at the other tracks in isolation (i.e., visual-visual, auditory-visual 2 & auditory-auditory) for the purpose of a correlational analysis.

This study involved typically developing children and comprised three experiments. Experiment 4a assessed a group (n = 6) of older children (7-8 years) for the first time on the TARPA. Older participants were included because it seemed likely that older children would reach the higher levels of the protocol (i.e. combinatorial entailment & TOF) and thus provide the necessary data to examine the changes made to TOF in this experiment. Experiment 4b (n = 10) primarily involved correlational analyses between the TARPA (e.g., overall scores, trials to criteria, etc.) and the other measures employed (i.e., in the case of all participants the SB5 and where participants were eligible, the PLS-4). Experiment 4c examined the effect of order of presentation of Stage 3 tracks on overall performance; this involved comparing the performance of participants in Experiment 4c who received the 'difficult-to-easy' order of tracks to participants in 4b who received the 'easy-to-difficult' order of tracks.

Experiment 4a

An older group of participants were recruited for this experiment. Specifically six children aged between 7 and 8 years were recruited to provide further data on the advanced levels of Stage 3 of the TARPA protocol and in particular to provide data on the amendment to the order of presentation of the TOF test. This change to the TOF test involved presenting the TOF test before the test for combinatorial entailment which, as explained earlier, has been argued to be more methodologically sound way of conducting this particular test (see Barnes & Keenan, 1993; Dymond & Rehfeldt, 2000; Rehfeldt & Hayes, 1998). While participants in this study did not contribute to the analysis of a track order effect, they did however receive the same 'easy-to-difficult' sequence of Stage 3 TARPA tracks (i.e., visual-visual; auditory-visual 1; auditory-visual 2; auditory-auditory) as participants in Experiment 4b in order to facilitate a later correlational analysis. These participants therefore were also tested using an abbreviated measure of intellectual functioning (ABIQ; Stanford-Binet 5; Roid, 2003).

Method

Participants

A prior power analysis indicated that a sample of six or more participants was needed to have 80% power for detecting a large sized effect at the traditional alpha value of .05. A large effect size was predicted based on Moran et al. (2010) and the results of the experiments previously reported in this thesis. Six typically developing children (3 male, 3 female; age range; 7 yrs 5 mths - 8 yrs 7 mths; see also Table 13) were recruited through a number of primary schools in Co. Mayo and Co. Galway (see Appendix D for a list of schools). One participant (P4) was an acquaintance of the experimenter. Consent for conducting the study was obtained from each of the school principals in the respective schools (with the exception of P4) and parental consent was also obtained in the case of each participant. None of the children had taken part in research before.

Materials

TARPA. The TARPA has been described in detail elsewhere (see Chapters 2 and 3 and Appendix A).

Stanford-Binet Intelligence Scales (Fifth Edition; SB5; Roid, 2003). The Stanford-Binet Intelligence Scales (Fifth Edition; SB5; Roid, 2003) is an individually administered test of cognitive abilities and intelligence, designed to assess individuals between 2 and 85 years of age. Reliabilities for the SB5 are very high; for the subtests, reliabilities range from 0.84 to 0.89 (Roid, 2003).The Abbreviated Battery IQ (ABIQ) scale of the SB5 used in this experiment consists of two subtests: one nonverbal (Object Series/Matrices) and one verbal (Vocabulary).

The nonverbal subtest (Object Series/Matrices) assesses what are referred to as ‘fluid reasoning skills’ including sequential reasoning and classic matrix items where the child selects the best alternative to complete a series or a matrix. The verbal subtest measures verbal knowledge and in particular vocabulary. Early items require a pointing response (e.g., items 1 and 3; 'touch your mouth'; 'touch your nose'). Items increase in difficulty requiring single word responses (e.g., item 12; 'What's happening in this picture?' – 'running'). Upper level items require that words be defined clearly (e.g., item 20; 'what is a parrot?').

Procedure

SB5. The first session involved SB5 assessment. The SB5 was administered to the child by the experimenter. This assessment was conducted in one sitting that lasted on average about 25 minutes per child.

TARPA. TARPA assessment began in the next session and was administered in the same manner as described in Chapter 3 with two adjustments to the procedure within Stage 3 of the protocol. First, children in Experiment 4a were presented with a specific order of Stage 3 tracks (i.e., visual-visual; auditory-visual 1; auditory-visual 2; auditory-auditory). A second change involved testing participants on TOF (Level 4) before combinatorial entailment (Level 3). Once a child passed the maintenance mixed (i.e., Section 5) in Level 3 they were then exposed to Level 4 (TOF training and testing). If the child met the criterion to pass Level 4 they were then exposed to the test for combinatorial entailment (i.e., Section 6; Level 3); if the child met the criterion to fail Level 4 they were then exposed to training within Level 3 (i.e., Section 1; Level 3)

Results & Discussion

Scores for the TARPA and the SB5 are shown in Table 13. Specific details on TARPA performance including information on number and length of sessions (in minutes) are also shown in Table 13. Scores for the TARPA were calculated as per the scoring table provided in Appendix B. Scores (raw and abbreviated IQ score) for the SB5 were calculated based on criteria outlined in the official SB5 manual. For the purpose of the correlational analyses raw scores on the SB5 (as opposed to age normative) scores were used. A Spearman's rank order correlation tests revealed non-significant correlations between scores on the TARPA and the total score on the SB5 ρ (4) = 0.3 (p > 0.05). Further tests revealed non-significant correlations between scores on the TARPA; and verbal scores on the SB5 ρ (4) = - 0.1 (p > 0.05); and non verbal scores on the SB5 ρ (4) = 0.67 (p > 0.05).

Table 13

Expt 4a. Older children. Participant sex, age and performance on the TARPA, PLS-4 and SB-5

Pt. No.	Age (Y: M)	Sex	No. TARPA Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested	TARPA Score	SB ABIQ	SB5 T Raw	SB5 Raw NV	SB5 Raw V
1	08:07	m	MD	MD	MD	S3L3 (CE)	S3L4 (TOF)	38	88	48	19	29
2	07:11	f	4	MD	MD	S3L3 (CE)	S3L4 (TOF)	34	76	41	16	25
3	07:09	f	MD	MD	MD	S3L4 (TOF)	S3L4 (TOF)	59	112	56	27	29
4	07:05	m	4	MD	MD	S3L4 (TOF)	S3L4 (TOF)	59	91	46	22	24
5	08:06	f	4	25	100	S3L4 (TOF)	S3L4 (TOF)	51	97	52	23	29
6	08:03	m	4	21	84	S3L4 (TOF)	S3L4 (TOF)	56	73	38	18	20

Note. Y = Year; M = Month; MD = Missing data; SB ABIQ = abbreviated IQ score; SB5 Raw NV = Raw non-verbal score; SB Raw V = Raw verbal score.

An older group of children were tested in this study to provide data on the advanced levels of Stage 3 of the protocol and in particular to examine whether children would show patterns consistent with TOF when this test was presented without a prior match-to-sample equivalence test.

A breakdown of individual TARPA performances for Experiment 4a is shown in Table 14 (Stage 3 only). All participants passed all levels of Stages 1 and 2. Three participants (P3, P4 & P5) passed all levels (i.e., passed TOF) across three tracks each. P6 passed all levels across two tracks and three levels across the two other tracks (i.e., passed combinatorial entailment). P1 passed three levels (i.e., passed combinatorial entailment) within two tracks, passed two levels (i.e., passed mutual entailment) within one track and failed to move beyond Level 1 in another track. P2 performed similarly to P1, she passed three levels (i.e., passed combinatorial entailment) within one track, passed two levels (i.e., passed mutual entailment) within two tracks and failed to move beyond Level 1 in another track.

Table 14

Experiment 4a. Performances of Individual Participants for each Level within Stage 3 of the TARPA

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)						
Part	P1	P2	P3	P4	P5	P6
Visual-Visual						
L1 CD	P	P	P	P	P	P
L2 ME	P	P	P	P	P	P
L3 CE	P	P	P	P	P	P
L4 TOF	F	F	P	P	P	P
Auditory-Visual 1						
L1 CD	P	P	P	P	P	P
L2 ME	P	P	P	P	F	P
L3 CE	P	F	P	P	n/a	P
L4 TOF	F	n/a	P	F	n/a	F
Auditory-Visual 2						
L1 CD	P	P	P	P	P	P
L2 ME	P	P	P	P	P	P
L3 CE	F	F	P	P	P	P
L4 TOF	n/a	n/a	F	P	P	P
Auditory-Auditory						
L1 CD	P	P	P	P	P	P
L2 ME	F	F	P	P	P	P
L3 CE	n/a	n/a	P	P	P	P
L4 TOF	n/a	n/a	P	P	P	F

Note. L = Level; P = Pass; F = Fail; n/a = not applicable

As can be seen in Table 14, the maximum number of instances of TOF that could have been observed was 24 (i.e., when each participants performance on each track is considered). Out of the 24 possible instances, there were 18 instances where participants met the criterion to be tested on Level 4 (TOF training and testing). In 11 out of 18 instances, patterns consistent with TOF were observed. This was a yield of 61%. In addition, these patterns were evident across different tracks. In Experiment 3 where TOF was tested after combinatorial entailment, the maximum number of instances of TOF that could have been observed was 40. Out of these there were 15 instances where participants were exposed to Level 4 (TOF training and testing) and in the case of 10 instances, patterns consistent with TOF were observed. This was a yield of 75%. Given that the participants in the current study were older it might be expected that they would show a higher yield of TOF than was seen in Experiment 3 where participants were younger. Dymond et al. (2000) argued that if TOF is tested after equivalence testing, then TOF-consistent responding may not be reflective of genuine TOF because the stimuli in derived relations are directly paired in the equivalence tests, and functions could thus transfer through a second-order respondent type process. When the results of Experiment 3 and 4a are compared they suggest that TOF is more likely to be observed if a child has succeeded on a prior equivalence test. TOF is (from an RFT perspective) a critical language skill as it gives derived relations psychological content (Hayes et al., 2001). As such the development of a valid test of this ability is a key aim of this work and the results presented here would suggest that presenting TOF before combinatorial entailment provides a more authentic test of this ability. For this reason this procedural change was maintained for subsequent experiments in this thesis.

It was also considered interesting that no participant in this experiment passed all stages and levels across the protocol (i.e., showed a ceiling effect); suggesting that the responding of older participants on the TARPA can provide important data on (i) the 'more difficult' tracks (e.g., auditory-auditory) as younger participants tend not to advance to the higher levels of these tracks and on (ii) higher levels across all tracks (i.e., TOF). For these reasons a small number of older participants were also included in Experiments 4b and 4c.

Experiment 4b

Experiment 4a tested a modification to the TOF test; the results seemed to suggest that this version of the TOF test may be a more valid test of this critical skill. Experiment 4a also indicated that older participants may provide useful data on the higher levels of the TARPA and suggested that the SB5 (Roid, 2003) may provide a broader basis of cognitive ability to compare with TARPA performance. Based on these findings, a further ten participants were recruited for Experiment 4b.

Participants in Experiment 4b were tested on the TARPA and the SB5; thus this experiment added an additional 10 participants to the overall sample tested, giving a total of 16 for the study as a whole, which increased the power of a potential correlational analysis. In addition, where eligible, participants in Experiment 4b were also tested on the PLS-4, this allowed comparisons between different correlations (e.g., correlation between the PLS-4 and SB5 versus the correlations between the PLS-4 and the TARPA and the SB5 and the TARPA).

Method

Participants

A prior power analysis indicated that a sample of six or more participants was needed to have 80% power for detecting a large sized effect at the traditional alpha value of .05. A large effect size was predicted based on Moran et al. (2010) and the results of the experiments reported thus far in this thesis. 10 typically developing children (3 male, 7 female; age range ; 4 yrs 8 mths - 8 yrs 2 mths; see also Table 15 for further participant information) were recruited through a number of primary schools in Co. Mayo and Co. Galway (see Appendix D for a list of schools). P12 was an acquaintance of the experimenter. Consent for conducting the study was obtained from the school principal where applicable. Parental consent was also obtained for each participant. Verbal consent was also obtained from each of the children themselves. None of the children had taken part in similar research before.

Materials

The TARPA, SB5 and the PLS-4 have been described elsewhere (see Experiment 4a).

Procedure

All ten participants were tested on the SB5 and participants who were eligible were also tested on the PLS-4 (n = 9). The TARPA assessment was administered in the same manner as in Experiment 4a (i.e., participants in this study were tested on TOF before combinatorial entailment and received the same 'easy-to-difficult' order of Stage 3 tracks [visual-visual; auditory-visual 1; auditory-visual 2; auditory-auditory]).

Results & Discussion

Table 15 provides demographic information and details on TARPA performance including information on number and approximate length of sessions (in minutes). All participants were tested on the TARPA and the SB5; nine participants (i.e., all those eligible) were also tested on the PLS-4. P7 exceeded the age range for this test (i.e. she was older than 6.5 years). P11 and P16 also exceeded the age range but their performance on the verbal subtest of the SB5 suggested a reduced verbal ability and they both were therefore tested on the PLS-4.

Table 15

Expt 4b. Demographic Information and TARPA Performance Parameters

Pt. No.	Age (Y: M)	Sex	No. Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested
7	07:10	f	4	19	76	S3L4 (TOF)	S3L4 (TOF)
8	06:00	f	4	22	88	S3L3 (CE)	S3L4 (TOF)
9	05:11	m	5	26	130	S3L2 (ME)	S3L3 (CE)
10	06:05	f	5	26	130	S3L4 (TOF)	S3L4 (TOF)
11	06:10	f	5	24	120	S3L2 (ME)	S3L3 (CE)
12	05:02	f	6	35	210	S3L3 (CE)	S3L4 (TOF)
13	05:05	f	5	20	100	S3L4 (TOF)	S3L4 (TOF)
14	04:08	m	5	19	95	S3L4 (TOF)	S3L4 (TOF)
15	05:03	f	7	16	112	S3L4 (TOF)	S3L4 (TOF)
16	08:02	m	6	25	150	S3L4 (TOF)	S3L4 (TOF)

Note. Y = Year; M = Months; CE = Combinatorial Entailment; TOF = transformation of function.

Individual scores (TARPA, SB5 & PLS-4) for participants in Experiment 4b are shown in Table 16. Scores for the TARPA were calculated as per the scoring table provided in the Appendix B. Scores for the SB5 (abbreviated, total, verbal & non-verbal raw scores) and PLS-4 (raw and age equivalent) were calculated per their respective manuals (Roid, 2003; Zimmerman et al. 2003).

Table 16

Expt 4b. TARPA, SB-5 and PLS-4 scores

Pt. Number	TARPA Score	PLS4 Raw AC	PLS4 Raw EC	PLS4 Total	PLS4 SS	PLS4 LAE	SB ABIQ	SB5 T Raw	SB5Raw NV	SB5 Raw V
7	59	n/a	n/a	n/a	n/a	n/a	115	57	26	31
8	41	62	68	130	116	>6:5	115	45	22	23
9	34	57	58	115	84	04:08	88	31	12	19
10	41	55	65	120	72	05:02	88	36	16	20
11	34	61	67	128	n/a	n/a	91	44	17	27
12	30	58	62	120	94	05:02	91	29	10	19
13	33	55	57	112	78	04:05	97	32	13	19
14	23	52	54	106	86	04:00	94	27	10	17
15	44	62	67	129	125	>6.5	112	37	13	24
16	41	59	58	117	n/a	n/a	52	27	8	19

Note. AC = Auditory Comprehension; EC = Expressive Communication; SS = Standardised Score; LAE = Language Age Equivalent; SB ABIQ = abbreviated IQ score; SB5 T Raw = Raw non-verbal score plus Raw verbal score; SB5 Raw NV = Raw non-verbal score; SB Raw V = Raw verbal score.

For the purposes of the correlational analyses raw (as opposed to age normative) scores on the SB5 and PLS-4 were used. A Spearman's rank order correlation tests revealed significant correlations with TARPA scores and total SB5 scores ρ (8) = 0.65 ($p < 0.05$) Further tests revealed a significant correlation with TARPA scores and verbal scores on the SB5 ρ (8) = 0.76 ($p < 0.05$) while the correlation with the non verbal score was not statistically significant ρ (8) = 0.55 ($p > 0.05$). The strong correlations between the TARPA scores and the verbal subtest of the SB5 are in line with previous research where verbal ability (e.g., PLS-4 & the communication domain of the VABS) have correlated with TARPA scores. The correlations between the TARPA and the non verbal portion of the SB5 approached statistical significance; thus as in Experiment 4a suggesting the relevance of the skills measured by the TARPA to cognitive ability more broadly.

Participants who were eligible in Experiment 4b were also assessed using the PLS-4 (all children in Experiment 4a exceeded the age range for this measure; nine participants in 4b were suitable for testing on this measure). The following section combines the data from Experiments 4a and 4b for the purpose of a larger correlational analysis. This section outlines correlations between the TARPA (i.e., overall scores, the ‘brief’ TARPA, trials to criterion, etc) and the SB5. To avoid repetition the correlations between the TARPA and the PLS-4 from Experiment 4b are detailed in the next section also.

Combined Analysis: Correlations between the TARPA and other measures

Data from Experiments 4a and 4b were combined for purposes of this correlational analysis. Spearman rank order correlations between performance on the TARPA (i.e., overall score, a ‘brief’ score, visual and auditory scores, individual tracks and trials to criterion data), the PLS-4 (total score and scores on subscales) and the SB5 (total scores, non verbal and verbal scores) are shown in Table 17.

Table 17

Expt. 4a & Expt 4b. Correlations between the TARPA and the PLS-4 and SB5

TARPA	Total TARPA Score	AV1 Brief TARPA	VV Vis. score (Stages 1-3)	AA Aud. score (Stages 1-3)	AV2 (Stage 3)	Trials to pass VV
SB5 T Raw	ρ= 0.69** (n = 16)	ρ= 0.46 (n = 16)	ρ= 0.3 (n = 16)	ρ= 0.57* (n = 16)	ρ= 0.75** (n = 16)	ρ= - 0.6 (n = 9)
SB5 Raw NV	ρ = 0.73** (n = 16)	ρ= 0.45 (n = 16)	ρ= 0.36 (n = 16)	ρ= 0.36 (n = 16)	ρ= 0.81** (n = 16)	ρ= - 0.44 (n = 9)
SB5 Raw V	ρ= 0.62* (n = 16)	ρ= 0.46 (n = 16)	ρ= 0.26 (n = 16)	ρ= 0.5* (n = 16)	ρ= 0.6* (n = 16)	ρ= - 0.74* (n = 9)
PLS-4 Total	ρ= 0.68 * (n = 9)	ρ= 0.83** (n = 9)	ρ= 0.03 (n = 9)	ρ= 0.23 (n = 9)	ρ= 0.77* (n = 9)	ρ= - 0.4 (n = 4)
PLS-4 Raw AC	ρ = 0.67* (n = 9)	ρ = 0.65 (n = 9)	ρ= 0.00 (n = 9)	ρ= 0.44 (n = 9)	ρ= 0.74* (n = 9)	ρ= - 0.25 (n = 4)
PLS-4 Raw EC	ρ= 0.67* (n = 9)	ρ= 0.79* (n = 9)	ρ= - 0.0 (n = 9)	ρ= 0.27 (n = 9)	ρ= 0.77* (n = 9)	ρ= - 0.4 (n = 4)

Note. "ρ" represents Spearmans Rank Order Correlation coefficient. Asterisks indicate a statistically significant correlation (2-tailed); * p < .05. ** p < .01. *** p < .001. 'n' represents the sample size.

Total TARPA score. A Spearman's rank order correlation showed a highly statistically significant correlation between the TARPA scores and the total scores on the SB5 ρ (14) = 0.69 (p < 0.01). Further tests revealed highly significant correlations between the TARPA and both the verbal scores ρ (14) = 0.62 (p < 0.01) and the non-verbal scores ρ (14) = 0.73 (p < 0.01) on the SB5.

For the nine participants who were tested on the PLS-4, a Spearman's rank order correlation showed a statistically significant correlation between the TARPA scores and the PLS-4 total scores ρ (7) = 0.68 ($p < 0.05$). Further tests using the two PLS-4 subscales revealed non-significant correlations with both Auditory Comprehension ρ (7) = 0.67 ($p > 0.05$) and Expressive Communication ρ (7) = 0.67 ($p > 0.05$). These correlations were just outside statistical significance

Brief TARPA score. A Spearman's rank order correlation revealed a non statistically significant correlation between the brief TARPA scores and the total scores on the SB5 ρ(14) = 0.46 ($p > 0.05$). Further tests showed that correlations between the brief TARPA scores and the non-verbal scores ρ (14) = 0.45 ($p > 0.05$) and verbal raw scores ρ (14) = 0.46 ($p > 0.05$) on the SB5 were also non significant.

Tests were also conducted to examine the relationship between the brief TARPA scores and the scores on the PLS-4. A highly significant correlation was shown between the brief TARPA scores and the PLS-4 total scores ρ (7) = 0.83 ($p < 0.01$). Further tests using the two PLS-4 subscales revealed significant correlations with Expressive Comprehension ρ (7) = 0.79 ($p < 0.05$) while the correlation with Auditory Communication ρ (7) = 0.65 ($p > 0.05$) was non significant.

Further correlations. Further correlational analyses were conducted using certain portions of TARPA performance. A 'visual' score was calculated based on performance on the visual levels within Stages 1 and 2 and the visual-visual track. Likewise, an 'auditory' score was calculated based on performance on the auditory levels within Stages 1 and 2 and the auditory-auditory track.

Spearman rank correlations tests indicated that neither the SB5 nor the PLS-4 were significantly correlated with the 'visual' score from the TARPA. Correlations with the 'auditory' TARPA score were stronger. A Spearman's rank order correlation showed a statistically significant correlation between the 'auditory' TARPA score and total score on the SB5 ρ (14) = 0.57 ($p < 0.05$). Further tests involving the subtests of the SB5 revealed a significant correlation between the 'auditory' TARPA score and the verbal score on the SB5 ρ (14) = 0.50 ($p < 0.05$). The correlation with the non verbal score on the SB5 was non significant ρ (14) = 0.36 ($p > 0.05$).

Scores on the auditory-visual 2 (AV2) track were also correlated with the SB5 and the PLS-4. A Spearman's rank order correlation showed a highly statistically significant correlation between AV2 scores and total scores on the SB5 ρ (14) = 0.75 ($p < 0.01$). Further tests indicated highly statistically significant correlations between AV2 and the SB5 subscales: both the non verbal ρ (14) = 0.81 ($p < 0.01$) and the verbal ρ (14) = 0.6 ($p < 0.01$).

Tests were also conducted to examine the relationship between AV2 performance and the scores on the PLS-4. A significant correlation was shown between TARPA scores and the PLS-4 total scores ρ (7) = 0.77 ($p < 0.05$). Further tests using the two PLS-4 subscales also revealed significant correlations with both Auditory Comprehension ρ (7) = 0.74 ($p < 0.05$) and Expressive Communication ρ (7) = 0.77 ($p < 0.05$).

Trials to criterion on the visual-visual track. Trials taken to pass a TARPA track may provide a more precise measure of TARPA performance in some instances (e.g., when there is a group of participants with the same overall score or who pass the same level or stage). Trials to criterion for a particular level was defined as the total accumulated number of training and testing trials taken to pass that level. In Experiments 4a and 4b, nine participants in total passed all portions of the visual-visual track (i.e., passed all levels from 1 to 4 and scored the maximum of 14 points for this track).

Trials taken to pass this track were correlated with the SB5 and its subscales. A Spearman's rank order correlation test showed a non-significant correlation with the full scale SB5 (ρ [5) = - 0.6; $p > 0.05$). There was a significant correlation with the verbal subtest (ρ [5] = - 0.74; $p < 0.05$) while the correlation with the non-verbal subtest was non-significant (ρ [5] = - 0.4; $p > 0.05$). A similar analysis across the other TARPA tracks was not possible because not enough participants passed these tracks outright.

Comparing correlations. Table 18 displays a correlation matrix of Spearman rank order correlations among the TARPA, SB5 and PLS-4. This table facilitates some interesting comparisons. First, the TARPA correlated significantly with the PLS-4 ρ (7) = 0.68 ($p = 0.05$) and this compares well to the correlation between the PLS-4 and scores on the verbal test of SB5 ρ (7) = 0.87 ($p < 0.05$). A stronger correlation between the PLS-4 and the verbal subscale of the SB5 would be predicted given the similarity between these measures. Second, the TARPA also correlated significantly with the verbal score on the SB5 ρ (14) = 0.62 ($p < 0.05$) and again this also compares well to the correlation between the PLS-4 (total scores) and the scores on the verbal test of SB5 (ρ (7) = 0.87 ($p < 0.05$). The third correlation is particularly interesting, the TARPA correlated significantly with the non verbal score on the SB5 ρ (14) = 0.73 ($p < 0.01$); while the correlation between PLS-4 (total scores) and the non verbal test of the SB5 (ρ (7) = 0.64 ($p > 0.05$) was just outside significance. This is not unexpected but the fact that the TARPA correlated with all measures suggests the importance of DRR as a core or underlying skill in language and cognition.

Table 18

Expt. 4a & Expt 4b.Correlations between the TARPA and the PLS-4 and SB-5 versus correlations between the PLS-4 and SB-5

Measures	Total TARPA score	SB5 T Raw	SB5 Raw NV	SB5 Raw V	PLS-4 Raw T	PLS-4 Raw AC
SB5 T Raw	ρ= 0.69** (n = 16)	-	-	-	-	-
SB5 Raw NV	ρ = 0.73** (n = 16)	ρ= 0.96*** (n = 16)	-	-	-	-
SB5 Raw V	ρ= 0.62* (n = 16)	ρ= 0.93*** (n = 16)	ρ= 0.84*** (n = 16)	-	-	-
PLS-4 Total	ρ= 0.68* (n = 9)	ρ= 0.79* (n = 9)	ρ= 0.64 (n = 9)	ρ= 0. 87* (n = 9)	-	-
PLS-4 Raw AC	ρ = 0.67 (n = 9)	ρ= 0.61 (n = 9)	ρ = 0.39 (n = 9)	ρ = 0.76* (n = 9)	ρ = 0.87** (n = 9)	-
PLS-4 Raw EC	ρ= 0.67 (n = 9)	ρ= 0.85** (n = 9)	ρ= 0.72* (n = 9)	ρ= 0.90** (n = 9)	ρ= 0.98** (n = 9)	ρ= 0.82** (n = 9)

Note. "ρ" represents Spearmans Rank Order Correllation coefficient. Asterisks indicate a statistically significant correlation (2-tailed); * p < .05. ** p < .01. *** p < .001. 'n' represents the sample size.T Raw = total raw score; Raw NV = Raw non-verbal score; Raw V = Raw verbal score; Raw T = Raw total score; Raw AC = Raw auditory comprehension score; Raw EC = Raw expressive communication score.

Individual TARPA performance

A detailed breakdown of individual performances within Stage 3 of the TARPA for Experiment 4b is shown in Table 19. All participants passed all levels of Stages 1 and 2. One of the older participants (P7) who was aged 7 years 10 months passed all levels (i.e., arbitrary conditional discrimination, mutual entailment, combinatorial entailment & transformation of function) across three tracks.

Table 19

Experiment 4b. Performances of Individual Participants for Each Level within Stage 3 of the TARPA

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Visual-Visual										
L1 CD	P	P	P	P	P	P	P	P	P	P
L2 ME	P	P	P	P	P	F	P	P	P	P
L3 CE	P	P	F	P	F	n/a	P	P	P	P
L4 TOF	P	F	n/a	P	n/a	n/a	P	F	P	P
Auditory-Visual 1										
L1 CD	P	P	P	P	P	P	P	P	P	P
L2 ME	P	P	P	P	P	P	F	F	P	P
L3 CE	P	P	F	P	F	P	n/a	n/a	P	F
L4 TOF	P	F	n/a	F	n/a	F	n/a	n/a	P	n/a

Note. L = Level; P = Pass; F = Fail; n/a = not applicable

Table 19. Continued.

Experiment 4b.Performances of Individual Participants for Each Level within Stage 3 of the TARPA

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Auditory-Visual 2										
L1 CD	P	P	P	P	P	P	P	P	P	P
L2 ME	P	P	P	P	P	P	P	F	P	P
L3 CE	P	F	F	F	F	F	F	n/a	F	F
L4 TOF	P	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Auditory-Auditory										
L1 CD	P	P	P	P	P	P	P	F	P	P
L2 ME	P	F	P	F	P	F	F	n/a	F	P
L3 CE	P	n/a	F	n/a	F	n/a	n/a	n/a	n/a	F
L4 TOF	F	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note. L = Level; P = Pass; F = Fail; n/a = not applicable

The majority of the children were between 5-6 years of age and a group composed mostly of this age range can be divided into two smaller groups based on performance. With regard to P8, P10, P15 and P16, all of whom were between 5-6 years except for P16 (8 years, 2 months), most passed Level 3 (combinatorial entailment) across two tracks (P16 passed Level 3 in the visual-visual track only) and three of them (P10, P15 and P16) passed Level 4 (transformation of function) also. P16, the one child who did not fit the age range here also had low or below average abbreviated IQ scores. P9, P11, P12 and P13 also performed within a similar range. They all passed Level 2 (mutual entailment) across at least two tracks. P12 and P13 also passed Level 3 (combinatorial entailment). Only P13 passed Level 4 (transformation of function) in the visual-visual track. All these participants were between 5-6 years old. Finally, the youngest participant, P14 passed Level 3 (combinatorial entailment) in one track (i.e., visual-visual) but failed to move beyond Level 1 on the two auditory-visual tracks and also failed to acquire the conditional discrimination in the auditory-auditory track.

The Effect of Stimulus Modality in Stage 3 Tracks (Experiments 4a & 4b)

Visual and auditory performance in the context of derived relational testing (i.e., Stage 3) was examined by comparing aggregate performance on each of the four tracks in terms of the number of levels passed. For children in Experiments 4a and 4b, the visual-visual track showed the highest number of passes (53), the auditory-auditory track showed the lowest number (31), while the two auditory-visual tracks showed intermediate numbers of passes (AV1 = 41; AV2 = 41).

Tracks were also compared in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function). There were nine participants who showed all three properties on the visual-visual track, three who showed all three on the auditory-visual 1 track, four who showed all three on the auditory-visual 2 track and three who showed all three properties on the auditory-auditory track.

When each track is considered in terms of the aggregate number of levels passed, the auditory-auditory track emerges clearly as the lowest scoring. This is consistent with the previous experiments. When tracks are compared in terms of the number of participants who showed all three properties of DRR (i.e. mutual entailment, combinatorial entailment and transformation of function) the pattern revealed is one in which a similar number of children passed the auditory-visual tracks (AV1 = 3; AV2 = 4) as did the auditory-auditory. However, it should be noted that those who passed the auditory-auditory track were older participants who demonstrated all three properties of DRR across three tracks each (e.g., P3, P4, P5 and P7).

Study 4 involved one more experiment (4c). Before moving on to the first of these, a very brief recap on the output from the first two experiments will be provided. Experiments 4a and 4b have provided some interesting additional analyses. They have advanced previous work by (i) facilitating further improvement of the validity of the TOF test (this involved presenting the TOF test before the test for combinatorial entailment); (ii) showing correlations between TARPA performance (i.e., total scores, trials to criterion & individual tracks) and cognitive as well as linguistic ability; (iii) conducting a comparative analysis of correlations between the TARPA and alternative standardised measures; and (iv) by examining trials to criterion as an additional measure of TARPA performance.

Experiment 4c

Based on findings from a previous experiment (see Chapter 3), it seemed that order of track presentation could affect TARPA scoring. The aim of Experiment 4c was to facilitate an investigation of this by presenting participants with a particular track order that could then be compared with the order received in a previous experiment in the present study (i.e., Experiment 4b). Whereas in Experiment 4b they had first been exposed to the track that participants tended to score highest on (i.e., visual-visual) and then been exposed to the two mid-level performance tracks (the auditory visual tracks, in the order AV1 followed by AV2) and finally received the track that participants tended to score lowest on (i.e., auditory-auditory), in this experiment they were exposed to the tracks in the reverse order. Performance on these two orders could then be systematically compared.

Method

Participants

A prior power analysis indicated that a sample of six or more participants was needed to have 80% power for detecting a large sized effect at the traditional alpha value of .05. A large effect size was predicted based on Moran et al. (2010) and the results of the experiments reported thus far in this thesis. Ten typically developing children (8 male, 2 female; age range 4 years 6 months - 8 years 7 months; see also Table 20) were recruited through a number of primary schools in Co. Mayo and Co. Galway (see Appendix D). Consent for conducting the study was obtained from the school principal for the primary school. Parental consent was obtained for each participant. Verbal consent was also obtained from each of the children themselves. None of the children had taken part in research before. Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland Galway Research Ethics Committee.

Materials

See Experiment 4b.

Procedure

All ten participants were tested on the SB5 and participants who were eligible were also tested on the PLS-4 (n = 7). The TARPA assessment was administered in the same manner as in Experiment 4b except for the order of tracks. Whereas in Experiment 4b they had first been exposed to the track that participants tended to score highest on (i.e., visual-visual) and then been exposed to the two mid-level performance tracks (the auditory visual tracks in the order AV1 and AV2) and finally received the track that participants tended to score lowest on (i.e., auditory-auditory), in this experiment they were exposed to the tracks in the reverse order.

Performance on these two orders referred to as easy-to-difficult and difficult-to-easy respectively, could then be systematically compared.

Results & Discussion

Demographic information and TARPA performance parameters for the ten participants in this experiment are shown in Table 20. Details on TARPA performance including information on number and approximate length of sessions (in minutes) are also shown in Table 20.

Table 20

Expt 4c. Demographic Information and TARPA Performance Parameters

Pt. No.	Age (Y: M)	Sex	No. Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested
17	08:07	m	5	25	125	S3L3 (CE)	S3L4 (TOF)
18	07:07	m	4	23	92	S3L3 (CE)	S3L4 (TOF)
19	06:08	m	4	16	64	S3L3 (CE)	S3L4 (TOF)
20	05:09	m	5	18	90	S3L4 (TOF)	S3L4 (TOF)
21	04:10	f	5	16	80	S3L3 (CE)	S3L4 (TOF)
22	04:08	m	4	19	76	S3L1 (CD)	S3L2 (ME)
23	04:06	f	4	15	60	S3L2 (ME)	S3L3 (CE)
24	05:01	m	3	24	72	S3L2 (ME)	S3L3 (CE)
25	05:04	m	3	24	72	S3L4 (TOF)	S3L4 (TOF)
26	04:09	m	3	18	54	S3L2 (ME)	S3L3 (CE)

Note. Y = Year; M = Months

Scores for the performance of the 4c ‘difficult-to-easy’ group on the TARPA, the SB5 and the PLS-4 are shown in Table 21. Spearman’s rank order correlation testing was used to examine the relationship between performance on the TARPA, the SB5 and the PLS-4. There was a non-significant correlation between the TARPA scores and the total scores on the SB5 ρ (8) = 0.57 (p > 0.05). Further tests revealed a highly significant correlation with the non-verbal score on the SB5 ρ (8) = 0.78 (p < 0.01) while the correlation with the verbal score was non-significant ρ (8) = 0.47 (p > 0.05). Tests were also conducted to examine the relationship between TARPA performance and the scores on the PLS-4. The correlation between TARPA scores and PLS-4 total scores were very clearly non-significant ρ (5) = - 0.1 (p > 0.05) as were the correlations with the PLS-4 subscales.

Table 21

Expt 4c. TARPA, PLS-4 and SB-5 scores

Pt. Number	TARPA Score	PLS-4 Raw AC	PLS-4 Raw EC	PLS-4 Raw T	PLS-4 SS	PLS-4 LAE	SB ABIQ	SB5 T Raw	SB5 Raw NV	SB5 Raw V
17	46	n/a	n/a	n/a	n/a	n/a	106	57	23	34
18	34	n/a	n/a	n/a	n/a	n/a	97	48	21	27
19	30	n/a	n/a	n/a	n/a	n/a	76	34	14	20
20	41	60	65	125	102	5:09-5:10	127	50	25	25
21	20	60	64	124	122	5:07-5:08	103	31	11	20
22	15	62	68	130	141	>6.5	112	35	12	23
23	16	57	62	119	108	5:0-5:01	121	38	18	20
24	23	56	62	118	88	04:11	100	31	13	18
25	26	58	59	117	89	04:10	94	30	14	16
26	16	57	60	117	104	04:10	97	29	9	20

Note. AC = Auditory Comprehension; EC = Expressive Communication; SS = Standardised Score; LAE = Language Age Equivalent; SB ABIQ = abbreviated IQ score; Raw SB5 T = Raw non verbal score plus raw verbal score; SB5 Raw NV = Raw non-verbal score; SB Raw V = Raw verbal score.

A Comparison between Experiment 4b and 4c: A Track Order Effect

Participants in Experiment 4b received the ‘easy-to-difficult’ sequence of Stage 3 TARPA tracks, while participants in Experiment 4c received the ‘difficult-to-easy’ sequence of Stage 3 tracks. Children in Experiment 4b had a mean age of 73.1 months (SD = 13.63) while those in Experiment 4c had a mean age of 69.3 months (SD = 15.8). Those in Experiment 4b had a mean total SB5 score of 36.5 (SD = 9.1) while those in Experiment 4c had a mean mean total SB5 score of 38.3 (SD = 9.3).

A Mann-Whitney *U* test was conducted to evaluate whether children in Experiment 4c (‘difficult-to-easy’ group) scored significantly lower on the TARPA, on average, than children in the Experiment 4b (‘easy-to-difficult’ group). The results of the test were in the expected direction and significant, $z = 2.04$, $p < 0.05$. Children in the ‘difficult’ group had an average rank of 7.8, while those in the ‘easy’ group had an average rank of 13.3.

The effect of track order is also evidenced by the contrast between correlations seen in Experiment 4b (‘easy-to-difficult’) and Experiment 4c (‘difficult-to-easy’). In both experiments scores on the TARPA were compared with the SB5 and the PLS-4. In Experiment 4b the TARPA correlated significantly with total scores on the SB5 $\rho(8) = 0.65$ ($p < 0.05$); in Experiment 4c the correlation between the TARPA and total SB5 scores was weaker and non significant $\rho(8) = 0.57$ ($p > 0.05$). With respect to the PLS-4, there is unambiguous difference between correlations between experiments; in Experiment 4b the TARPA correlated significantly with the PLS-4 $\rho(7) = 0.68$ ($p = 0.05$), while in Experiment 4c the correlation between the TARPA and PLS-4 scores was very clearly non significant $\rho(5) = -0.1$ ($p > 0.05$). It must be noted that all participants were tested on the SB-5 while not all participants in Experiments 4b and 4c were eligible for testing on the PLS-4. It seems that the SB5 correlation in Experiment 4c was boosted by a number of older participants (with high SB5 scores) who were apparently less affected by the track order. The participants in the PLS-4 correlation were younger and had lower scores on the SB5 and this group appeared to be more affected by receiving the 'difficult-to-easy' track order. These points will be further elaborated upon in the next section when individual performances are discussed.

Individual TARPA performance

A detailed breakdown of individual performances on Stage 3 of the TARPA for participants in Experiment 4c is shown in Table 22. They all passed all levels of Stages 1 and 2. They were then exposed to the 'difficult-to-easy' track order in Stage 3. Six participants (P21, P22, P23, P24, P25 and P26) in this group aged between 4 years 6 months and 5 years 4 months demonstrated a very similar pattern of responding and will be discussed together.

All but one of these children passed Level 2 (mutual entailment) in one track. P22 failed to pass Level 2 in any track and P24 passed it in two tracks. With regard to Level 3 (combinatorial entailment) only two children (P21 & P25) passed this and they did so only in the visual-visual track. In the visual-visual track, P25 from this group also passed Level 4 (transformation of function).

Given the age of these children and their PLS-4 scores their performance on the TARPA is quite poor. Performance appears to be affected by the sequence in which tracks were presented. Only P24 passed Level 2 (mutual entailment) in any of the first three tracks and he did so for only one track (auditory-visual 2).

Participants tended to show an improvement in performance on the last track on which they were tested which was the visual-visual track. All but one passed

Level 2 (mutual entailment), two passed Level 3 (combinatorial entailment) and as mentioned above P25 also passed Level 4 (transformation of function) in this track.

Interestingly some of the children actually commented on the 'difficulty' of the TARPA. For example, throughout exposure to the first three tracks, P21 remarked that the task was 'too hard' and 'that it should be for children in sixth class'. However, on the visual-visual track she commented that she now understood and that it was 'easy'.

Four other children (P17, P18, P19 & P20) in this experiment also received this sequence of tracks. These children were older than those in the group just discussed and ranged in age from 5 years 9 months to 8 years 7 months. This group did not show the very clearly diminished performance that was evident in the younger group. All of these children passed Level 2 (mutual entailment) in at least two tracks and two of them (P17 and P20 who had the highest total SB5 scores in this experiment, with scores of 57 and of 50 respectively) passed this level across all tracks.

In terms of combinatorial entailment, there was some variation across this older group. All passed Level 3 (combinatorial entailment) in at least one track. P17 and P18 passed at this level in a number of tracks (P17 in three and P18 in two). One participant (P20) passed Level 4 (transformation of function) in the auditory-visual 1. The impact of track sequence on overall TARPA performance wasn't as obvious in this group. However, when their performance is compared with that of children of a similar age and ability (as measured on the SB5 and/or PLS-4) in Experiment 4a and 4b some perceptible differences are observed. For example, P17's score was poor when compared with those of P3 and P5 in Experiment 4a who were of a similar age and has similar SB5 scores. P17 did not pass Level 4 (transformation of function) on any track while P3 and P5 showed patterns consistent with transformation of function in three out of four tracks. Interestingly, P3 and P5 both passed Level 4 (transformation of function) in their first track which was visual-visual and went on to show this pattern in a further two tracks. Conversely, P17 performed weakest on his first track, the auditory-auditory track and subsequently did not show transformation of function in any track. It seems as though for P3 and P5 showing all three properties of DRR (i.e., ME, CE and TOF) in the visual-visual modality facilitated similar level of performance in subsequent and arguably more ‘difficult’ tracks.

Table 22

Experiment 4c.Performances of Individual Participants for Each Level within Stage 3 of the TARPA

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26
Auditory-Auditory										
L1 CD	P	P	P	P	P	P	F	P	F	P
L2 ME	P	F	P	P	F	F	n/a	F	n/a	F
L3 CE	F	n/a	F	F	n/a	n/a	n/a	n/a	n/a	n/a
L4 TOF	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Auditory-Visual 2										
L1 CD	P	P	P	P	F	F	F	P	P	F
L2 ME	P	F	F	P	n/a	n/a	n/a	P	F	n/a
L3 CE	P	n/a	n/a	F	n/a	n/a	n/a	F	n/a	n/a
L4 TOF	F	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 22. Continued.

Experiment 4c. Performances of Individual Participants for Each Level within Stage 3 of the TARPA.

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)										
Part	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26
Auditory-Visual 1										
L1 CD	P	P	P	P	F	P	P	F	P	F
L2 ME	P	P	F	P	n/a	F	F	n/a	F	n/a
L3 CE	P	P	n/a	P	n/a	n/a	n/a	n/a	n/a	n/a
L4 TOF	F	F	n/a	P	n/a	n/a	n/a	n/a	n/a	n/a
Visual-Visual										
L1 CD	P	P	P	P	P	P	P	P	P	P
L2 ME	P	P	P	P	P	F	P	P	P	P
L3 CE	P	P	P	F	P	n/a	F	F	P	F
L4 TOF	F	F	F	n/a	F	n/a	n/a	n/a	P	n/a

Summary

The key pattern of results for Experiment 4c was finding a strong order effect. Specifically, receiving the difficult-to-easy track order appeared to reduce level of TARPA performance for participants in comparison with that of another earlier group who received the reverse order. This effect is clearly evident at both a group as well as an individual level. Possible explanations for this effect as well as its implications are discussed in the General Discussion (below). In light of the strength of the order effect, strong correlations with the alternative measures (i.e., SB5 & PLS-4) were not predicted in this particular experiment. The correlation between TARPA scores and the PLS-4 scores was unambiguously outside significance and this in clear contrast with the correlation between these same measures seen in Experiment 4b. However, the correlation with the total SB5 was strong and borderline significant. As mentioned above, it is likely that the 'track order' effect is moderated by a participant's language/cognitive ability. More specifically, across this research, children with higher scores on tests of linguistic and cognitive ability have been more likely to pass levels on difficult tracks (e.g., auditory-auditory); while children with lower language/cognitive scores have been less likely to pass levels on these tracks. It seems reasonable then that the children who fall into the latter group will be more likely to be negatively affected by receiving a 'difficult-to-easy' track order. Potential evidence for this is seen in the correlation between the TARPA and SB5 in Experiment 4c; it appears that this correlation was boosted by a number of older participants (with high SB5 scores).

General Discussion

At its current stage of development, the TARPA protocol is primarily a means of assessing emergent coordinate DRR or relational framing. The aim of the current program of research is to evaluate the TARPA as a tool in this regard and the current study has extended this research program in a number of important respects.

DRR and Linguistic and Cognitive Ability

Experiment 4a and 4b are important in that they correlated the current version of the TARPA with a standardised measure of intellectual functioning (i.e., SB5). This is the first time that the TARPA has been correlated with a measure of intellectual / cognitive functioning as opposed to a measure of language. Where participants in experiments 4a and 4b were eligible they were also tested on the PLS-4 and so these experiments afforded a novel degree of insight into how the TARPA compares with other measures as a means of gauging intellectual performance. A key outcome was that TARPA was found to correlate significantly with both the SB5 and PLS-4. The correlations with the PLS-4 replicate the findings of Experiments 1-3 while the correlations with the SB5 further underline the relevance of DRR and of the TARPA as a measure of this repertoire to a broader range of cognitive abilities.

In this experiment, further correlational analyses were also conducted using certain portions of TARPA performance (i.e., individual Stage 3 tracks) and the SB5 and PLS-4. Interestingly, these analyses revealed that the auditory-visual 2 track correlated most strongly of the four tracks with verbal scores on the SB5 while the correlation with the visual-visual track was the weakest. Similarly, analyses revealed that the auditory-visual 1 track correlated most strongly of the four tracks with the total scores on the PLS-4, while again the correlation with the visual-visual track was the weakest. These are interesting finding, explanations for which might include the foundational importance of audio-visual relations in language learning and the prevalence of audio-visual relations within typical socio-verbal contexts.

The current study also advanced on previous work by examining trials to criterion (i.e., total training and testing trials to pass a certain level in the TARPA) as a measure of TARPA performance. Nine participants passed all levels in the visual-visual track. A correlational analysis revealed a strong and significant correlation between trials to criterion on visual-visual track and verbal scores on the SB5. This has implications for the development of the TARPA protocol. When ‘points’ scored is used as the sole measure of TARPA performance this group of nine participants are considered equal, but trials to criterion and its correlations with the SB5 suggest that there are important differences between these participants. Participants with higher verbal scores on the SB5 take significantly less trials to achieve criterion on the track. An analysis of TARPA performance that takes account of trials taken to reach criterion may be particularly informative when dealing with an apparently more homogeneous sample.

In future work trials to criterion could be used to inform a more subtle breakdown of TARPA scoring. Previous versions of the TARPA (e.g., Moran et al., 2011; Chapter 2 of this thesis) allowed participants only one exposure to tests of derived relations (e.g., tests for mutual entailment, combinatorial entailment & TOF). Later versions (Experiments 3-4) have allowed participants a second chance on these tests. The current data suggest that participants who pass tested sections on their first attempt (i.e., take less trials to reach criterion) should receive more points for doing so. Other researchers have highlighted issues with relying on accuracy data alone in learning tasks (e.g., Binder, 1996) and in DRR research (e.g., Whelan & Schlund, 2013). While trials to criterion data helps us to differentiate between participants who met a pre defined criterion related based on accuracy, fluency (accuracy plus speed) may provide an even better measure of TARPA performance. Future work will need to investigate this and this avenue of research is particularly important as the TARPA is developed as a remediation tool.

The Effect of Stimulus Modality on Stage 3 Tracks

The current program of research has provided significant evidence of the effect of stimulus modality on derived relational performance. For example, throughout the research program, all children, whether typically developing (e.g., Experiments 2 & 3) or with ASD (Experiment 1) have tended to show lowest scoring on the auditory-auditory track. With respect to performance on the other tracks, differences have been more marginal and less consistent, but there are still detectable patterns. For instance, children in Experiments 1 and 2 were most likely to show

DRR consistent performance in the visual-visual track, while children in Experiment 3 tended to show higher levels of performance on the auditory-visual 1 track.

Experiments 4a and 4b of the current study have provided further data that is relevant to this issue. One analysis reported earlier compared aggregate number of levels passed across the four tracks. This data indicated that participants tended to pass most levels in the visual-visual track (53) and the lowest number of levels on the auditory-auditory track (31). Tracks were also compared in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function). Nine participants showed all the derivation patterns in the visual-visual track while between three and four participants showed these patterns in the other tracks (AV1 = 3; AV2 = 4 & AA = 3). In this study, both analyses indicated that the visual-visual track was the highest scoring track relative to tracks involving auditory stimuli. However, while the analysis that looked at the number of levels passed indicated the exclusively auditory track to be the lowest scoring (replicating previous work); the same number of participants passed the auditory-auditory track outright as did the two auditory-visual tracks. In the previous experiments, relatively less participants passed this track outright. However, as mentioned previously there were a number of older participants in this study who demonstrated all three properties of DRR (i.e. mutual entailment, combinatorial entailment and transformation of function) across three tracks each (e.g., P3, P4, P5 and P7).

Previous research in this area (e.g., Green, 1990; Smeets & Barnes-Holmes, 2005) has suggested that persons with developmental disabilities and typically developing young children demonstrate DRR more readily in the auditory-visual than in the visual-visual domain. The current study provides data that bears on this question; more participants in Experiments 4a and 4c showed patterns consistent with DRR when the relations involved all visual stimuli (i.e., the visual-visual track in Stage 3). However, the differences between the visual-visual track and the auditory-visual tracks found in this study are not very substantial and when this research is considered as a whole these difference are inconsistent (for example, in Experiment 3, there were more levels passed in the auditory-visual 1 track than the visual-visual). What does seem clear, however, is that children tend to show better DRR performance when relations involve at least some visual stimuli in comparison with relations that involve only auditory stimuli; as the auditory-auditory track has consistently been revealed to be the lowest scoring track.

A Track Order Effect

Based on the unusual patterns of responding of one participant in Chapter 3 it was speculated that receiving Stage 3 tracks in a particular order might affect overall performance on the TARPA. This participant despite having advanced language abilities performed very poorly on the TARPA. It was thought that starting with the apparently most difficult track (e.g., auditory-auditory) and progressing through the other tracks to the apparently least difficult (e.g., visual-visual) may have resulted in a reduced overall score on the protocol for this participant. The results of the current study lend support to this proposition; six out of ten children in Experiment 4c who received the ‘difficult-to-easy’ track order showed a very clear decrement in performance relative to their age, level of functioning and the performance of others under the same conditions. No child of a comparable age or level of functioning in Experiment 4b (who received the ‘easy-to-difficult’ track order) showed these weak patterns of responding on the protocol. This issue hasn’t been extensively investigated in DRR research; only one previous study showing a similar effect was identified (i.e., Connor et al., 2009) and while O'Connor et al. (2009) found an 'order-of-difficulty' effect, their study was different in that the difficulty level was based on the variables of nameability and familiarity of stimuli and not stimulus modality as was the case in the current work.

As discussed previously, this effect appears to be somewhat moderated by age and level of functioning. The four older children in Experiment 4c did not demonstrate the same breakdown in performance that was evident in the other children in this group but arguably when they are compared to similar participants in Experiment 4a they appear to have performed at a level lower than would be expected. Receiving the apparently most difficult sequence of tracks appears to render the protocol more difficult for participants. Why might this occur? One explanation provided by RFT suggests that correspondence in a relational network (i.e., verbal coherence) functions as a powerful reinforcer for relational activity itself (e.g., Roche, Barnes-Holmes, Barnes-Holmes, Stewart, & O'Hora, 2002). In this study participants who received the 'easier' presentation of tracks showed appropriate derivation patterns early on, and while these derivation patterns were shown under conditions of extinction (there was no contingent reinforcement for responses), it is possible that 'verbal coherence' functioned as a reinforcer which subsequently strengthened responding across subsequent tracks. Indeed, the overt verbal behavior of some of the children during derivation tests seemed to indicate that showing the correct derivation patterns was functioning as some sort of conditioned reinforcer. For example, during derivation tests children often made statements such as 'Yes- I'm getting them right!' or 'I'm winning!'; these statements seem to indicate that children were self-discriminating that they were doing well.

Conversely, the experience of receiving the 'difficult' sequence of tracks and getting things wrong may perhaps have punished responding. At an applied level, this finding would lend further empirical support to the importance of designing educational programmes that carefully match learners current level of functioning. These results would suggest that engaging a child in a difficult task where they do not contact reinforcement (extrinsic or intrinsic) hinders performance and learning on subsequent easier tasks.

Conclusion

The purpose of this research programme is to develop the TARPA as a methodology for assessing DRR and its precursor skills. The current study extended this work in a number of important directions; including further refining the TOF test and showing correlations between the TARPA and cognitive as well as linguistic abilities. The inclusion of the SB5 indicated the relevance of the TARPA to a broader range of cognitive skills and lends further support to the TARPA as a RFT-based protocol for assessing DRR. The current study also provided new insight with respect to track order effect. In particular, the finding of a robust track order effect provides an explanation for the anomalous performance of a child in Experiment 3.

Chapter 5

Testing A Brief Version of the Adjusted TARPA with Children with ASD

One key aim of this work has been to contribute to the development of a protocol for measuring DRR, which RFT sees as the key skill underlying language that might be useful in the measurement of this repertoire in children with ASD and/or developmental delay. By developing a reliable and valid protocol for measuring derived relational ability in children who might be deficient in this ability, practitioners might have access to a means of identifying such deficiency and targeting the repertoire at issue with appropriate training.

Given the importance of this aim, it was decided that the final study in this thesis would use the TARPA protocol, which has been developed over the course of a number of the previously reported studies, to measure DRR in a relatively large sample of children ASD and to correlate DRR measured in this way with language and cognitive performance assessed using the standardised assessments also featuring in some of these previous studies. Given the importance of the previous work that has contributed to the development of the protocol to be used in this final study, it seems worthwhile to first recap on some of this research.

The first published study to use the TARPA also focused exclusively on children with ASD. This study by Moran et al. (2010) measured the performance of a small sample of five children on a preliminary version of the TARPA as well as on the Vineland Adaptive Behavior Scales (VABS; Sparrow, Cicchetti & Balla, 2005). Findings showed a strong and significant correlation between these instruments. Data from this study additionally showed that the communication subscale of the VABS correlated highly and at borderline significance with the TARPA, while other subscales showed lower and more clearly non-significant correlations, coherent with the theoretical link between DRR and language. Moran et al (2010) also showed variation in DRR performance based on stimulus modality. For instance, a number of participants showed weaker responding on TARPA sections involving exclusively auditory stimuli than on sections involving at least some visual stimuli, irrespective of the strength of their overall performance.

The experiments presented in Chapter 2 of the current thesis advanced this early work in a number of ways. First, they used a more comprehensive and systematic TARPA protocol; for example, this expanded version included several different tracks that tested for derivation of relations in different modalities. Second, they correlated the TARPA with a performance-based measure of language ability, namely the Preschool Language Scale (4^{th} edition) (PLS-4; Zimmermann, Steiner & Pond, 2002) and in a much bigger and more diverse sample; more specifically, in ten children with ASD (Experiment 1) and thirteen typically developing children (Experiments 2). The results revealed strong and significant correlations between the TARPA performance and the PLS-4. Third, in order to investigate the appropriateness of the ordering sequence of TARPA sub-stages, a formal hierarchical analysis of the protocol was conducted (e.g., Krus, Bart, & Airasian, 1977; Marion et al., 2003) which showed that the structure was, for the most part, robust. In addition, findings shed further light in relation to the issue of modality: Experiments 1 and 2 provided further evidence that children with ASD (across all stages of the TARPA) and typically developing children (across Stage 3) show better performance with visual than with auditory stimuli.

Experiment 3 with typically developing children focused on improving the validity of the TARPA and this involved two procedural modifications. One was the number of exposures to tests of derived relations. Multiple exposures to derivation tests are common in the equivalence literature (e.g., Cassidy, 2008; O'Connor, Rafferty, Barnes-Holmes & Barnes-Holmes, 2009). However, in the early studies presented in this thesis participants were allowed only one exposure to each derivation test. This was changed in Experiment 3 so that participants were provided with one additional exposure to any derived test that they failed. The single subject data from this and subsequent experiments suggested that this adjustment allowed a more robust test of DRR. The second adjustment concerned the transformation of function (TOF) test. Hierarchical analysis performed in Chapter 2 indicated a number of instances in which participants passed the TOF test despite failing the previous test of combinatorial entailment, contrary to the expected order. To make passing on this basis less likely, the protocol was changed so that participants had to show TOF with two separate sets of stimuli. Subsequent testing with this modification showed no further such disconfirmations.

Experiment 4, also with typically developing children, advanced previous work in a number of respects. First, this study made a further change to the TOF test. This involved presenting it before the test for combinatorial entailment, an arrangement which has been argued to be more methodologically sound (see Dymond & Rehfeldt, 2000; Rehfeldt & Hayes, 1998; Barnes & Keenan, 1993). The findings from Experiment 4 seemed to support this change. Second, Experiment 4 used a test of intellectual/cognitive functioning (the Stanford-Binet 5 [SB5]; Roid, 2003) in TARPA based research for the first time. A key aspect of RFT is the idea that DRR not only underlies language as narrowly understood but that it underlies a whole gamut of cognitive abilities including those measured by traditional tests of intellectual functioning (e.g., Cassidy et al., 2010; Cassidy et al., 2011). The strong and significant correlations between the TARPA and (i) verbal and (ii) non verbal IQ scores presented in Chapter 4 provided support for this position and for the TARPA as a measure of DRR. Third, Experiment 4 investigated the effect of the order of presentation of TARPA tracks on participants overall score, a strong track order effect was evident in that presentation of the auditory track first seemed to negatively affect subsequent performance.

As indicated above, the aim of the study reported in this chapter was to use an adjusted version of the TARPA protocol, modified over the course of a number of the experiments reported in this thesis, to assess DRR in a relatively large sample of children with ASD. The current study employed a larger group of participants (n = 20) than either of the previous experiments with children with ASD. As well as adding power to the correlational analyses, this larger sample size also enabled a more intensive analysis of the hierarchy of levels within the early stages of the protocol to be conducted. As mentioned above the hierarchy of the protocol has been assessed previously and certain weaknesses have been addressed. However, only Experiments 1 and 2 have provided data in terms of Stages 1 and 2 of the protocol and this data was limited because only certain patterns of responding contribute to an order analysis. Specifically, data from participants who pass or fail both levels in any one comparison doesn't add to the analysis and most children in Experiments 1 and 2 passed all levels in Stages 1 and 2. Thus, in the order analysis presented in Chapter 2, the number of relevant comparisons is low for these stages and therefore a number of these comparisons were statistically non-significant (i.e., in certain cases the hierarchy was not confirmed). For comparisons in these stages (e.g., S1L1 and S1L2; S1L2 and S1L3; etc) the question arises as to what further data might indicate regarding the predicted order. Also, the order analysis presented in Chapter 2, was based on the amalgamation in the cases of Stages 1 and 2 of visual and auditory level data. The increased sample size in this study allowed a more intensive analysis of the hierarchy within these stages and looked at the visual and auditory levels separately, thus focusing on the difference between auditory and visual stimuli.

The study reported here also used a more extensive range of assessment tools in the context of testing children with ASD. In addition to the VABS and PLS-4, which have been used in previous TARPA research with this population, it used a measure of intellectual / cognitive functioning (SB5; Roid, 2003). The results of Experiment 4 with typically developing children indicated that the TARPA correlated not only with language but also with intellectual ability more broadly (e.g., non verbal scores on the SB5) and so this study extended this line of inquiry to children with ASD. A measure of autism severity (the Gilliam Autism Rating Scale, 2nd ed.; Gilliam, 2006) was also included in this study.

The expanded battery of assessment tools employed in this study, as well as enabling a more extensive correlational analysis with the TARPA; also allowed a comparative analysis of different correlations. More specifically, comparisons could be made between correlations involving the TARPA (e.g., TARPA & PLS-4) versus correlations involving the other measures (e.g., PLS-4 & SB5). This allowed additional information as regards the validity of the TARPA. A similar analysis was conducted in Chapter 4; however it was limited because not all participants were tested on all measures (e.g., some children were ineligible for testing on the PLS-4). All participants in the current study were tested on all measures which provided a better basis for comparisons.

One other aspect of the current study was that it involved administering a brief version of the TARPA; specifically only the auditory-visual (AV) 1 track of the protocol was used (together with Stages 1 and 2). The key reason for doing this was to speed up test administration so as to enable a greater sample size to be tested for correlational purposes. AV1 was chosen because in investigations of DRR as a key process underlying language, the AV tracks are particularly useful because normal language development obviously involves both the visual and auditory modalities in combination. In addition correlational analysis in Experiment 1 which involved children with ASD indicated that out of the Stage 3 tracks involving auditory stimuli, the AV1 track showed the highest levels of correlation with language ability as measured by the PLS-4. A correlational analysis in Experiment 4 also showed strong and statistically significant correlations between the AV1 only portion of the TARPA and PLS-4 scores; and borderline significant correlations with the SB5.

Method

Participants

Prior ethical approval for recruitment of participants was obtained from the National University of Ireland Galway Research Ethics Committee. A prior power analysis indicated that a sample of sixteen or more participants was needed to have 80% power for detecting a 0.6 sized effect at the traditional alpha value of .05. While the results of Moran et al. (2010) and the studies reported in this thesis indicated a large effect size; the current study used a brief version of the TARPA, thus it was speculated that this might produce a weaker effect size.

Twenty children (17 male, 3 female; age range 3 years 8 months - 15 years 0 months; see also Table 23), each of whom had an independent diagnosis of autism provided by a licensed psychologist, were recruited through special schools, ABA classes and through acquaintances of the experimenter (see Appendix E for a breakdown of individual participants school placements). Fifteen were attending a specialised ASD class within a mainstream school, three were being taught in the context of a home program and the remaining two were attending a specialised school for learners with ASD. None of the children had previously taken part in research. Consent for conducting the study was obtained from the school principle or the director of education in the respective schools. Parental consent was obtained for each child who participated and verbal consent (where possible) and assent was also obtained from each of the participants.

Materials

TARPA. The TARPA has been described elsewhere (see Chapter 2). Only the auditory-visual 1 track (see Appendix A for details) from Stage 3 of the TARPA protocol was used in this study.

PLS-4/SB5. The PLS-4 and SB5 have also been described in detail elsewhere. Two new measures, the Vineland Adaptive Behavior Scales (VABS; Sparrow et al., 2005) and the Gilliam Autism Rating Scale (2nd Edition; GARS; Gilliam, 2006) were used in this study and will now be described.

Vineland Adaptive Behavior Scales. The VABS is a rating scale administered to caregivers to assess a child's daily living skills in three domains: communication (receptive, expressive and written), daily living skills (personal, domestic and community) and socialisation (interpersonal relationships, play and leisure time, and coping skills). The VABS has excellent levels of split-half, inter-rater and test-retest reliability for each domain (Sparrow et al., 2005). A four point rating scale is used to score the response that best describes the behaviour of the child that is being assessed on any given item (i.e., DK = don't know, 0 = never, 1 = sometimes and 2 = usually). An example of a VABS item (from the communication domain) is 'points to at least three major body parts when asked (for example, nose, mouth, hands, feet, etc.).'

Gilliam Autism Rating Scale - 2 edition. The GARS-2 is a norm-referenced instrument that consists of 42 items describing the characteristic behaviours of individuals with ASD. The items are grouped into three subscales (stereotyped behaviours, communication and social interaction). The respondent uses a likert scale to rate the frequency of the subscale items as 0 (never observed), 1 (seldom observed), 2 (sometimes observed), or 3 (frequently observed). The obtained raw scores are then converted into standard scores (mean = 10, SD = 3) which, when totalled, provide an Autism index (mean 100, SD= 15) for each individual. An Autism Index of 85 or higher, means the possibility of the individual having autism is ‘very likely’, a score between 70 and 84 means that it is ‘possible’ that the individual has autism, and a score of 69 or lower means that it is unlikely that the individual has autism (Gilliam, 2006).

Procedure

The first session involved PLS-4 assessment. The PLS-4 was administered to the child by the experimenter. This assessment was usually conducted in one sitting (with breaks) that lasted on average about 35 minutes per child. In the second session, the SB5 was administered and this assessment lasted about 25 minutes. TARPA assessment began in the next session and was administered in the same manner as in Experiment 4. Caregivers completed the VABS and GARS for each child in their own time and returned the completed questionnaires to the experimenter. Session lengths were dependent on duration of on-task behaviour, the length of time for which each child was available in that particular session and TARPA progress.

Results & Discussion

Demographic Information and TARPA performance parameters are shown in Table 23.

Table 23

Demographic Information and TARPA Performance Parameters

Pt. No.	Age	Sex (Y: M)	No . Sessions	Mean Session Length (mins)	Total Testing Length (mins)	Highest Level Passed	Highest Level Tested
1	04:02	m	4	15	60	S3L1	S3L2
2	07:02	f	5	17	85	S3L1	S3L2
3	07:07	m	3	15	45	S3L4	S3L4
4	06:08	m	3	22	66	S3L1	S3L2
5	06:03	m	4	13	52	S3L2	S3L3
6	08:11	m	2	23	46	S3L3	S3L4
7	06:01	m	4	14	56	S2L3 V & A	S3L1
8	15:00	m	4	11	44	S3L1	S3L2
9	05:05	f	4	16	64	S1L3 V	S2L3
10	10:02	m	3	14	42	S3L2	S3L3
11	05:09	m	4	13	52	S3L2	S3L3
12	07:01	m	3	18	54	S3L4	S3L4
13	06:10	m	4	12	48	S2L3 V	S3L1
14	04:08	m	4	11	44	S3L1	S3L2
15	09:04	m	4	10	40	S3L3	S3L4
16	07:10	m	4	8	32	S3L1	S3L2
17	05:09	m	3	20	60	S2L3 V & A	S3L1
18	09:07	f	3	17	51	S3L1	S3L2
19	08:05	m	4	13	52	S3L2	S3L4
20	03:08	m	6	10	60	S1L3 V	S2L2 V & A

Note. Y = Year; M = Month; S = Stage; L = Level; A = Auditory; V = Visual.

Scores for the TARPA, PLS-4, SB5, GARS and VABS assessments are provided in Table 24. Scores for the TARPA were calculated as per the scoring table provided in Appendix B. Scores for the PLS-4 (raw and age equivalent), SB5 (total, verbal and non-verbal), VABS and GARS were calculated per their respective manuals (Zimmermann et al., 2002; Roid, 2003; Sparrow, Cicchetti & Balla, 2005; Gilliam, 2006).

Correlations between the TARPA and other measures

All correlations are based on raw (as opposed to age normative) scores. Table 25 displays a matrix of Spearman rank order correlations among the key measures. This shows a correlation between the TARPA and PLS-4 total scores ($\rho = 0.72$; $p < 0.001$) as well as with both auditory comprehension ($\rho = 0.68$; $p < 0.01$) and expressive communication ($\rho = 0.69$; $p < 0.01$) PLS-4 subscales. The TARPA was also found to be correlated with total scores (i.e., verbal plus non verbal scores) on the SB5 ($\rho = 0.79$; $p < 0.01$) and with both the SB5 subscales. The correlation with the non-verbal score was 0.69 ($p < 0.01$) while that with the verbal score was 0.68 ($p < 0.01$). A Spearman's rank correlation test also revealed a highly significant correlation ($\rho = 0.6$; $p < 0.01$) between performance on the TARPA and communication subscale of the VABS.

Table 24

Scores for GARS, VABS, TARPA, PLS-4 & SB5 assessments.

Pt. Number	GARS Autism	VABS Comp	VABS Comm. Total Raw	TARPA Score	PLS-4 Raw AC	PLS-4 Raw EC	PLS-4 Total	PLS-4 SS	PLS-4 LAE	SB5 ABIQ	SB5 T Raw	SB5 Raw NV	SB5 Raw V
1	109	64	38	9	34	36	70	55	02:03	82	18	6	12
2	100	73	128	6.66	37	42	79	NA	02:07	52	24	8	16
3	91	69	134	20	57	61	118	NA	04:11	67	34	12	22
4	94	MD	MD	6.33	53	43	94	NA	03:04	55	24	8	16
5	66	78	130	13	55	61	117	59	04:10	94	35	18	17
6	89	72	113	17	41	40	81	NA	02:08	70	34	21	13
7	96	66	120	6	42	39	81	55	02:08	55	18	6	12
8	102	36	59	8	43	37	80	NA	02:08	47	29	16	13
9	79	59	55	1.66	28	30	58	55	01:09	70	19	11	8
10	89	MD	MD	13	58	64	122	NA	05:04 - 05:05	52	35	16	19

Note. MD = Missing data; NA = not applicable; AC = Auditory Comprehension; EC = Expressive Communication; SS = Standardised Score; LAE = Language Age Equivalent; ABIQ = Abbreviated Intelligence Quotient Score; SB5 T Raw = Raw non-verbal scores plus Raw non-verbal scores; Raw V = Raw Verbal; Raw NV = Raw Non-Verbal.

Table 24. Continued.

Scores for GARS, VABS, TARPA, PLS-4 & SB5 assessments.

Pt. Number	GARS Autism	VABS Comp	VABS Comm. Total Raw	TARPA Score	PLS-4 Raw AC	PLS-4 Raw EC	PLS-4 Total	PLS-4 SS	PLS-4 LAE	SB5 ABIQ	SB5 T Raw	SB5 Raw NV	SB5 Raw V
11	79	85	110	12.66	32	39	71	55	02:03	67	20	11	9
12	83	85	159	20	58	65	123	NA	05:06	79	38	15	23
13	65	74	129	5	35	41	76	NA	02:06	52	22	7	15
14	66	82	79	8.66	38	41	79	55	02:07	94	27	10	17
15	79	81	146	17	60	64	124	NA	05:07 - 05:08	61	37	15	22
16	72	72	125	8.666	49	55	104	NA	03:11	58	31	13	18
17	87	55	60	6	36	37	73	55	02:04	67	21	9	12
18	106	66	92	7.166	51	41	92	NA	03:03	55	28	14	14
19	76	74	145	13	61	55	116	NA	04:09	82	42	25	17
20	85	69	66	1.66	27	33	60	55	01:10	79	16	5	11

Note. MD = Missing data; NA = not applicable; AC = Auditory Comprehension; EC = Expressive Communication; SS = Standardised Score; LAE = Language Age Equivalent; ABIQ = Abbreviated Intelligence Quotient Score; SB5 T Raw = Raw non-verbal scores plus Raw non-verbal scores; Raw V = Raw Verbal; Raw NV = Raw Non-Verbal.

Table 25

Spearman's rank order correlations (n = 20) between assessment measures

	Total TARPA	SB5 T Raw	SB5 Raw NV	SB5 Raw V	PLS-4 Raw T	PLS-4 Raw AC	PLS-4 Raw EC
SB5 T Raw	ρ = .79***	---	---	---	---	---	---
SB5 Raw NV	ρ = .69**	ρ = .86***	---	---	---	---	---
SB5 Raw V	ρ = .68**	ρ = .83***	ρ = .47*	---	---	---	---
PLS-4 Raw T	ρ = .72**	ρ = .87***	ρ = .63**	ρ = .88***	---	---	---
PLS-4 Raw AC	ρ = .68**	ρ = .88***	ρ = .67**	ρ = .83***	ρ = .96***	---	---
PLS-4 Raw EC	ρ = .69**	ρ = .83***	ρ = .52*	ρ = .94***	ρ = .92***	ρ = .86***	---
VABS T Raw Comm.	ρ = .60**	ρ = .73**	ρ = .43	ρ = .79***	ρ = .81***	ρ = .73**	ρ = .91***

Note. "ρ" represents Spearmans Rank Order Correllation coefficient. Asterisks indicate a statistically significant correlation (2-tailed); * p < .05. ** p < .01. *** p < .001. SB5 T Raw = Raw non-verbal scores plus Raw non-verbal scores; Raw NV = Raw non-verbal score. Raw V = Raw verbal score. Raw T = Raw total score. Raw AC = Raw auditory comprehension score. Raw EC = Raw expressive communication score; VABS T Raw Comm = Raw total score on the communication domain of the VABS (note this correlation had an 'n' of 18).

A Comparison of Correlations

TARPA correlations with both the PLS-4 and SB5 can be compared with SB5-PLS-4 correlations as another approach to examining TARPA validity. The TARPA-PLS-4 (total scores) correlation (ρ (18) = 0.72; p < 0.01) compares well to the correlation between the PLS-4 (total scores) and scores on the verbal test of SB5 (ρ (18) = 0.88 (p < 0.001). A stronger correlation between the PLS-4 and the verbal subscale of the SB5 would be predicted given the similarity between these measures. The TARPA also correlated significantly with the verbal scores on the SB5 $\rho\,(18) = 0.68$ ($p < 0.001$) and this compares well to the correlation between the PLS-4 (total scores) and the scores on the verbal test of SB5 ($\rho\,(18) = 0.88$ ($p < 0.001$). A third correlation is also interesting; the TARPA correlated significantly with the non verbal scores on the SB5 $\rho\,(18) = 0.69$ ($p < 0.01$). The correlation between PLS-4 (total scores) and the non verbal test of the SB5 ($\rho\,(18) = 0.63$ ($p < 0.01$) was also significant. This is not unexpected but the fact that the TARPA correlated with all measures replicates the findings reported in Chapter 4 and further supports the conceptualisation of DRR as a core or underlying skill in language and cognition.

Trials to Criterion. Trials taken to pass certain levels of the TARPA may provide a more precise measure of DRR performance in some instances (e.g., where participants receive the same overall score on the protocol or where a group of participants pass the same stage). Number of trials taken to pass certain parts of the TARPA were correlated with PLS-4 (total) scores and SB5 (non-verbal and verbal subtest) scores. In this experiment, all 20 participants passed all three levels in the visual track of Stage 1. Total trials taken to pass these levels were correlated with the PLS-4 (total score) and the SB5 (non verbal and verbal subtests scores). Spearman's rank order correlation tests between total trials to criterion (S1 Visual) and PLS-4 total scores and the non-verbal subtest of the SB5 approached significance while the verbal subtest of the SB5 was more clearly non significant (see Table 26).

A Spearman's rank order correlation between trials to criterion (S1 Auditory) and scores on the PLS-4 (total) was statistically significant $\rho\,(17) = -0.47$ ($p < 0.05$). Trials to criterion (S1 Auditory) and scores on verbal subtest of the SB5 was also statistically significant ρ (17) = - 0.55 (p < 0.05). The correlation with the non verbal subtest of the SB5 was non-significant.

A Spearman's rank order correlation between trials to criterion (S2 Visual) and scores on the PLS-4 (total) was statistically significant ρ (17) = - 0.5 (p < 0.05). Spearman's rank order correlation tests between trials to criterion (S2 Visual) and scores on the SB5 (non verbal and verbal) were outside statistical significance. Spearman rank order correlations between trials taken to pass auditory levels within Stage 2 and scores on the PLS-4 and SB5 were also non-significant.

Table 26

Correlations between trials to criterion for visual and auditory sections of Stages 1 and 2 of the TARPA respectively and each of PLS-4 raw total score, SB5 raw non-verbal score and SB5 raw verbal score

	Trials to criterion Stage 1 V	Trials to criterion Stage 1 A	Trials to criterion Stage 2 V	Trials to criterion Stage 2 A
PLS-4 Total Raw	ρ= - .4 (n = 20)	ρ= -.47* (n = 19)	ρ= - .5* (n = 18)	ρ= - .25 (n = 18)
SB5 Raw Non Verbal	ρ= - .4 (n = 20)	ρ= - .35 (n = 19)	ρ= - .42 (n = 18)	ρ= - .31 (n = 18)
SB5 Raw Verbal	ρ= - .37 (n = 20)	ρ = - .55** (n = 19)	ρ= - .39 (n = 18)	ρ= - .28 (n = 18)

Note. V = Visual; A = Auditory. "ρ" represents Spearmans Rank Order Correllation coefficient. Asterisks indicate a statistically significant correlation (2-tailed); * p < .05. ** p < .01. *** p < .001.

Eight participants passed Levels 1 and 2 of the AV 1 track (i.e., showed patterns consistent with mutual entailment and as such scored 7 points on the track). Trials taken to pass the track were correlated with scores on the PLS-4 and the SB-5 (verbal and non verbal subtests). There were no significant correlations found. However, correlations between trials to criterion and verbal ability (i.e., scores on the PLS-4 and verbal portion of the SB5) were in the predicted direction. Specifically, Spearman's rank tests suggested a negative relationship between trials to criterion and verbal ability; as measured by the total scores on the PLS-4 $\rho\,(6) = -\,0.47$ ($p > 0.05$) and verbal subtest scores of the SB5 $\rho\,(6) = -\,0.45$ ($p > 0.05$). It wasn't possible to perform a similar analysis of trials to criteria data across higher levels of Stage 3 (i.e., Level 3 or Level 4), however, as an insufficient number of participants passed these levels.

Individual TARPA Performance

TARPA performance per participant is shown in Table 27. There was quite a bit of variation across participants. P9 and P20 passed the three visual sections of Stage 1 and passed the first two auditory levels within Stage 1. However, they both failed the third auditory level in Stage 1 (S1L3 auditory) and failed all subsequent levels in Stage 2 to which they were exposed. P13 passed all levels within Stages 1 and 2 with the exception of the test for generalised identity matching of auditory stimuli (S2L3 auditory). P7 and P17 both passed all levels in Stages 1 and 2 but failed to pass any level in Stage 3. All the remaining participants passed at least some levels of Stage 3 but there was also significant variation within this stage. P1, P2, P4, P8, P14, P16 and P18 passed Level 1 (arbitrary conditional discrimination) but failed to move beyond this level. While these participants passed Level 1 of Stage 3 all of them (with the exception of P1) failed at least one auditory level in Stage 1 or 2. Further discussion of this pattern will be provided in the next section which will focus on differences between performance in visual and auditory modalities.

P5, P10, P11 and P19 passed Level 2 in Stage 3 (mutual entailment) but failed to pass Level 3 (combinatorial entailment). Of these participants, P11 failed a level at an earlier stage involving auditory stimuli (S1L3 auditory); P6 and P15 both passed Level 3 (combinatorial entailment); and P3 and P12 passed Level 3 (combinatorial entailment) and Level 4 (transformation of function).

Table 27

TARPA Performances of Individual Participants on Each Stage and Level

Stage 1: Simple discrimination

Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Vis. L1	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Vis. L2	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Vis. L3	P	P	P	P	P	P	P	P	P	P	P	P	F/**P**	P	P	P	P	P	P	P
Aud. L1	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Aud. L2	P	P	P	F	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Aud. L3	P	F	P	F	P	P	P	P	F	P	F	P	P	F	P	F	P	F	P	F

Stage 2: Non Arbitrary conditional discrimination (i.e., formally SIMILAR stimuli)

Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Vis. L1	P	P	P	P	P	P	P	P	F	P	P	P	P	P	P	P	P	P	P	F
Vis. L2	P	P	P	P	P	P	P	P	F	P	P	P	P	P	P	P	P	P	P	F
Vis. L3	P	P	P	P	P	P	P	P	F	P	P	P	P	P	P	P	P	P	P	F
Aud. L1	P	F	P	F	P	P	P	P	F	P	P	P	P	P	P	P	P	P	P	F
Aud. L2	P	F	P	F	P	P	P	P	F	P	P	P	P	P	P	P	P	F	P	F
Aud. L3	P	F	P	n/a	P	P	P	F	F	P	P	P	F	P	P	P	P	F	P	F

Note. Vis. = Visual; Aud. = Auditory; L = Level; CD = Conditional discrimination; ME = Mutual entailment; CE = Combinatorial entailment; TOF = Transformation of function.. F followed by bold P indicates that the pt. showed serious attention deficits during the first test and passed after retesting with new stimuli.

Table 27. Continued.

Stage 3: Arbitrary conditional discrimination (i.e., formally DISSIMILAR stimuli)																				
Part	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Auditory-Visual 1																				
L1 CD	P	P	P	P	P	P	F	P	n/a	P	P	P	F	P	P	P	F	P	P	n/a
L2 ME	F	F	P	F	P	P	F	F	n/a	P	P	P	n/a	F	P	F	n/a	F	P	n/a
L3 CE	n/a	n/a	P	n/a	F	P	n/a	n/a	n/a	F	F	P	n/a	n/a	P	n/a	n/a	n/a	F	n/a
L4 TOF	n/a	n/a	P	n/a	n/a	F	n/a	n/a	n/a	n/a	n/a	P	n/a	n/a	F	n/a	n/a	n/a	F	n/a

Note. Vis. = Visual; Aud. = Auditory; L = Level; CD = Conditional discrimination; ME = Mutual entailment; CE = Combinatorial entailment; TOF = Transformation of function.. F followed by bold P indicates that the pt. showed serious attention deficits during the first test and passed after retesting with new stimuli.

Auditory versus Visual performance

Number of Levels passed. Aggregate number of visual levels passed was compared with aggregate number of auditory levels passed for both Stages 1 and 2. In Stage 1, the visual track showed 60 passes while the auditory showed 51, while in Stage 2, the visual track showed 54 passes while the auditory showed 44. These analyses indicated a general pattern whereby participants seemed to perform better on levels involving visual stimuli compared to levels involving auditory stimuli. At an individual level, some participants demonstrated a marked difference between visual and auditory modalities. For example, P4 passed levels that tested for a simple discrimination involving one, two and three comparisons but failed to demonstrate these discrimination abilities within the corresponding auditory levels. P9, P11, P14, P16, P18 and P20 passed Stage 1 Level 3 that tested for a simple discrimination involving three visual comparisons but failed to demonstrate this discrimination ability within the corresponding auditory level. P18 also passed all visual levels in Stage 2 but failed two corresponding levels in the auditory track (i.e., S2L2 and S2L3 auditory). P2 demonstrated the ability to match identical visual stimuli with two and three comparisons and showed generalised identity matching but failed these three levels in the auditory modality. Similarly, P8 and P13 showed generalised identity matching with visual stimuli (S2L3) but failed the same level in the auditory track.

Trials to criterion. A Wilcoxon signed-ranks test indicated a significant difference between trials taken to pass Stage 1 Visual compared with Stage 1 Auditory ($Z = -3.04$, $p < .001$). There was also a significant difference between trials taken to pass Stage 2 Visual and Stage 2 Auditory ($Z = -1.75$, $p < 0.05$).

Order Analysis

Order analysis (Kerr et al., 1977; Krus, Bart, & Airasian, 1977, Marion et al., 2003) was used in previous experiments to evaluate the hierarchical structure of the TARPA. A similar analysis was conducted in this experiment but with a particular emphasis on Stages 1 and 2. Table 28 compares each of the six TARPA levels in Stages 1 and 2 of the visual track with each of the other levels and Table 29 compares each of the six TARPA levels in Stages 1 and 2 of the auditory track with each of the other levels. Table 30 compares each of the four levels in Stage 3 of the auditory-visual 1 track with each of the other levels. In the case of each comparison it shows (i) number of confirmations [C] and disconfirmations [D] for those two levels, where a confirmation is a pattern in which the lower level was passed and the higher level was failed while a disconfirmation is a pattern in which the higher level was passed and the lower level was failed; (ii) a z-score measurement for that pattern; and (iii) a p-value corresponding to that z-score.

Table 28

Order analysis of TARPA Stages 1 and 2 for each of the levels in the Visual track based on the data provided by the 20 participants.

Visual – Stages 1 & 2										
Stage/Level	S1 L2		S1 L3		S2 L1		S2 L2		S2 L3	
	C	D	C	D	C	D	C	D	C	D
S1 L1	0	0	0	0	2	0	2	0	2	0
	Z = 0		Z = 0		Z = 1.41		Z = 1.41		Z = 1.41	
S1 L2	-		0	0	2	0	2	0	2	0
			Z = 0		Z = 1.41		Z = 1.41		Z = 1.41	
S1 L3	-		-		2	0	2	0	2	0
					Z = 1.41		Z = 1.41		Z = 1.41	
S 2 L1	-		-		-		0	0	0	0
							Z = 0		Z = 0	
S2 L2	-		-		-		-		0	0
									Z = 0	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 28 depicts the results of the order analysis conducted within the visual track of Stages 1 and 2. As can be seen, in the majority of the cells in Table 28 (i.e., 9/15) there were a small number of confirmations and no disconfirmations of the predicted order (i.e., participants typically passed these levels and as such their scores did not contribute to the analysis). None of these patterns is statistically significant even though the trend is in the predicted direction. The reason for this is the low number of recorded comparisons. Meantime, in the remaining, six cells in

Table 28 (i.e., S1L1 vs. S1L2; S1L1 vs. S1L3; S1L2 vs. S1L3; S2L1 vs. S2L2; S2L1 vs. S2L3; S2L2 vs S2L3) there were no confirmations or disconfirmations available.

Table 29

Order analysis of TARPA Stages 1 and 2 for each of the levels in the Auditory track based on the data provided by the 20 participants.

Auditory – Stages 1 & 2										
Stage/Level	S1 L2		S1 L3		S2 L1		S2 L2		S2 L3	
	C	D	C	D	C	D	C	D	C	D
S1 L1	1	0	8	0	4	0	5	0	6	0
	Z = 1		Z = 2.82**		Z = 2*		Z = 2.24**		Z = 2.45**	
S1 L2	-		7	0	3	0	4	0	6	0
			Z = 2.64**		Z = 1.73*		Z = 2*		Z = 2.45**	
S1 L3	-		-		0	4	0	3	2	3
					Z = - 2.0		Z = -1.73		Z = -0.45	
S2 L1	-		-		-		1	0	3	0
							Z = 1		Z = 1.73*	
S2 L2	-		-		-		-		2	0
									Z = 1.41	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * p < .05. ** p < .01. *** p < .001.

Table 29 depicts the results of the order analysis conducted within the auditory track of Stages 1 and 2. In the majority of the cells in Table 29 (i.e., 12/15) there were more confirmations than disconfirmations of the predicted order. In addition, in 9/15, there is a statistically significant effect in favour of the predicted hierarchy for the comparison between number of confirmations and number of disconfirmations. In the case of the cells where there is not a significant effect, the trend is in the predicted direction but the number of recorded comparisons is low and this is likely the factor responsible for the lack of significance. There are three cells in this category and these include the comparisons between S1L1 and

Table 30

Order analysis of TARPA Stage 3 for each of the levels in the Auditory Visual 1 track based on the data provided by the 20 participants.

	Stage 3 Auditory-Visual 1					
Levels	S3 L2 Mutual Entailment		S3 L3 Comb. Entailment		S3 L4 TOF	
	C	D	C	D	C	D
S3 L1	7	0	4	0	3	0
Cond. Dis	Z = 2.6457**		Z = 2*		Z = 1.732*	
S3 L2.	-		4	0	3	0
Mutual Entailment			Z = 2*		Z = 1.732*	
S3 L3	-		-		2	0
Comb. Entailment					Z = 1.4142	

Note. A confirmation (C) is recorded when a participant passes a lower level and fails a higher level. A disconfirmation (D) is recorded when a participant fails a lower level and passes a higher level. All significance values are based on a (one-tailed) right-tailed P-value. * p < .05. ** p < .01. *** p < .001.

S1L2; S2L1 and S2L2; S2L2 and S2L3. There are three cells in this table that show trends not in the predicted directions (i.e. there are more disconfirmations than confirmations). The particular comparisons involved are S1L3 versus S2L1; S1L3 versus S2L2 and S1L3 versus S2L3. This issue will be returned to in the general discussion.

Table 30 shows the results of the order analysis conducted within the auditory-visual 1 track of Stage 3. In all the cells in Table 30 (i.e., 6/6) there were more confirmations than disconfirmations of the predicted order. In addition, in all but one cell, there is a statistically significant effect in favour of the predicted hierarchy for the comparison between number of confirmations and number of disconfirmations. In the case of the one cell (i.e., S3L3 vs. S3L4) without a significant effect, the trend is in the predicted direction but the number of recorded comparisons is low and this is likely the reason for the lack of significance.

General Discussion

This study was similar to previous TARPA research in that it correlated a version of the protocol with alternative standardised measures of responding in children with ASD. However, it advanced previous work by (i) investigating correlations between TARPA performance and cognitive in addition to linguistic ability; (ii) indicating particular auditory deficits in children with ASD; (iii) conducting a more intensive hierarchical analysis of Stages 1 and 2 of the protocol.

Correlations between the TARPA and other measures

This study extended previous TARPA research with children with ASD by correlating a brief version of the protocol with a number of alternative measures of responding. A strong and significant set of correlations was found between the (i) TARPA and PLS-4; (ii) TARPA and SB5 and (iii) TARPA and the communication subscale of the VABS. The correlations with the VABS and PLS-4 replicate the findings of Moran et al. (2010) and of Experiment 1 of this thesis. The correlations between the TARPA and the SB5 support the relevance of DRR to cognitive / intellectual abilities. In addition, the correlations between the TARPA and these alternative measures (e.g., the PLS-4 and SB5) compare well with correlations between the alternative measures (e.g., between the PLS-4 and SB5) thus providing further support for the validity of the TARPA.

Trials to criterion (i.e., training and testing trials taken to pass a certain level) emerged from Experiment 4 as a possible important metric of TARPA performance.

The findings reported in this study also revealed a relationship between verbal ability (as measured by the PLS-4 and the verbal portion of the SB5) and trials to criterion on certain levels of the TARPA. Specifically, participants with higher scores on the PLS-4 and higher verbal scores on the SB5 took significantly less trials to achieve criterion on certain visual and auditory learning tasks in Stage 1 (simple discrimination) and Stage 2 (conditional discrimination based on physical similarity). Though non-significant in the context of a small sample the correlations found between verbal ability (PLS-4 and verbal subtest of SB5) and trials to criterion on Level 2 (mutual entailment) of Stage 3 also suggested that participants with higher verbal ability demonstrated DRR skills more readily. This is similar to the findings of Experiment 4 where verbal scores on the SB5 (and not non verbal scores) were found to correlate with DRR; children with more advanced verbal skills passed all four levels of the visual-visual track in fewer trials. In terms of the development of the TARPA this could inform a more subtle breakdown of points scored (as suggested in Experiment 4). At a practical level, it might suggest that clinicians should (i) test a child’s DRR skills and also (ii) analyse the trials taken to perform these tasks, as both these metrics appear to be very relevant to language and cognitive performance more generally.

Visual versus Auditory performance

The experiments presented in this thesis as well as previous work (e.g., Moran et al., 2010) have suggested that children (both typically developing as well as with ASD) show better performance with visual than with auditory stimuli. The data provided by this study allowed a further examination of this issue particularly at the earlier stages of the TARPA and the findings were similar. Participants tended to perform better on levels that involved visual stimuli and this was evident in terms of the (i) higher aggregate number of visual levels passed versus aggregate number of auditory levels passed (across Stages 1 and 2) and (ii) in the significantly lower number of trials taken to meet criteria on visual levels as compared with auditory levels (again across Stages 1 and 2). Moran et al. (2010) suggested that this pattern of results might be explained on the basis of the difference in behavioral control obtained by simultaneous and successive discriminations. In Chapter 2 it was also speculated that situations involving the types of auditory discrimination required in the TARPA protocol are not very common in the everyday socio-verbal environment and that it was possible therefore that at least some of the trials involving auditory stimuli might require more training than other types of trials because discrimination of auditory stimuli presented in the context of the TARPA is not as easily learned. Both these explanations remain plausible.

It is interesting that some patterns of responding seem to indicate particular auditory deficits in children with ASD. An analysis of the performances of individual participants identified some patterns that might be described as indicative of unbalanced visual and auditory repertoires. As mentioned previously, P4 passed levels in Stage 1 that required simple discriminations with one, two and three comparisons but failed the corresponding levels in the auditory track. Similarly, P2 passed all levels in Stage 2 in the visual track (including a test for generalised identity matching) but failed all three levels of Stage 2 in the auditory track. Research has identified the importance for more complex language skills of acquiring these basic skills across modalities (e.g. Chavez-Brown, 2005; Greer & Ross, 2008; Keohane & Greer, 2005; Marion et al. 2003). For example, Greer and Ross (2008) argue that developing ‘a capacity for sameness’ (i.e., matching across all senses including visual, auditory, tactile, etc.) is a fundamental step towards becoming verbal. Of course, Greer's definition of verbal is not identical with that of RFT (on which the TARPA is largely based); nevertheless, given the importance from an RFT point of view of non-arbitrary relational responding as a foundational element of learning relational framing, RFT would tend to agree with this suggestion.

Children in this study who performed in the advanced levels of the TARPA (e.g., at a minimum showed mutual entailed responding) tended to pass all levels within Stages 1 and 2 of the protocol. This makes sense since it seems likely that performing certain discriminations as assessed in Stages 1 and 2 across visual and auditory stimuli is fundamental to DRR ability and language. The TARPA offers an efficient and precise way of measuring these skills and with further development, it could also potentially play an important role in training these abilities. These implications will be further discussed in the next chapter.

Order Analysis

Order analysis (Kerr et al., 1977; Krus, Bart, & Airasian, 1977, Marion et al., 2003) was conducted and the large and varied data set provided by the ASD sample allowed a more thorough examination of the ordering of levels within Stages 1 and 2. The results of this analysis indicated that the hierarchical structure of the TARPA was generally robust.

The one exception to this pattern involved S1L3 within the auditory track. In four cases participants failed S1L3 and subsequently passed the higher level S2L1. S1L3 (auditory) assesses a participant's ability to make a simple discrimination involving three auditory stimuli; while S2L1 assesses a participant's ability to learn two non arbitrary conditional discriminations (i.e., match auditory stimuli) involving auditory stimuli. The four participants in question were retested on this level (S1L3 auditory) with novel stimuli and failed to pass, thus making it less likely that this was simply an attentional or other issue that might have only affected the first exposure. Three out of four of the affected participants also passed S2L2 and S2L3 within the auditory track accounting for the disconfirmations in the two further cells (i.e., comparisons S1L3 versus S2L2 and S1L3 versus S2L3). This is an interesting finding that seems to suggest that a three comparison discrimination involving all auditory stimuli is 'more difficult' than a two and three comparison conditional discrimination involving auditory stimuli. However, this is not a serious issue for the scoring or the hierarchical order of the TARPA since passing even one level in Stage 1 allows progress to Stage 2. The latter was originally done in order to allow for eventualities such as this.

Conclusion

The purpose of this research programme is to develop the TARPA as a methodology for assessing DRR and its precursor skills. From an RFT perspective these skills are fundamental to language and cognitive ability. The current study extended this work in a number of important directions. Previous work with children with ASD, using smaller sample sizes, has correlated performance on the protocol with the VABS (e.g., Moran et al., 2011) and with a performance based language assessment (e.g., Experiments 1). The work reported in this chapter had a much larger sample size than either of the previous studies and in addition compared TARPA performance with both linguistic and cognitive ability. The inclusion of the SB5 as a test of cognitive ability indicated the relevance of the TARPA to a broader range of cognitive skills and lends further support to this protocol as an RFT-based instrument for assessing DRR.

Chapter 6

General Discussion

Overview of Research Area and Aims

The aim of this thesis has been to test and further develop the TARPA, a protocol for the assessment of emergent DRR (DRR) abilities. From an RFT perspective, DRR is important because (i) it appears to be a key skill that underpins an array of linguistic and cognitive abilities (e.g., Rehfeldt & Barnes-Holmes, 2009; Stewart, Tarbox, Roche & O'Hora, 2013) including the ability to produce novel linguistic and cognitive performances (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004; Stewart, Barnes-Holmes & Roche, 2004; Wulfert & Hayes, 1988), and (ii) it appears to be amenable to training (i.e., it is an operant). These two points suggest that DRR should be a key target for behaviour analysts interested in improving linguistic and cognitive abilities in a number of different populations including typically developing children and children with Autism Spectrum Disorder (ASD) or other forms of developmental delay.

At its current stage of development the TARPA is primarily focused on multi-modality assessment of derived coordinate relations and the purpose of the research reported in this thesis has been to examine the validity of the protocol as a means of doing this and to contribute to improving the protocol in this respect. The evaluation of the TARPA involved: (i) examining for quantitative correlations between the TARPA and alternative linguistic and cognitive measures and (ii) testing the proposed hierarchical sequence of stages and sub-stages using order analysis (e.g., Krus, Bart & Airasian, 1975). The results of these analyses generally provided support for the validity of the TARPA. However, they also informed further adjustments to the protocol and additional avenues of investigation that were carried out as part of this research. This work was conducted over a series of five experiments. A brief overview of each experiment will now be provided.

Summary of Thesis Findings

The experiments presented in Chapter 2 aimed to validate the TARPA as a methodology for assessing coordinate DRR skills and their prerequisites. This was achieved by (i) correlating scores on the TARPA with performance on a standardised direct measure of language (i.e., the Preschool Language Scale 4th Edition; Zimmermann, Steiner & Pond, 2002) and (ii) formally evaluating the structure of stages and levels within the TARPA using order analysis (e.g., Krus et al., 1975). In an earlier study, Moran, Stewart, McElwee and Ming (2010) using a beta version of the protocol showed a correlation between TARPA performance and the Vineland Adaptive Behavior Scales (VABS; Sparrow, Cicchetti, & Balla, 2005) in several children with ASD. The work reported in Chapter 2 advanced this work by testing a more comprehensive and systematic version of the TARPA with ten children with ASD (Experiment 1) and 13 typically developing children (Experiment 2). Results showed (i) a strong and significant correlation between TARPA performance and PLS-4 full-scale and subscale scores and (ii) evidence to support the hierarchical structure of the TARPA. A number of further avenues for investigation were also indicated.

One issue which arose from the findings presented in Chapter 2 was the validity of the transformation of functions test. Though the hierarchical analysis performed in Experiments 1 and 2 generally supported the ordering of TARPA stages, it indicated a number of instances in which participants passed the transformation of function (TOF) test despite failing the previous test of combinatorial entailment, contrary to (and contributing to a disconfirmation of) the expected order. Participants who showed this pattern appeared to be responding under unanticipated stimulus control that still produced consistent correct responding.

In order to reduce the chances of this happening again, the TOF test was adjusted so that it required showing a correct pattern with two sets of stimuli rather than just one from then on.

The third experiment, presented in Chapter 3, involved ten typically developing children and was similar to Experiments 1 and 2 in that it compared performance on the TARPA and PLS-4. However, the former also featured several adaptations based on previous findings. One was an adaptation to deal with the TOF issue as just described. A second was related to the pass/fail criteria employed on certain levels of the TARPA. In the protocol employed in Experiments 1 and 2, children had only one chance to pass each test of derived relations (i.e., mutual and combinatorial entailment & TOF). This criterion was originally adopted in order to minimise administration time. However, multiple exposures to derivation tests are relatively common in the literature on DRR (e.g., Cassidy, 2008; O'Connor, Rafferty, Barnes-Holmes & Barnes-Holmes, 2009) and it is possible that administration of only one test of each DRR skill might under-estimate a child's ability to derive relations. Hence it was decided in this experiment to provide each participant with two exposures to each DRR test (as well as to pre-requisite training) rather than just one.

The results presented in Chapter 3 suggested that both the changes made in Experiment 3 improved the validity of the protocol. First, there were no further disconfirmations of the hierarchy, suggesting that the modified TOF testing was more valid. Second, the single subject data seemed to indicate that the second exposure to derivation tests made the protocol a more robust test of DRR. Specifically, a number of participants in Experiment 3 passed higher levels and additional derivation tests in certain tracks given the second exposure. For example, in auditory-visual 1, P3 failed Level 2 (mutual entailment) on her first attempt before passing with retraining. She then went on to pass Level 4 (TOF) within this track. Similarly, in visual-visual, P7 and P8 failed Level 2 (mutual entailment) on their first attempt before passing with retraining. They went on to pass Level 4 (TOF) and Level 3 (combinatorial entailment) respectively within this track. This single subject data clearly suggests that the second exposure to derivation tests may make the protocol a more robust test of DRR.

Despite the apparent improvements of the protocol in certain respects, the performance of one participant in Experiment 3 indicated that there might still be issues to address. This child scored highly on the PLS-4 while receiving a very low score on the TARPA, which was incongruous with the otherwise relatively consistent correlation between language and DRR. As it happened, she had received a particular sequence of tracks during Stage 3 that no other participant in this experiment had received and that involved exposure to the tracks starting with the most difficult (i.e., auditory-auditory) and progressing through progressively less difficult ones to the easiest (i.e., visual-visual). In order to examine whether this particular order might have affected her performance, she was subsequently retested using a completely new set of stimuli across the whole protocol and with a reverse of this sequence (i.e., ‘easy-to-difficult’). The score she received on this second assessment was higher and appeared to correspond more closely with her PLS-4 scores. These findings suggested that the influence of track order needed to be further examined and this issue was investigated in Experiment 4.

Experiment 4, described in Chapter 4, advanced previous work in a number of respects. First, it investigated the possible correlation not just between the TARPA and the PLS-4, a measure of purely linguistic performance, but also between the TARPA and the Stanford-Binet (SB5; Roid, 2003), a measure of cognitive or intellectual functioning more broadly. Second, it systematically examined the effect of the order of presentation of TARPA tracks (easy-to-difficult versus difficult-to-easy) on participants overall score. Third, it also involved a further slight modification to TOF testing. In this experiment, the TOF test was presented before the test for combinatorial entailment, on the basis that this has been argued to be a more methodologically sound way of conducting this test (e.g., Barnes & Keenan, 1993; Dymond & Rehfeldt, 2000; Rehfeldt & Hayes, 1998).

The results further supported the TARPA as a measure of DRR ability. They showed strong and significant correlations between the TARPA and both the PLS-4 and SB5 which were comparable with correlations between the two latter. The correlations with the PLS-4 replicated the findings in previous experiments while the correlations with the SB5 provided further evidence of the relevance of DRR with respect to cognition (i.e., further to the results of previous published studies). Experiment 4 also for the first time examined the relationship between trials to criterion (i.e., total number of training and testing trials taken to pass different levels of the TARPA) and scores on the SB5. The correlations observed between these measures suggested that trials taken on certain levels may provide an additional important metric of performance on the TARPA and this capacity on the part of the latter arguably further increases its relevance with respect to the measurement of intellectual ability.

Track order was also investigated. A strong track order effect was evident in that presentation of the “difficult-to-easy” sequence seemed to negatively affect performance in comparison with presentation of the “easy-to-difficult” sequence. A statistical analysis indicated a significant difference in TARPA scores between the two groups. In addition, while six out of ten children who received the difficult-to-easy sequence showed a very clear decrement in performance relative to their age, level of functioning as measured by the SB5, no child who received the 'easy-to-difficult' sequence showed a comparable pattern.

Experiments 1 to 4 provided useful data on the TARPA as a measure of DRR, some of which resulted in further development of the initial protocol. In Experiment 5 (Chapter 5), the final experiment in the thesis, this adjusted version of the protocol was used once more to test children with ASD. Experiment 5 aimed to extend previous studies with children with ASD in several respects. First, a larger sample (n = 20) of children with ASD was employed. Second, a more extensive range of assessment tools was used. In addition to the VABS and PLS-4, which had been employed in previous TARPA research with children with ASD, this study also included a measure of autism severity (the Gilliam Autism Rating Scale, 2nd edition; Gilliam, 2006) as well as the SB5, the test of intellectual / cognitive functioning first used in Experiment 4.

Results showed strong correlations between TARPA performance and scores on the other instruments, including the full scale and subscales of the PLS-4, SB5 and the communication domain of the VABS. Similar to the findings presented in Chapter 4, correlations also indicated a relationship between trials to criterion on certain visual and auditory tasks in Stages 1 and 2 of the TARPA and verbal ability (in this case, as measured by the PLS-4 and the verbal portion of the SB5), so that children with more advanced verbal abilities took significantly fewer trials to achieve criterion. Though, it must be noted that these trials to criterion correlations, while significant, should be interpreted with caution given the small – medium correlation coefficients observed. Other interesting patterns were also observed in that children performed substantially better on primarily visual tasks that seemed to indicate particular auditory deficits in children with ASD.

Having thus reviewed the empirical findings of the current programme of research, next a number of theoretical issues and implications arising from the research will be explored.

Theoretical Issues and Implications

DRR and Linguistic and Cognitive Abilities

The focus of this thesis has been the multi-modality assessment of derived coordinate or equivalence relations. There have been many years of work showing the potential theoretical importance of derived coordinate relations, which RFT sees as the earliest and most fundamental form of DRR, with respect to linguistic and cognitive ability (e.g., Barnes, McCullagh, & Keenan, 1990; Cassidy, 2008; Devany, Hayes, & Nelson, 1986; Dugdale & Lowe, 2000; O'Connor, Rafferty, Barnes-Holmes & Barnes-Holmes, 2009). For example, research has shown that while humans with minimal verbal ability can demonstrate stimulus equivalence, non-verbally able humans and non-humans typically do not demonstrate it, at least not readily (e.g., Barnes, McCullagh, & Keenan, 1990; Dugdale & Lowe, 2000). More recently, O'Connor et al. (2009) showed that the verbal ability of children with ASD correlated with derived equivalence performance; while Cassidy (2008) showed that scores on tests of intellectual functioning (Wechsler Intelligence Scale for Children; WISC-IIIUK) correlated with derived equivalence performance in typically developing children. The work presented in this thesis has further supported the RFT position that derived coordinate relations are integral to linguistic and cognitive ability by demonstrating this relationship across several experiments.

However, this set of experiments also has a number of strengths and advantages in relation to previous work in this domain.

First, the procedures employed in this research have arguably been more systematic and comprehensive with respect to the use of the auditory modality. For example, most previous work testing DRR has used both visual samples and comparisons (i.e., 'visual-visual' formats). There is much less work using a combination of auditory and visual stimuli in the testing format and little or no work involving all auditory stimuli. The protocol used in this research tested all three properties of derived relations as conceptualised within RFT in each of four different modality combinations (i.e., all visual, all auditory, and two separate auditory-visual modalities). Given the importance of the auditory modality in many human contexts, it could be argued that this allowed a more comprehensive and in-depth measure of derived relational ability. It also facilitated a new level of between-modality comparison with respect to derived relational testing, especially with respect to correlation of derived relations with linguistic and cognitive ability. For example, the correlational analyses involving typically developing children in Experiment 4 revealed that the auditory-visual 2 track correlated most strongly of the four tracks with the verbal scores on the SB5 while the correlation with the visual-visual track was the weakest. This is an interesting finding, explanations for which might include the foundational importance of audio-visual relations in language learning and the prevalence of audio-visual relations within typical socio-verbal contexts.

A second strength of this work as an examination of the relationship between derived relations and linguistic/intellectual capacity was the use of computer based procedures for the examination of derived relational performance. Much previous work investigating relations between derived relations and linguistic or cognitive ability typically has employed table-top procedures. Computer-based procedures such as used in this work offer a number of advantages (e.g., Whalen, Liden, Ingersoll, Dallaire, & Liden, 2006). For example, they minimise the issue of inadvertent prompting that is often associated with table top procedures delivered by an experimenter. Hence, they may be argued to be better controlled and thus more reliable procedures for gauging the extent of a child's derived relational repertoire. This is not the only research to have used computer based procedures to test derived relational ability of course or even to correlate it with linguistic and cognitive ability. For example, Cassidy (2008) using a computer-based procedure compared derived equivalence performance and scores on WISC-IIIUK. However, the current work is different to the study just mentioned in important ways; as outlined above the computer based procedures employed in this research allowed a multi-modality assessment of DRR whereas Cassidy (2008) assessed DRR in the visual modality only. In addition, the work presented in this thesis involved younger children, who ranged in age from 2- 8 years, whereas Cassidy (2008) employed older children (i.e., 8 – 12 years). Thus, this work has added considerably to the existing literature in this area.

A third strength of this work was the fact that it compared derived relational ability with both linguistic and cognitive ability, allowing a useful additional level of exploration and comparison. Previous work correlated DRR with either linguistic or cognitive ability. In two studies reported in this thesis (e.g., Experiment 4 & 5) both linguistic and cognitive ability were measured. The inclusion of tests of linguistic and cognitive ability provided support for DRR as a skill relevant to language as traditionally defined and to cognitive skills more broadly. Additionally, the inclusion of multiple tests allowed an interesting level of comparison; for example, the finding that trials to criterion on certain levels of the TARPA correlated with verbal ability (i.e., PLS-4 & verbal scores on SB5) but not non-verbal ability (i.e., non verbal scores on the SB5) emerged from this work.

Fourth, this work employed larger samples of young children both typically developing and with ASD. Previous studies in this area (e.g., Cassidy, 2008; Moran et al., 2010) have tended to use small sample sizes which can arguably limit the generalisability of the findings. For example, Cassidy et al. (2008) tested 12 typically developing children while Moran et al. (2010) tested just five children with ASD. The sample sizes used in individual studies of this research compare well with the studies just mentioned, for example Experiment 5 in this thesis involved 20 children with ASD. In addition, the sample, size involved in the project as a whole could be considered a particularly advantageous aspect of this work. The number of children engaged across the five experiments in this programme of research included a total of 79, at different ages and, with respect to those with ASD, levels of autism. The fact that high correlations with language and intellectual performance were seen consistently over the course of the five experiments constitutes a particularly strong demonstration of the relevance of derived relations to linguistic and cognitive ability.

On the whole the results presented in this thesis are considered particularly encouraging given that the current version of the TARPA measures the earliest form of DRR (i.e., coordination), that emerges first and, in typically developing children quite early (e.g., Lipkens et al.,1993). Thus, even though the correlations between the TARPA and the measures of linguistic and cognitive ability were predicted, the strength of the correlations was not necessarily predicted, especially the strong correlations between the parts of these standardised tests that measure more advanced language skills such as tests of vocabulary (e.g., the verbal subtests of the

SB5). In order to get higher correlations with more sophisticated instruments for measuring language and cognition, the TARPA would need to tap into an array of different patterns of DRR not just coordination (e.g., Cassidy, 2011). Future work will involve expanding the TARPA to include other frames (e.g., comparison, difference, hierarchy, etc). In the meantime however, it is a encouraging that even just with a test of emergent coordinate derived relations, there are strong and consistent correlations with a range of measures of linguistic and cognitive ability.

The TARPA as a Protocol for Measuring DRR

Evaluating the TARPA. In assessing the contribution of this work it is important to re-consider the primary aim of the thesis. This thesis aimed to contribute to the development of a protocol for measuring DRR. DRR is implied in some assessments of language skills that look for novel responding and response generalisation (e.g., Sundberg, 2008); however, direct testing of this fundamentally important capacity is not yet typical in behavioural intervention programmes. A protocol when fully developed that enables the direct assessment of these skills will constitute a novel and significant tool for use with children with ASD and/or developmental delay. Therefore, the analyses conducted as part of this thesis differed from previous more basic work that has looked at DRR and linguistic or cognitive ability, in the sense that this work was translational (e.g., Critchfield, 2011; McIlvane, 2009; Wacker, 1996; 2003) and as such the analyses conducted in this work have basic and applied implications which will now be outlined.

The correlational analyses reported in this thesis have indicated that TARPA performance is related to linguistic and cognitive ability. These results have (i) provided important support for the RFT position that DRR is central to language and cognitive performance (which has already been discussed above) and (ii) perhaps more importantly, given the aims of this project, demonstrate the potential utility of the TARPA as a RFT-based protocol for predicting and remediating DRR capacity for children with ASD and/or developmental delay.

Apart from the evidence provided by the correlational analyses, there is also evidence from the order analyses conducted (Krus et al.,, 1975; Marion et al. 2003). These analyses have supported the sequence of stages and levels within the protocol. Furthermore, where they have not supported them, they have pointed to ways in which the protocol might be improved. For example, the order analyses conducted in Chapter 2 identified a weakness in the test for TOF; it indicated a number of instances in which participants passed this test despite failing the previous test of combinatorial entailment. This test was subsequently changed and further analyses indicated that the test was now more valid. One of the inspirations for the TARPA was the Assessment of Basic Learning Abilities (ABLA) measure (Kerr et al., 1977). Research over the years has provided solid empirical support for the hierarchical organisation of basic behavioural skills suggested by the ABLA. The TARPA protocol is also a hierarchically structured measure of key behavioral capacities. However, it incorporates a much more extensive range of skills including derived relational abilities and it does so in accordance with more explicit theoretical considerations including most centrally RFT. The results from this research project provide preliminary support for the protocol including both its measurement of DRR as well as its proposed sequencing of stages and levels. With further testing and use in relevant research projects, it is envisaged that a similar evidence base might be built up for the TARPA as already exists for the ABLA.

Children with ASD. A substantial proportion of the empirical work conducted as part of this project, including the very final experiment which used the most up to date version of the protocol, was carried out with children with ASD as participants and thus the support for the TARPA protocol also has important implications for intervention with children with ASD and/or developmental delay. The findings that (i) emergent DRR skills as measured by the TARPA (scores and trials to criterion) are related to linguistic and cognitive skills and (ii) that these emergent DRR skills are hierarchically related, have very clear implications for the way in which progress is tracked in behavior analytic programmes for children with ASD.

Traditionally, progress in programmes for children with ASD has been measured through the acquisition of content; for example, number of tacts and mands, or the number of nouns, verbs, and prepositions, and so on (e.g.,*Verbal Behavior Milestones Assessment and Placement Program*; Sundberg, 2008). However, on the basis of findings presented in this thesis, the ability to (i) demonstrate emergent DRR skills (i.e., prerequisite discrimination and DRR skills) and (ii) demonstrate these skills readily (i.e., with limited training) appears to be critical to more advanced linguistic and cognitive performance. This obviously supports the proposition that progress should be systematically assessed not just in terms of content, but also in terms of these critical language learning processes. For example, knowing the level of derived relational ability that a student has reached (e.g., combinatorially entailed coordinate relations) can point to more targeted and efficient lesson programming for acquisition or expansion of a range of skills. A number of recent studies (e.g., Rehfeldt & Root, 2005; Halvey & Rehfeldt, 2005) have capitalised on participants' existing DRR repertoires to expand functional communication skills. For example, Rehfeldt and Root (2005) expanded derived requesting (mand) skills in three adults with disabilities by training specific conditional discriminations. Participants were first taught to request preferred items using pictures; they were then taught conditional discriminations between pictures and their dictated names and between dictated names and their corresponding text. In subsequent tests, all three participants requested preferred items using text (i.e., demonstrated derived mands). In a similar study, Halvey and Rehfeldt (2005) showed that training specific conditional discriminations resulted in untrained vocal requesting (derived mand) skills in three adults with severe developmental disabilities. Both these studies expanded mand repertoires via DRR indicating that this may be an effective and efficient means of programming for the emergence of a number of new skills. An assessment tool like the TARPA would be useful in terms of determining whether a student had the requisite skills to benefit from this sort of instructional design.

Moreover, knowing that a student is able to demonstrate particular relational skills might be an indicator of when to stop specifically targeting particular skills—for example, if a student is able to demonstrate combinatorially entailed coordinate or equivalence relations, it is probably less useful to continue targeting specific nouns, verbs, and so on for teaching both as a listener discrimination and as a tact. Rather such content could be taught in one mode only and then assessed in other modes. Also, if a student cannot demonstrate particular DRR skills (whether mutual entailment or combinatorial entailment within a frame of coordination, or within more advanced relational frames), then arguably curricular programming should focus on establishing those skills through multiple exemplar training on the relevant pattern of responding (e.g., see Luciano, Rodriguez, Manas, , Ruiz, Berens, & Valdivia-Salas, 2009). These points will be further elaborated upon in the following section.

Emergent DRR and Stimulus Modality

Rehfeldt (2011) recently highlighted the relative dearth of research investigating DRR involving auditory stimuli and considered this noteworthy given how fundamental the formation of auditory stimulus relations is for understanding spoken language. As mentioned previously, most previous work testing DRR has used 'visual-visual' formats and there is much less work using a combination of auditory and visual stimuli in the testing format. Examples of work involving a combination of auditory and visual stimuli include the very first empirical demonstration of equivalence by Sidman (1971), amongst others (e.g., Almeida-Verdu et al., 2008; De Rose, De Souza, & Hanna, 1996; Gast, VanBiervliet, & Spradlin, 1979; Green, 1990; Groskreutz, Karsina, Miguel, & Groskreutz, 2010; Kelly, Green, & Sidman, 1998; Sidman & Cresson, 1973; Smeets & Barnes-Holmes, 2005; Ward & Yu, 2000). With respect to the investigation of DRR using solely auditory stimuli, only two studies have been conducted thus far (e.g., Dube, Green, and Serna, 1993; Stewart & Lavelle, 2013) and these studies employed typically developing adult participants. The current work has advanced this area of research considerably by examining both auditory and auditory-visual relations in young typically developing children and children with ASD; the inclusion of these testing tracks in addition to the more conventional all visual track in the TARPA has enabled the investigation of responding to auditory stimuli in the context of training and testing for DRR and allowed comparison of all visual, all auditory and auditory-visual stimulus mixes in this respect. The findings of the current project have provided some interesting data as regards auditory stimuli and emergent DRR.

Stimulus Modality in Stages 1 and 2 of the TARPA. In each of the studies reported in this thesis visual and auditory performances were compared. In terms of

Stages 1 and 2 of the TARPA, this comparison was made by analysing aggregate performance on the visual and auditory track in terms of the number of levels passed. Children with ASD (Experiment 1 & 5) provided most of the data for these analyses because typically developing children tended to readily pass these early stages (i.e., with minimal trials). This data indicated that the visual track showed the highest number of overall passes (160) while the auditory track showed fewer overall passes (127). The divergence in performance on discrimination tasks involving visual and auditory stimuli was further supported by the analysis of trials to criterion data presented in Chapter 5. Children tended to take fewer trials to pass visual levels in comparison with auditory levels and the difference was found to be statistically significant at a group level.

Stimulus Modality in Stage 3 of the TARPA. The effect of auditory stimuli in the context of DRR was also evident in the advanced stages and levels of the TARPA that involved multi-modality assessments of arbitrary conditional discriminations and DRR skills (i.e., mutual entailment, combinatorial entailment, TOF). Across the work reported in this thesis, visual and auditory performances for Stage 3 alone were analysed by comparing aggregate performances on each of the four tracks in terms of the number of levels passed. For the children with ASD in Experiment 1, this revealed that the visual-visual track showed the highest number of passes (16), while the auditory-auditory and auditory-visual 1 track showed the lowest number (9). When the data from the typically developing children in Experiments 2, 3 and 4 were collated, the visual-visual track showed the highest number of passes (98), the two auditory-visual tracks showed a similar although slighter lower number of passes (AV1 = 92; AV2 = 83) while the auditory-auditory track showed the lowest number (57).

Tracks were also compared in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and TOF) on each. In the case of the children with ASD it was found that in the visual-visual track, three participants showed all three properties whereas in the case of each of the other tracks which involved at least some auditory stimuli only one participant showed all three properties. Out of the typically developing children there were fourteen participants who showed all three properties on the visual-visual track, nine who showed all three on the auditory-visual 1 track, seven who showed all three on the auditory-visual 2 track and three who showed all three on the auditory-auditory track (see Appendix F for a detailed breakdown).

Previous research in this area (e.g., Green, 1990; Smeets & Barnes-Holmes, 2005) has suggested that persons with developmental disabilities and typically developing young children demonstrate DRR more readily in the auditory-visual than in the visual-visual domain. The current research provides data that bears on this question. It should be pointed out that here are some important methodological differences between the protocols used in the current research and previous work and so this complicates comparison. For example, both the studies cited at the start of this paragraph used table-top procedures while the TARPA is an automated computer based procedure. However, there is some degree of difference between the findings of this work and those found in those previous studies with respect to the visual-visual/auditory-visual comparison, since in the current research more participants showed patterns consistent with DRR when the relations involved all visual stimuli (i.e., the visual-visual track in Stage 3). Nevertheless, the differences between the visual-visual track and the auditory-visual tracks found during this research program were not very substantial. What does seem clear from the current project at the least, however, is that children either with ASD or typically developing, tend to show better DRR performance when relations involve at least some visual stimuli in comparison with relations that involve only auditory stimuli.

Moran et al. (2010) suggested that this pattern of responding in which children perform better when the trained and tested relations involve at least some visual stimuli might be explained on the basis of the difference in behavioural control obtained by simultaneous and successive discriminations. Visual stimuli can be presented simultaneously, while auditory stimuli must be presented successively, and simultaneous discriminations are learned more readily than successive ones due to the more time-limited availability of successively presented stimuli. This simultaneous-successive difference seems to be one key contributing factor. Another explanation might be that scenarios involving the types of all auditory conditional discriminations required in the TARPA protocol are probably not as common in the everyday socio-verbal environment as scenarios requiring audio-visual conditional discriminations, (i.e., in natural language interactions young children receive more training in audio-visual discriminations than they do in auditory-auditory discriminations).

Children with ASD. Some other interesting patterns of responding by children with ASD with respect to auditory stimuli were noted also. First, those who showed even the simplest form of DRR (i.e., mutual entailment) invariably passed all levels within Stages 1 and 2 (which was not a ‘given’ since even passing one level of these stages meant that the person was at least tested on the next stage up). Meanwhile those children with ASD who did not reach this criterion often failed levels within these stages particularly those involving auditory stimuli. Second, even when children with ASD passed levels involving auditory stimuli they often took a greater number of trials to pass these levels in comparison to the same levels within the visual track. Some of these children also demonstrated what might be described as unbalanced (based on the TARPA assessment) visual and auditory repertoires. For example, P4 in Experiment 5 could do a simple discrimination involving one, two and three comparisons but failed to show any simple discriminative abilities at all on the corresponding auditory levels. Similarly, P2 demonstrated the ability to match identical visual stimuli with two and three comparisons and also showed generalised identity matching but failed all these Stage 2 levels in the auditory modality. The TARPA has been developed as a precise and efficient way of measuring important auditory discrimination skills and the results presented here (e.g., performances of P2 and P4) suggest that the protocol in its current form fulfills this function and provides a fine-grained analysis of a child’s discrimination repertoire across auditory and visual modalities.

Within behavioural analytic programming, teaching auditory discrimination skills (i.e., identity matching with auditory stimuli) is rarely discussed (with the exception of Chavez-Brown, 2005; Greer & Ross, 2008), while the importance of visual discrimination skills (e.g., generalised identity matching, etc.) is widely disseminated among behavioural clinicians as important to the developmental of more advanced language and cognitive skills (e.g., Leaf & McEachin, 1999; Sundberg, 2008; Williams & Jackson, 2009). One reason for the lack of emphasis on assessing and teaching auditory discrimination skills could be the difficulty in arranging instructional programmes involving only auditory stimuli. A number of studies with persons with developmental disabilities have suggested that an auditory-auditory identity matching task may be a worthwhile addition to the ABLA test. The task was presented as follows; a tester said a word (e.g., “pen”) on some trials and a different word (e.g., "block") on other trials, while one assistant said "pen" and a second assistant said "block." Across trials, the assistants randomly alternate as to who says which word and who speaks first. On each trial, the correct response was to point to the assistant who spoke the word that matched that of the tester. The research presented in this thesis has provided the first demonstration that the TARPA when fully-developed might provide a better means of testing and training these critical skills.

As well as the TARPA providing a means of measuring increasingly complex discrimination skills, it also could be argued that the results of a TARPA assessment provide a clear pathway and conceptual framework for remediating identified deficits across discrimination abilities and stimulus modalities. For the purposes of providing an example the performance of Participant 2 from Experiment 5 is used.

P2's performance within Stages 1 and 2 of the TARPA suggest that her visual discrimination skills were better developed than her auditory discrimination skills. Intervention within the auditory track, would involve specific training of these deficient skills such as matching auditory stimuli with two and three comparison where the initial objective would be the establishment of a generalised identity matching repertoire with auditory stimuli. A more long term objective in this case would be the establishment of a DRR repertoire with exclusively auditory stimuli which may map onto the child's intraverbal repertoire.

Concurrently, the child's visual or auditory-visual repertoire could be targeted but at a different level. For example, P2 demonstrated generalised identity matching with visual stimuli and also acquired an arbitrary conditional discrimination with auditory-visual stimuli but failed to demonstrate mutual entailed responding with the same stimuli (i.e., the reversal of the trained relation). Therefore, within these modalities initially mutual entailed responding would be targeted for remediation. This could involve two programmes, where mutual entailed responding could be targeted with (i) with exclusively visual stimuli and with (ii) auditory-visual stimuli. Specifically, the learner would be exposed to MET. During MET, the learner would be directly taught to respond bidirectionally to the stimuli (i.e., A↔B). Once the child acquired these directly trained targets they would be taught another set but only in one direction (i.e., A→B) and the mutually entailed response would be probed (i.e., B→A). This process of teaching and probing would continue until the child could readily demonstrate this repertoire (i.e., acquire an arbitrary conditional discrimination and derive the untaught relation with minimal trials), Subsequent to this combinatorial entailment would be targeted within these modalities.

While it is envisaged that the TARPA protocol itself (when fully developed) will teach these skills, ideally these skills would also be targeted as part of a wider 'TARPA' based curricula. For example, it was mentioned above that the training of auditory skills might map on to training intraverbal responding skills. Similarly, the remediation plan for mutual entailment within the visual repertoire might be targeted concurrently within certain 'visual' table top programmes such as an association matching programme (e.g., baby goes with pram, etc.). The remediation plan for the auditory-visual repertoire might also be targeted within audiovisual programmes, for example, like naming programmes (i.e., listener-tact) that often feature in behavioural intervention and typically involved different content areas (e.g., nouns, verbs, prepositions, etc.). The example just provided lays out how an intervention based on the TARPA assessment may look, and while there is a body of work that supports multiple exemplar training as a means of teaching DRR skills (e.g., Rehfedlt & Barnes-Holmes, 2009; Cassidy, Roche & Hayes, 2011), the merits of this approach in the context of the TARPA is as of yet untested. Future research will need to investigate the efficacy of TARPA based procedures for remediation training and this will likely be in the context of using appropriate single subject experimental designs (e.g., Kazdin, 1982). While acknowledging the preliminary nature of this work and speculative nature of some of the suggestions outlined above, across this current work, strong correlations between DRR skills and language/cognitive abilities were repeatedly observed. Based on these and other findings (e.g., Cassidy et al., 2011) it is argued that an approach that concurrently targets (i) DRR skills as a process as well as (ii) specific content could provide a more powerful approach to behavioural intervention for children with developmental delay and/or ASD and to education more broadly.

A Track Order Effect

While the effect of stimulus modality on DRR is interesting in itself, another related effect, namely an order effect, is also interesting and has potential basic and applied implications. The possibility that the order of presentation of Stage 3 tracks might impact overall scores on the TARPA protocol (i.e., a track order effect) emerged from an inexplicable pattern of responding demonstrated by one participant in Chapter 3. This participant despite having advanced language abilities performed very poorly on the TARPA. It was speculated that starting with the apparently most difficult track (e.g., auditory-auditory) and progressing through the other tracks to the apparently least difficult (e.g., visual-visual) may have resulted in a reduced overall score on the TARPA. The possible effect was subsequently tested and the results of Experiment 4 supported the proposition that receiving the most difficult sequence of tracks rendered the protocol more difficult for participants. This finding was important as it clearly outlines the importance of administering the TARPA, and in particular tracks in Stage 3 in a specific way. At a more basic level, this issue hasn't been extensively investigated, only one previous study showing a similar effect was identified (i.e., O'Connor et al., 2009).

An explanation of this 'order effect' also presents an interesting question. RFT suggests that coherence in a relational network functions as a powerful reinforcer for relational activity itself (see e.g., Roche, Barnes-Holmes, Barnes-Holmes, Stewart, & O'Hora, 2002). In Experiment 4, most participants who received the 'easier' presentation of tracks showed appropriate derivation patterns early on, and while these derivation patterns were shown under conditions of extinction (responses were not followed by a contingent reinforcement), it is possible that 'verbal coherence' functioned as a reinforcer which subsequently strengthened responding across subsequent tracks. Conversely, participants who failed to show these patterns early on (i.e., participants who received the 'difficult' auditory-auditory track first) missed out on this potential reinforcement which perhaps weakened their responding across later tracks. Indeed, the overt verbal behavior of some of the children during derivation tests seemed to indicate that showing the correct derivation patterns was functioning as some sort of conditioned reinforcer. For example, during derivation tests children often made statements such as 'Yes- I'm getting them right!' or 'I'm winning!' Of course, the possibility that 'verbal coherence' played some part in the observed track order effect is purely speculative and future research needs to examine these issues further.

At an applied level, this finding of a track order effect would lend further empirical support to the importance of designing educational programmes that carefully match learners' current level of functioning. It suggests that engaging a child in a task where they do not contact reinforcement (extrinsic or intrinsic) hinders learning on subsequent tasks, even tasks that are within their repertoire. The importance of tailoring educational programmes to a learner's current level of functioning is well documented within applied behavior analysis as is the importance of using appropriate schedules of reinforcement and the findings presented here lend further support to this idea.

The track order effect observed in Experiment 4 was just discussed in terms of DRR becoming a 'conditioned reinforcer' that sustains learning in environments in which direct reinforcement isn't freely available. Other researchers have made similar suggestions in relation to DRR (e.g., Luciano, et al., 2009). Establishment of skills that will allow a child with ASD to succeed in more naturalistic settings is a key goal of behavioral intervention. As such, the importance of establishing 'learning to learn' skills in children with ASD is well documented. For example, Holth, Vandbakk, Finstad, Grønnerud, and Akselsen Sørensen (2009) recently outlined procedures for establishing social behavior (e.g., eye gaze, gestures, nodding, smiling, etc.) as conditioned reinforcers in order to avoid a reliance on contrived reinforcers that are not always available in the natural environment. Arguably, establishing DRR as a conditioned reinforcer could have similar benefits for the child with ASD.

Limitations and Directions for Future Research

Throughout this chapter limitations and directions for future research have been highlighted where appropriate. Some further points in these respects will now be considered.

TARPA: A Limited Measure of Relational Framing

One important point relates to the development of the TARPA as a comprehensive measure of relational framing. The version of the protocol used in this programme of research only examined DRR at the level of a frame of coordination or equivalence. This perhaps could be conceived as a criticism of this work as a key feature of RFT is that it encompasses a far broader range of relational skills than predicted by equivalence research (e.g., Sidman, 1994). However, this work is based on a version of the TARPA that represents only an early step in its development. The development of the TARPA deliberately began with the simplest and most foundational level of DRR and the one into which there is already most empirical investigation (i.e., coordinate DRR or equivalence). However the aim of the TARPA project is to gradually extend the reach of the protocol in terms of its capacity to examine DRR and hence future work will involve the development of stages and levels for the examination of higher and more complex frames (e.g., comparison, difference, etc.). The inclusion of these further frames will no doubt help the TARPA to differentiate between participants of different ability with more depth and precision in future research. At this stage in its development, however, the TARPA is still focused on coordinate DRR. At the same time, given the importance of coordination as the first form of DRR, and given that the TARPA assesses pre-cursors to this ability as well as its appearance across different modalities, the TARPA even at this point is still a comprehensive and useful instrument and as such it is appropriate that it be tested and improved in its current range, which was the aim of the current project.

Methods Employed to Investigate DRR

Another potential limitation of the TARPA and hence of this research which is based on this protocol is related to the methods employed to investigate DRR. The TARPA uses (i) a matching-to-sample preparation, where (ii) a test score based on accuracy is the primary dependent variable. Potential issues with both these features of the TARPA will now be individually discussed.

Matching-to-Sample Procedures. MTS protocols have been criticised on a number of points. First, it is claimed that they often involve extensive training and testing and as such are an inefficient means of looking at DRR (e.g., Horne & Lowe, 1997; Whelan & Schlund, 2013). Second, MTS preparations have been said to lack face and ecological validity as they are not typically present in the natural learning environment. Third, MTS is considered limited as it is more suited to the examination of relations of equivalence or coordination than other types of stimulus relations (e.g., Hayes & Barnes, 1997; Horne et al., 1996).

There are a number of existing non MTS procedures that assess DRR skills, specifically the Relational Evaluation Procedure (REP; Stewart, Barnes-Holmes, & Roche, 2004) and the Relational Completion Procedure (RCP; Dymond, Roche, Forseyth, Whealan, & Rhoden, 2007, 2008; Dymond & Whelan, 2010). The REP was developed to study multiple stimulus relations; for example Same and Different relations (e.g., Stewart et al., 2004) and Before and After relations (Barnes-Holmes et al., 2001; O'Hora, Barnes-Holmes, Roche, & Smeets, 2004). The defining feature of the REP is that it allows participants to evaluate, or report on, different stimulus relations. That is, in the REP participants confirm or deny the applicability of particular stimulus relations to other sets of stimulus relations. For example, a participant might be presented with a contextual cue for OPPOSITE and three or more arbitrary stimuli that are specified within that trial as participating in an Opposite relation. The participant is then required to select one of two arbitrary shapes that function as True or False. So, for instance, selecting True in the presence of the OPPOSITE contextual cue, and selecting False in the presence of the SAME contextual cue would be considered correct, reinforced responses.

The RCP is also used within RFT research and it is based on the REP. In the RCP, stimuli are presented in sequence from left to right, starting with the sample and followed (1 s later) by a contextual cue, a blank space and up to five comparisons. The participant's task is to "complete the sentence" by dragging and dropping one of several comparisons into the blank space and confirming each selection. A key strength of REP and RCP is that they can efficiently train and test multiple stimulus relations and it is likely that these procedures will be adapted and used when more advanced frames are added to the TARPA protocol.

However, while the REP and RCP offer advantages over MTS procedures they also have certain elements that made them unsuitable for use in the current version of the TARPA. For example, it was important to minimise the possibility that procedural variables would render some children unsuitable for testing on the TARPA and a key element of the REP method involves the participant affirming and disconfirming relations using Yes / No or True / False. However, while this repertoire would already be established in typically developing participants, it is something that may need to be taught in many children with ASD. Similarly, the RCP requires a certain topographical response (i.e., dragging and dropping comparisons and subsequently confirming choices) that may not be in the repertoire of a child performing in the early stages of the TARPA. The only prerequisites a child requires for testing on the TARPA are basic learner behaviours such as sitting and responding to adult prompts as well as the ability to point or touch the screen to select a stimulus.

The children with ASD who participated in the current research varied greatly in terms of level of functioning and it could be considered a relative strength of the TARPA protocol that all children were testable on this measure. While the use of the alternative measures just mentioned offer potential advantages in terms of efficiency they could arguably have precluded some participants who did not possess certain prerequisites, participants who provided rich data particularly in terms of the early stages of the TARPA (e.g., auditory discrimination skills).

Two-Choice Conditional Discrimination Preparations. A related issue is the use of only two choices in all conditional discriminations. The possibility of false positives is increased due to the use of a two choice preparation (e.g., Carrigan & Sidman, 1992; Sidman, 1987). This version of the TARPA was developed based on Moran et al. (2010) and was subject to pilot testing before it was used in this research. Initially, this version involved both two- and three-choice comparisons. A key reason why it was subsequently decided during pilot testing (unpublished) with typically developing children (aged between 2 years 6 months and 10 years 0 months) to transition to what is now primarily a two choice protocol is that it became apparent that children took extensive training to acquire relations involving three comparisons.

With three auditory comparisons, even older children had difficulty learning the conditional discriminations. For example, a 10 year old child during pilot testing with three-choice comparisons took extensive training to acquire the requisite conditional discriminations and even then didn't demonstrate equivalence. The initial inclusion of three-choice comparisons was based on an argument against the use of two choice procedures in studies of emergent matching to sample (e.g., Carrigan et al., 1992; Sidman, 1987) which has been refuted somewhat in more recent literature (e.g., Boelens, 2002; Saunders, Chaney & Marquis, 2005). The decision to use just two choice tests throughout the protocol was taken in order to minimise training times and thus make TARPA assessment as brief and efficient as possible.

The two choice format was adopted throughout most of the TARPA but as suggested this also has potential issues. The key weakness of this format is that it can increase the possibility of false positive results. Indeed, this is what appears to have happened in Experiments 1 and 2 when some participants passed transformation of function (i.e., Level 4) having failed combinatorial entailment (i.e., Level 3). Experiment 4 addressed the issue of ‘false positives’ in the transformation of function test by requiring correct performance with more than one stimulus set. Even though the order analyses conducted across this work did not show a similar weakness with respect to levels other than transformation of function, the possibility that participants might have passed other levels in a similar way cannot be discounted. However, given the multiple stages and levels across the protocol, it is unlikely that ‘false positives’ might allow a participant to attain a much higher score than they would otherwise attain. One can also argue that the use of multiple tracks helps to increase the validity of the test. However, of course if one relies on the provision of multiple testing tracks as a guard against false positives affecting overall performance, then one can be less confident with respect to ones measurement of performance on any one individual track.

Response Accuracy as a Measure of Performance. Another potential limitation is related to the use of response accuracy data to measure performance on the TARPA. When response accuracy data is used, the proportion of test trials correct is the measure by which emergence of derived relations is inferred (Dymond

& Rehfeldt, 2001; Whelan et al., 2013). Despite the widespread use of this measure, Whelan et al. (2013) cautions that the emergence of derived relations and a test score (e.g., 7/8 consecutive trials correct in the case of the TARPA for derivation tests) are not one and the same thing. They also warn that accuracy as a dependent measure may overlook important subtleties in responding.

While this research used test scores based on accuracy as a dependent variable throughout, in some of the later experiments another metric of performance was examined in the form of trials to criteria data (which involved total training and testing trials). While this measure was also derived from accuracy data, it did at least allow more fine grained differentiation of performance. Trials to criteria showed correlations with scores on both the PLS-4 as well as verbal portions of the SB5. The further development of the TARPA as an assessment and a training tool might involve exploring the use of fluency data (i.e., accuracy plus speed) as a measure of performance on the TARPA.

The Administration of Multiple Tests by One Experimenter

It could also be argued that the administration of all tests (i.e., the TARPA and alternative linguistic/cognitive measure or measures) by the same experimenter is a methodological weakness of this work. In future research different experimenters should administer the TARPA and the alternative measure(s) of functioning. Ideally this should involve blind testing also. The resources (e.g., additional experimenter) needed to implement these controls were not available to this research. However, the TARPA is a computer based protocol with strict rules in relation to administration and scoring so arguably the experimenter in this research had little or no influence on participants performance on the protocol.

Predictive Validity

A final issue relates to the way in which the validity of the TARPA was examined; across the experiments presented in this thesis, the overall validity of the TARPA was assessed by correlating performance on the TARPA with scores on various standardised measures of language, IQ and adaptive behaviour. It would have also been desirable to measure the predictive validity of the TARPA; arguably the failure to do this is a weakness of this work. Predictive validity assesses the extent to which a score on a scale or test predicts scores on some criterion measure. In terms of the TARPA, this would involve measuring behaviour that a TARPA score might predict. For example, a score within Stage 2 of the TARPA (e.g., passing tasks involving non arbitrary conditional discriminations but failing tasks involving arbitrary conditional discriminations) should predict performance on certain early behavioural intervention programmes (e.g., performing well on matching tasks but struggling with listener or tact programmes). While performance at the highest level of the TARPA (e.g., passing transformation of function in Stage 3 of the protocol) should predict a certain degree of language generativity (e.g., produce novel and untaught sentences). Future work could examine the predictive validity of the TARPA in this way.

Summary and Conclusion

This thesis has provided the first thorough examination of the TARPA as a methodology for assessing DRR. The overarching goal of this current work has been the development of the TARPA as an assessment tool and the results have been encouraging: (i) the TARPA has been found to correlate consistently with both linguistic and cognitive ability and (ii) the structure of the protocol appears to be appropriate. At a more basic level, this research has also provided insight into: (i) how different Pass/Fail criteria affect DRR performance; (ii) how stimulus modality affects DRR performances and (iii) the effect of order of presentation of DRR tasks on overall DRR performance. As well as elucidating important issues that surround the TARPA, these additional analyses have provided data that will benefit DRR research more broadly.

A number of possible directions for future studies have also been indicated. One avenue might involve addressing deficits in participants DRR repertoire as identified by the TARPA through multiple exemplar training and following this training, examining the effects of the latter on performance on the TARPA and other tests of DRR as well as on standardised measures of language and cognitive ability. In addition, future work will likely involve the expansion of the TARPA in a number of directions. This might include (i) developing the protocol to examine forms of DRR other than sameness (equivalence) relations (e.g., opposition, distinction, comparison, hierarchical etc.) and employing (ii) more subtle dependent measures of DRR (e.g., fluency).

In summary, the work presented in this thesis is promising for the future development of the TARPA and gives an early indication that the protocol could eventually provide a comprehensive tool for the assessment and training of core linguistic and cognitive abilities for use with a number of different populations, including typically developing children and children with ASD or other forms of developmental delay. This no doubt would constitute a key addition to existing tools for intensive behavioral intervention.

References

Almeida-Verdu, A. C., Huziwara, E. M., Souza, D. G., Rose, J. C., Bevilacqua, M. C., Lopes, J. & McIlvane, W. J. (2008). Relational learning in children with deafness and cochlear implants. *Journal of the experimental analysis of behavior*, *89*(3), 407-424. doi:10.1901/jeab.2008-89-407.

Anderson, S. R., Avery, D. L., DiPietro, E. K., Edwards, G. L., & Christian, W. P. (1987). Intensive home-based intervention with autistic children. *Education and Treatment of Children*, 10, 352–366.

Barnes, D., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behavior, 8*, 1-11.

Barnes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, *59*, 61-81. doi: 10.1901/jeab.1993.59-61

Barnes, C. S., Grannan, L., Lovett, S., & Rehfeldt, R. A. (2012). Behavior Analytic Interventions for Individuals with Autism Spectrum Disorders. *Psicología, Conocimientoy Sociedad, 2*, 25-53.

Barnes, C., & Rehfeldt, R.A. (2013).Advances in language interventions based on Relational Frame Theory for individuals with developmental disorders. In S. Dymond, & B. Roche (Eds.), *Advances in relational frame theory: Research and applications*. Oakland, CA: New Harbinger Publications.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of functions in accordance with symmetry: II. *The Psychological Record, 51,* 589-604.

Barnes-Holmes, Y., Barnes-Holmes, D., McHugh, L., & Hayes, S. C. (2004). Relational Frame Theory: Some implications for understanding and treating human psychopathology. *International Journal of Psychology and Psychological Therapy*, *4*, 355-376.

Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P.M. (2004). Establishing relational responding in accordance with opposite as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 559-586.

Barnes-Holmes, D., & Murphy, C. (2007a). Addressing the generativity of language: A late reply to Chomsky. In L. B. Zhao (Ed.), *Autism research advances*. NY: Nova Science Publishers.

Barnes-Holmes, D., & Murphy, C. (2007b). Towards establishing generative language with children with autism. In B. S. Mesmere (Ed.), *New autism research developments.* NY: Nova Science Publishers.

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The implicit relational assessment procedure (IRAP) as a response-time and event related potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record, 58*, 497-516.

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: evidence for a relational operant. *Journal of Applied Behavior Analysis, 40,* 45-71. doi: 10.1901/jaba.2007.7-06

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, *19*, 163-197.

Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch Early Intervention Program after 2 years. *Behaviour Change*, *10,* 63-74.

Boelens, H. (2002). Studying stimulus equivalence: Defense of the two-choice procedure. *The Psychological Record*, *52*, 305-314.

Bzoch, K. R., & League, R. (1991). *Receptive-expressive emergent language scale* (2nd ed.). Austin, TX: Pro-Ed.

Carbone, V. (2000). Teaching verbal behavior to children with autism and related disabilities.

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, *58*, 183-204. doi: 10.1901/jeab.1992.58-183

Casey, L., & Kerr, N. (1977). Auditory-visual discrimination and language prediction. *Rehabilitation Psychology*, *24*, 137-155.

Cassidy, S. (2008). *Relational Frame Theory and human intelligence: A conceptual and empirical analysis*. (Unpublished doctoral thesis). National University of Ireland, Maynooth.

Cassidy, S., Roche, B., & O'Hora, D. (2010). Relational Frame Theory and human intelligence. *European Journal of Behavior Analysis*, *11*, 37-51.

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.

Chavez-Brown, M. (2005). *The effects of the acquisition of a generalized auditory word match-to-sample repertoire on the echoic repertoire under mand and tact conditions.* (Doctoral Dissertation, 2005, Columbia University). Abstract from: UMI Proquest Digital Dissertations [on-line]. Dissertations Abstracts: AAT 31559725.

Choi, J. (2011). Effects of mastery of auditory match-to-sample instruction on echoics, emergence of advanced listener literacy, and speaker as own listener cups by elementary school students with ASD and ADHD. (Unpublished doctoral dissertation). Columbia University, NY.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*,*35*, 26-58.

Coolican, J., Bryson, S. E., & Zwaigenbaum, L. (2008). Brief Report: Data on the Stanford–Binet Intelligence Scales in Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *38*, 190-197. doi: 10.1007/s10803-007-0368-2

Clark, H. H. (1994). Discourse in production. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* London: Academic Press.

Cooper, J.O., Heron, T.E., & Heward, W.I. (2007). A*pplied Behavior Analysis* (2nd ed.). NY: MacMillan.

Cowley, B. J.. Green. G.. & Braunling-McMorrow, D.(1992). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis, 25*, 461-475.

Critchfield, T. S. (2011). Translational contributions of the experimental analysis of behavior. *The Behavior Analyst, 34,* 3-17.

Cullinan, V., &Vitale, A.(2008). The contribution of Relational Frame Theory to the development of interventions for impairments of language and cognition. *Journal of Speech-Language Pathology and Applied Behavior Analysis, 2*(4)-*3*,122-135.

Delgado, J. A. P., & Oblak, M. (2007). The effects of daily intensive tact instruction on the emission of pure mands and tacts in non-instructional settings by three preschool children with developmental delays. *Journal of Early & Intensive Behavior Intervention, 4*, 392-411.

De Rose, J. C., de Souza, D. G., & Hanna, E. S. (1996). Teaching reading and spelling: Stimulus equivalence and generalization. *Journal of Applied Behavior Analysis, 29,* 451–469.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243-57. doi: 10.1901/jeab.1986.46-243

Dickins, D.W., Singh, K.D., Roberts, N., Burns, P., Downes, J.J., Jimmieson, P., & Bentall, R.P. (2001). An fMRI study of stimulus equivalence. *NeuroReport, 12,* 2-7. doi: 10.1097/00001756-200102120-00043

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior,62*, 331-351. doi: 10.1901/jeab.1994.62-331

Dougher, M. J., Hamilton, D., Fink, B., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior, 88,* 179-197. doi: 10.1901/jeab.2007.45-05

Dube, W. V., Green, G., & Serna, R. W. (1993). Auditory successive conditional discrimination and auditory stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, *59*, 103-114. doi: 10.1901/jeab.1993.59-103

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73,* 5-22. doi: 10.1901/jeab.2000.73-5

Dymond, S., & Barnes, D. (1994). A transfer of self-discrimination response functions through equivalence relations. *Journal of the Experimental Analysis of Behavior*, *62*, 251-267. doi: 10.1901/jeab.1994.62-251

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, *23*, 239-254.

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, *19*, 8-12.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2007). Transformation of avoidance response functions in accordance with same and opposite relational frames. *Journal of the Experimental Analysis of Behavior*,*88*, 249-262. doi: 10.1901/jeab.2007.22-07

Dymond, S., Roche, B., Forsyth, P., Whelan, R., & Rhoden, S. (2008). Derived avoidance learning: Transformation of avoidance response functions in accordance with the relational frames of same and opposite. *The Psychological Record*, *58*, 271-288.

Dymond, S., & Whelan, R. (2010). DRR: a comparison of match-to-sample and the relational completion procedure. *Journal of the Experimental Analysis of Behavior*, *94*, 37-55. doi: 10.1901/jeab.2010.94-37

Dymond, S., & Alonso-Álvarez, B. (2011). The selective impact of Skinner's Verbal Behavior on empirical research: A reply to Schlinger (2008). *The Psychological Record*, *60*, 10.

Dymond, S., & Roche, B. (Eds.). (2013). *Advances in Relational Frame Theory & contextual behavioral science: Research & application*. Oakland, CA: New Harbinger Publications.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: the role of naming. *Journal of the Experimental Analysis of Behavior*, *58*, 123-133. doi: 10.1901/jeab.1992.58-123

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7 a comparison controlled study. *Behavior Modification*, *31*, 264-278. doi: 10.1177/0145445506291396

Fenske, E. C., Zalenski, S., Krantz, P. J., & McClannahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis and Intervention in Developmental Disabilities*, 5, 49-58. doi: 10.1016/S0270-4684(85)80005-7

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior*, *42*, 143-157. doi: 10.1901/jeab.1984.42-143

Fiorile, C. A., & Greer, R. D. (2007). The induction of naming in children with no prior tact responses as a function of multiple exemplar histories of instruction. *The Analysis of Verbal Behavior, 23*, 71-87.

Gast, D. L., VanBiervliet, A., & Spradlin, J. E. (1979). Teaching number-word equivalences: A study of transfer. *American Journal of Mental Deficiency*. *83*, 524–527.

Goldsmith, T. R., LeBlanc, L. A., & Sautter, R. A. (2007). Teaching intraverbal behavior to children with autism. *Research in Autism Spectrum Disorders, 1*, 1-13. doi: 10.1016/j.rasd.2006.07.001

Gómez, S., Lopez, F., Martin, C. B., Barnes-Holmes, Y., & Barnes-Holmes, D. (2010). Exemplar training and a derived transformation of functions in accordance with symmetry and equivalence. *The Psychological Record*, 57, 273-294.

Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation, 95*, 260-270.

Greer, R. D., Stolfi, L., Chavez-Brown, M., & Rivera-Valdes, C. (2005). The emergence of the listener to speaker component of naming in children as a function of multiple exemplar instruction. *The Analysis of Verbal Behavior,* 21, 123-134.

Greer, R. D., Stolfi, L., & Pistoljevic, N. (2007). Emergence of naming in preschoolers: A comparison of multiple and single exemplar instruction. *European Journal of Behavior Analysis,* 8, 109-131.

Greer, R. D., & Ross, D. E. (2008). *Verbal behavior analysis: Inducing and expanding new verbal capabilities in children with language delays.* New York: Allyn & Bacon.

Greer, R. D., & Speckman, J. (2009). The integration of speaker and listener responses: A theory of verbal development. *The Psychological Record, 59,* 449-488.

Groskreutz, N. C., Karsina, A., Miguel, C. F., & Groskreutz, M. P. (2010). Using complex auditory—visual samples to produce emergent relations in children with autism. *Journal of applied behavior analysis*, *43*, 131-136.

Halvey, C., & Rehfeldt, R. A. (2005). Expanding vocal requesting repertoires via relational responding in adults with severe developmental disabilities. *The Analysis of Verbal Behavior, 21*, 13-25.

Harapiak, S. M., Martin, G. L., & Yu, D. (1999). Hierarchical ordering of auditory discriminations and the Assessment of Basic Learning Abilities test. *Journal on Developmental Disabilities*, *6*, 32-50.

Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four-to six-year follow-up. *Journal of Autism and Developmental Disorders*, *30*, 137-142.

Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule- governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 153-190). New York: Plenum. .

Hayes, S. C. (1991). A relation control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 19-40). Reno, NV: Context Press.

Hayes, S.C. (1992). Verbal relations, time, and suicide. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 109-118). Reno, NV: Context Press.

Hayes, S. C. (1994). Relational frame theory: A functional approach to verbal events. In S. C. Hayes, L. J. Hayes, M. Sato & K. Ono (Eds.), *Behavior analysis of language and cognition*. Reno, NV: Context Press

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior, 68*, 235-244. doi: 10.1901/jeab.1997.68-235

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Plenum Press.

Hayes, S.C., Fox, E., Gifford, E., Wilson, K., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In S.C. Hayes, D. Barnes-Holmes & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition* (pp.21-50). New York: Plenum Press.

Holth, P., Vandbakk, M., Finstad, J., Gronnerud, E. M., & Sorensen, J. M. A. (2009). An operant analysis of joint attention and the establishment of conditioned social reinforcers. *European Journal of Behavior Analysis*, *10*, 143-158.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*(1), 185-241. doi: 10.1901/jeab.1996.65-185

Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior*, *68,* 271-296. doi: 10.1901/jeab.1997.68-271

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*, 359-383. doi: 10.1016/j.ridd.2004.09.005

Howell, D. C. (1998). *Statistical methods in human sciences*. New York: Wadsworth.

Ingvarsson, E. T., Tiger, J. H., Hanley, G. P., & Stephenson, K. M. (2007). An evaluation of intraverbal training to generate socially appropriate responses to novel questions. *Journal of Applied Behavior Analysis*, *40*, 411-429. doi: 10.1901/jaba.2007.40-411

Jackson, M., Williams, W. L., & Biesbrouck, J. (2006). Conditional discrimination ability, equivalence formation and mental retardation: Implications for development in children with developmental disabilities. *Journal of Speech Language Pathology and Behavior Analysis*, *1*, 27-42.

JE, G. (2006). *Gilliam Autism Rating Scale 2nd Edition (GARS-2).* Austin, TX: Pro-Ed.

Jennett, H. K., Harris, S. L., & Delmolino, L. (2008). Discrete trial instruction vs. mand training for teaching children with autism to make requests. *The Analysis of Verbal Behavior, 24*, 69-85.

Keintz, K. S., Miguel, C. F., Kao, B., & Finn, H. E. (2011). Using conditional discrimination training to produce emergent relations between coins and their values in children with autism. *Journal of Applied Behavior Analysis, 44*(4), 909. doi: 10.1901/jaba.2011.44-909

Kelly , S., Green , G., & Sidman , M. (1998). Visual identity matching and auditory-visual matching: A procedural note. *Journal of Applied Behavior Analysis, 31*, 237–243. doi:10.1901/jaba.1998.31-237

Keohane, D. D., & Greer, R. D. (2005). Teachers' use of a verbally governed algorithm and student learning. *International Journal of Behavioral Consultation and Therapy*, *1*, 252-271.

Kerr, N., Meyerson, L., & Flora, J. A. (1977). The measurement of motor, visual, and auditory discrimination skills [Monograph]. *Rehabilitation Psychology, 24*, 95–112.

Krus, D.J., Bart, W.M., & Airasian, P.W. (1975). *Ordering theory and methods*. Los Angeles: Theta Press.

Leaf, R. B., & McEachin, J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. New York: DRL Books.

LeBlanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, J. E. (2003). The effects of three stimulus-equivalence testing conditions on emergent US geography relations of children diagnosed with autism. *Behavioral Interventions, 18*, 279-289. doi: 10.1002/bin.144

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology*, *56*, 201-239. doi: 10.1006/jecp.1993.1032

Lord, C., & McGee, J.P. (Eds.). (2001). *Educating children with autism*. Washington DC: National Academy Press.

Lovaas, O. I. (1977). *The autistic child: Language development through behavior modification*. New York, NY: Irvington.

Lovaas, O.I. (1981). *The ME book: Teaching developmentally disabled children*. Austin, TX: Pro-Ed Inc .

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55,* 3-9. doi: 10.1037//0022-006X.55.1.3

Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: PRO-ED, Inc.

Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior*, *69*, 327-354. doi: 10.1901/jeab.1998.69-327

Lowenkron, B. (2004). Meaning: A verbal behavior account. *The Analysis of Verbal Behavior*, *20*, 77-97.

Luciano, C., Rodriguez, M., Manas, I., Ruiz, F., Berens, N., & Valdivia-Salas, S. (2009). Acquiring the earliest relational operants: coordination, distinction, opposition, comparison and hierarchy. In R. A. Rehfeldt & Y. Barnes-Holmes. (Eds.), *Derived relational responding: Applications for learners with autism and other developmental disabilities*. CA: New Harbinger.

Luciano, C., Gomez-Becerra, I., & Rodriguez-Valverde, M. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior*, *87*, 349-365. doi: 10.1901/jeab.2007.08-06.

Martin, G. L., & Yu, D. C. (2000). Overview of research on the Assessment of Basic Learning Abilities test. *Journal on Developmental Disabilities*. 7, 10-36.

Marion, C., Vause, T., Harapiak, S., Martin, G. L., Yu, C. T., Sakko, G., & Walters, K. (2003). The hierarchical relationship between several visual and auditory discriminations and three verbal operants among individuals with developmental disabilities. *The Analysis of Verbal Behavior, 19,* 91–105.

Martin, G.L., Yu, D., Quinn, G., & Patterson, S. (1983). Measurement and training of AVC discrimination skills: independent confirmation and extension. *Rehabilitation Psychology, 28*, 231-7. doi: 10.1037/h0090974

Mayer, G.R., Sulzer-Azaroff, B., & Wallace, M. (2011). *Behavior analysis for lasting change*. Cornwall-on-Hudson, NY: Sloan Educational Publishing.

MacCorquodale, K. (1970). On Chomsky's review of Skinner's Verbal behavior. *Journal of the Experimental Analysis of Behavior*, *13*, 83-99. doi: 10.1901/jeab.1970.13-83

McDonagh. E. C, McIlvane, W. J., & Stoddard, L T. (1984). Teaching coin equivalences via matching to sample. *Applied Research in Mental Retardation.* 5, 177-197. doi: 10.1016/S0270-3092(84)80001-6

McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation*, *97*, 359-359.

McGee, G. G., Krantz, P. J., & McClannahan, L. E. (1985). The facilitative effects of incidental teaching on preposition use by autistic children. *Journal of Applied Behavior Analysis*, *18*, 17-31. doi: 1985.18-17.

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *Psychological Record*, *54*, 115-144.

McIlvane, W. J., Dube, W. V., Kledaras, J. B., Iennaco, F. M., & Stoddard, L. T. (1990). Teaching relational discrimination to individuals with mental retardation: Some problems and possible solutions. *American Journal on Mental Retardation*, *95*, 283-296.

McIlvane, W. J. (2009). Translational behavior analysis: From laboratory science in stimulus control to intervention with persons with neurodevelopmental disabilities. *The Behavior Analyst*, *32*, 273-280.

Miguel, C. F., Yang, H. G., Finn, H. E., & Ahearn, W. H. (2009). Establishing derived textual control in activity schedules with children with autism. *Journal of Applied Behavior Analysis*, *42*, 703-9. doi: 10.1901/jaba.2009.42-703

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2010). Brief report: The training and assessment of relational precursors and abilities (TARPA): A preliminary analysis. *Journal of Autism and Developmental Disorders, 40*, 1149-53. doi: 10.1007/s10803-010-0968-0

Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2005). Derived manding in children with autism: Synthesizing Skinner's verbal behavior with relational frame theory. *Journal of Applied Behavior Analysis, 38*, 445-462. doi: 10.1901/jaba.2005.97-04

Murphy, C., & Barnes-Holmes, D. (2009a). Derived more-less relational mands in children diagnosed with autism. *Journal of Applied Behavior Analysis, 42*. doi: 10.1901/jaba.2009.42-253

Murphy, C. & Barnes-Holmes, D. (2009b). Establishing derived manding for specific amounts with three children: An attempt at synthesizing Skinner's Verbal Behavior with relational frame theory. *The Psychological Record, 59*, 75-92.

Murphy, C., & Barnes-Holmes, D. (2010). Establishing complex derived manding in children with and without a diagnosis of autism. *The Psychological Record, 60*, 489-504.

Noell, G. H., Connell, J. E., & Duhon, G. J. (2006). Spontaneous response generalization during whole word instruction: Reading to spell and spelling to read. *Journal of Behavioral Education, 15*, 121-130. doi: 10.1007/s10864-006-9016-8

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.

O'Donnell, J., & Saunders, K. J. (2003). Equivalence relations in individuals with language limitations and mental retardation. *Journal of the Experimental Analysis of Behavior, 80*, 131-157. doi: 10.1901/jeab.2003.80-131

Ogawa, A., Yamazaki, Y., Ueno, K., Cheng, K., & Iriki, A. (2010). Neural correlates of species-typical illogical cognitive bias in human inference. *Journal of Cognitive Neuroscience, 22*, 2120-2130. doi: 10.1162/jocn.2009.21330

O'Hora, D., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2004). Derived relational networks and control by novel instructions: A possible model of generative verbal responding. *The Psychological Record, 54*, 437-460.

O'Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the WAIS-III. *The Psychological Record*, *55*, 155-175.

O'Hora, D., Paláez, M., Barnes-Holmes, D. Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, *58*, 569-584.

Osborne, J. G., & Gatch, M. B. (1989). Stimulus equivalence and receptive reading by hearing-impaired preschool children. *Language, Speech, and Hearing Services in Schools*, *20*, 63-75.

O'Toole, C., & Barnes-Holmes, D. (2009). Three chronometric indices of relational responding as predictors of performance on a brief intelligence test: The importance of relational flexibility. *The Psychological Record*, *59*, 119-132.

Partington, J. W., Sundberg, M. L., Newhouse, L., & Spengler, S. M. (1994). Overcoming an autistic child's failure to acquire a tact repertoire. *Journal of Applied Behavior Analysis*, *27*, 733-734. doi: 10.1901/jaba.1994.27-733

Partington, J. W., & Sundberg, M. L. (1998). *The Assessment of Basic Language and Learning Skills (the ABLLS): Scoring instructions and IEP development guide: the ABLLS guide*. Pleasant Hill, CA: Behavior Analysts, Inc.

Peláez, M., Gewirtz, J. L., Sanchez, A., & Mahabir, N. M. (2000). Exploring stimulus equivalence formation in infants. *Behavioral Development Bulletin*, *9*, 23-28.

Peláez, M. (2009). Joint attention and social referencing in infancy as precursors of derived relational responding. In R.A. Rehfeldt & Y. Barnes-Holmes (Eds.), *Derived Relational Responding: Applications for learners with autism and other developmental disabilities.* Oakland: New Harbinger Publications, Inc.

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. doi: 10.1016/j.rasd.2011.12.007

Plavnick, J. B., & Ferreri, S. J. (2012). Collateral effects of mand training for children with autism. *Research in Autism Spectrum Disorders*, *6*, 1366-1376. doi: 10.1016/j.rasd.2012.05.008

Premack, D. (1976). *Intelligence in apes and man*. Hillsdale, NJ: Lawrence Erlbaum.

Reese, H. W. (1968). *The perception of stimulus relations: Discrimination learning and transposition.* New York: Academic Press.

Rehfeldt, R. A., & Root, S. L. (2005). Establishing derived requesting skills in adults with severe developmental disabilities. *Journal of Applied Behavior Analysis, 38*, 101-105. doi: 10.1901/jaba.2005.106-03

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M., & Kowalchuk, R. K. (2007). Assessing relational learning deficits in perspective-taking in children with high functioning autism spectrum disorder. *Psychological Record, 57*, 23-47.

Rehfeldt, R. A., & Barnes-Holmes, Y. (2009). *Derived relational responding: Applications for learners with autism and other developmental disabilities*. Oakland, CA: New Harbinger Publications, Inc.

Rehfeldt, R. A. (2011). Toward a technology of derived stimulus relations: An analysis of articles published in the journal of applied behavior analysis, 1992-2009. *Journal of Applied Behavior Analysis, 44*, 109-119. doi: 10.1901/jaba.2011.44-109

Richards, D. F., Williams, W. L., & Follette, W. C. (2002). Two new empirically derived reasons to use the Assessment of Basic Learning Abilities. *Journal Information, 107*, 329-339.

Roche, B., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I., & O'Hora, D. (2002). Relational Frame Theory: A new paradigm for the analysis of social behavior. *The Behavior Analyst, 25*, 75-91.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior*, *67*, 275-301. doi: 10.1901/jeab.1997.67-275

Roid, G. H. (2003). *Stanford Binet intelligence scales* (5th ed.). Itasca, IL: Riverside Publishing.

Rosales, R. & Rehfeldt, R. A. (2007). Contriving transitive conditioned establishing operations to establish derived manding skills in adults with severe developmental disabilities. *Journal of Applied Behavior Analysis, 40*, 105-121. doi: 10.1901/jaba.2007.117-05

Rosales, R., & Rehfeldt, R. A. (2009). Extending functional communication through relational framing. In R. A. Rehfeldt & Y. Barnes-Holmes (Eds.), *DRR applications for learners with autism and other developmental disabilities: A progressive guide to change* (pp. 237-256). Oakland, CA: New Harbinger.

Sakko, G., Martin, T. L., Vause, T., Martin, G. L., & Yu, C. T. (2004). Visual-visual non-identity matching assessment: A worthwhile addition to the assessment of basic learning abilities test. *American Journal on Mental Retardation, 109,* 45–52.

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence Class Establishment with Two-, Three-, and Four-Choice Matching to Sample by Senior Citizens. *Psychological Record, 55*, 539.

Schauffler, G., & Greer, R. D. (2006). The effects of intensive tact instruction on audience-accurate tacts and conversational units. *Journal of Early and Intensive Behavior Intervention*, *3*, 121-134.

Schlinger Jr, H. D. (2010). The long good-bye: Why BF Skinner's Verbal Behavior is alive and well on the 50th anniversary of its publication. *The Psychological Record*, *58*, 1.

Schreibman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders, 30*, 373-378.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (Zalophus californianus) is capable of forming equivalence relations. *Psychological Record*, *43*, 823-839.

Serna, R. W., Dube, W. V., & McIlvane, W. J. (1997). Assessing same/different judgments in individuals with severe intellectual disabilities: A status report. *Research in Developmental Disabilities*, *18*, 343-368.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5-13.

Sidman , M., & Cresson , O., Jr. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, *77*, 515–523.

Sidman, M., Cresson, O., & Willson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior*, *22*, 261-273. doi: 10.1901/jeab.1974.22-261

Sidman, M. (1986). The measurement of behavioral development. In N. A. Krasnegor, D. B. Gray, & T. Thompson (Eds.), *Developmental Behavioral Pharmacology, Vol 5*. Hillsdale, NJ: Lawrence Erlbaum.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22*, 11-18.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*, 127-146. doi: 10.1901/jeab.2000.74-127

Skinner, B. F. (1957). *Verbal behavior* . New York: Appleton-Century-Crofts.

Smeets, P. M., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes*, *69*(3), 281-293. doi: 10.1016/j.beproc.2004.12.009

Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, *105*, 269-285. doi: 10.1352/0895-8017(2000)105<0269:RTOIEI>2.0.CO;2

Stewart, I. Barnes-Holmes, D., & Roche, B. (2004). A functional analytic model of analogy using the Relational Evaluation Procedure. *The Psychological Record, 54* , 531-552.

Stewart, I., & McElwee, J. (2009). Relational responding and conditional discrimination procedures: An apparent inconsistency and clarification. *The Behavior Analyst*, *32*, 309-317.

Stewart, I., McElwee, J., & Ming, S. (2013). Language generativity, response generalization, and derived relational responding. *The Analysis of Verbal Behavior*, *29*, 137.

Stewart, I., Roche, B., O'Hora, D., & Tarbox, J. (2013). Education, intellectual development, and relational frame theory. In S. Dymond & B. Roche (Eds.), *Advances in Relational Frame Theory: Research & application.*. Oakland, CA: New Harbinger.

Sparrow S. S., Cicchetti D.V., & Balla, D. A. (2005). *Vineland II Adaptive Behavior Scales*. (2nd ed.) Circle Pines, MN: American Guidance Service, Inc.

Stewart, I., & Lavelle, N. (In press). Auditory stimulus equivalence and non-arbitrary relations. *The Psychological Record*.

Sundberg, M. L., & Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities*. Danville, CA: Behavior Analysts, Inc.

Sundberg, M. L., Endicott, K., & Eigenheer, P. (2000). Using intraverbal prompts to establish tacts for children with autism. *The Analysis of Verbal Behavior*, *17*, 89-104.

Sundberg, M. L., & Michael, J. (2001). The benefits of Skinner's analysis of verbal behavior for children with autism. *Behavior Modification, 25*, 698-724. doi: 10.1177/0145445501255003

Sundberg, M. L. (2008). *VB-MAPP: Verbal Behavior Milestones Assessment and Placement Program - protocol*. Concord, CA: AVB Press.

Vause, T., Martin, G. L., Yu, C. T., Marion, C., & Sakko, G. (2005). Teaching equivalence relations to individuals with minimal verbal repertoires: Are visual and auditory-visual discriminations predictive of stimulus equivalence? *The Psychological Record*, *55*, 197-218.

Ward, R., & Yu, D. C. (2000). Bridging the gap between visual and auditory discrimination learning in children with autism and severe developmental disabilities. *Journal on Developmental Disabilities*. 7, 142-155.

Wacker, D. P. (1996). Behavior analysis research in JABA: A need for studies that bridge basic and applied research. *Experimental Analysis of Human Behavior Bulletin, 14*, 11–14.

Wacker, D. P. (2003). Bridge studies in behavior analysis: Evolution and challenges in JABA. *The Behavior Analyst Today, 3*, 405–411.

Wechsler, D. (1997). Manual for the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III). San Antonio, TX: The Psychological Corporation.

Wesolowski, M. D., Zencius, A. H., McCarthy-Lydon, D., & Lydon, S. (2005). Using behavioral interventions to treat speech disorders in persons with head trauma. *Behavioral Interventions, 20*, 67-75. doi: 10.1002/bin.169

Whalen, C., Liden, L., Ingersoll, B., Dallaire, E., & Liden, S. (2006). Behavioral improvements associated with computer-assisted instruction for children with developmental disabilities. *The Journal of Speech and Language Pathology-Applied Behavior Analysis*, *1*, 11-26.

Whelan, R., & Schlund, M. W. (2013). Reframing Relational Frame Theory research: Gaining a new perspective through the application of novel behavioral and neurophysiological methods. In S. Dymond & B. Roche (Eds.), *Advances in relational frame theory: Research and applications* (pp. 73-96). Oakland, CA: New Harbinger Publications.

Williams, W. L., & Jackson, M. L. (2009). The Assessment of Basic Learning Abilities (ABLA) and its relation to the development of stimulus relations in persons with autism and other intellectual disabilities. In R.A. Rehfeldt & Y. Barnes-Holmes (Eds.), *DRR: Applications for learners with autism and other developmental disabilities* (pp. 25-39). Oakland: New Harbinger Publications.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of behavior*, *50*, 125-144.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scale (4th edition)*. San Antonio, TX: Harcourt Assessment.

Appendix A

TARPA structure and stimuli

TARPA 4.3: Structure

Stage 1: Basic Discrimination

Track: Visual

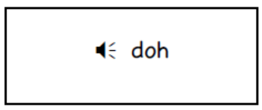
Level 1: Blank vs Visual

Level 2: 2 Visual Comparisons

Level 3: 3 Visual Comparisons

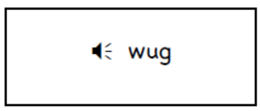
Track: Auditory

Level 1: Blank vs Auditory



Level 2: 2 Auditory Comparisons

Level 3: 3 Auditory Comparisons



Stage 2: Conditional Discrimination (Formally SIMILAR Stimuli)

Track: Visual

Level 1: 2 Visual Comparisons

Level 2: 3 Visual Comparisons

Level 3: 3 Visual Identity Matching

Track: Auditory

Level 1: 2 Auditory Comparisons

Level 2: 3 Auditory Comparisons

Level 3: 3 Auditory Identity Matching

Stage 3: Conditional Discrimination (Formally DISSIMILAR Stimuli)

Tracks: Visual-Visual; Auditory-Visual 1; Auditory-Visual 2; Auditory-Auditory.

Level 1: Conditional Discrimination

Level 2: Mutual Entailment

Section 1: Training A→B

Section 2: Maintenance A→B

Section 3: Derivation (Test B→A)

Level 3: Combinatorial Entailment

Section 1: Training A→B

Section 2: Maintenance A→B

Section 3: Training C→B

Section 4: Maintenance C→B

Section 5: Maintenance Mixed A→B, C→B

Section 6: Derivation (Test C→A, A→C)

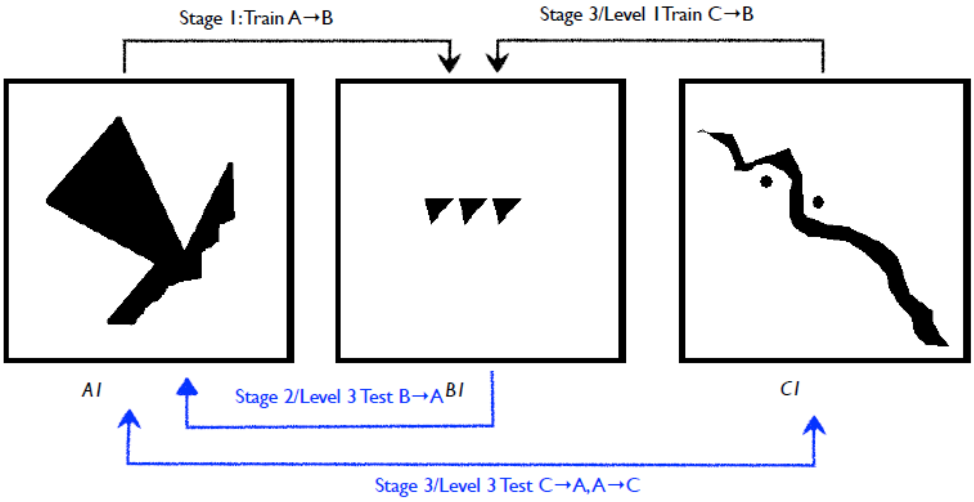
Level 4: Transformation of Function

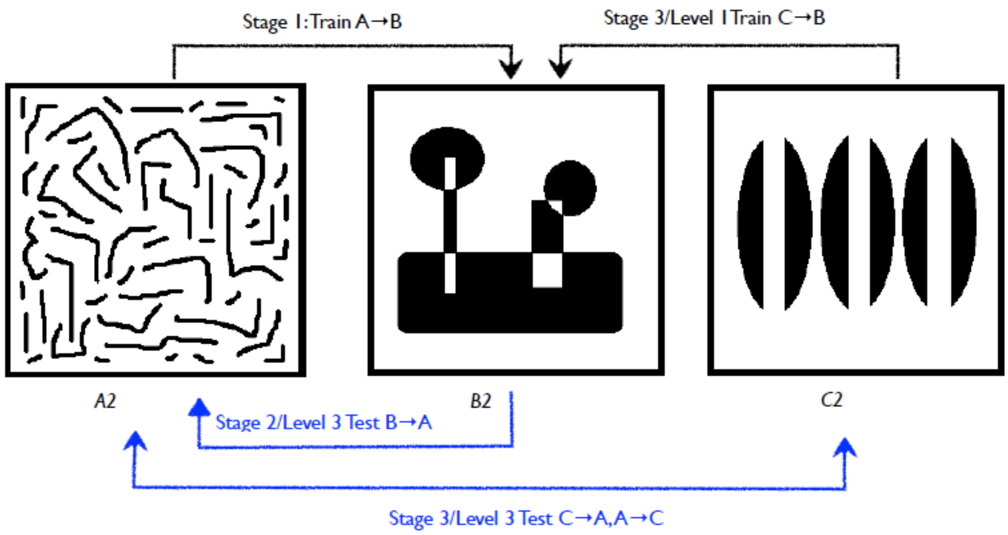
Section 0: CE Maintenance

Section 1: Task Training

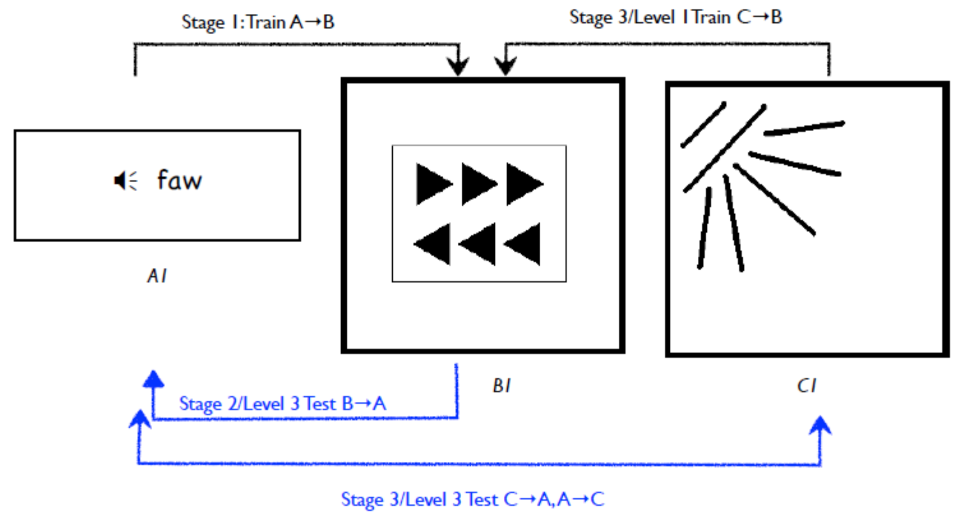
Section 2: Transformation of Function Testing

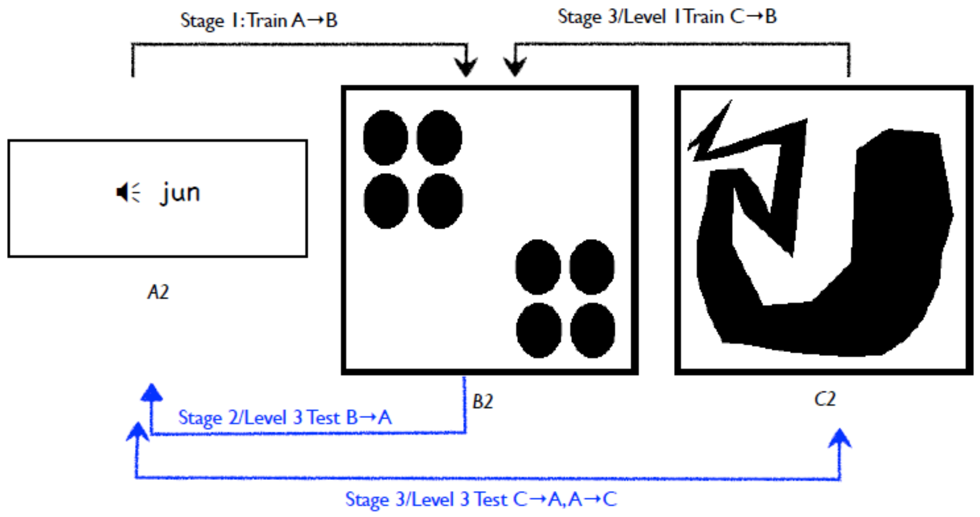
Visual-Visual Track



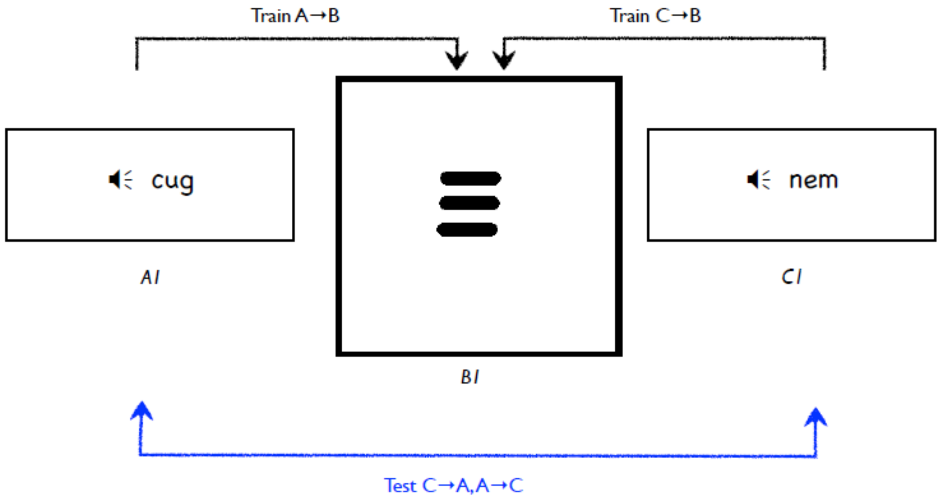


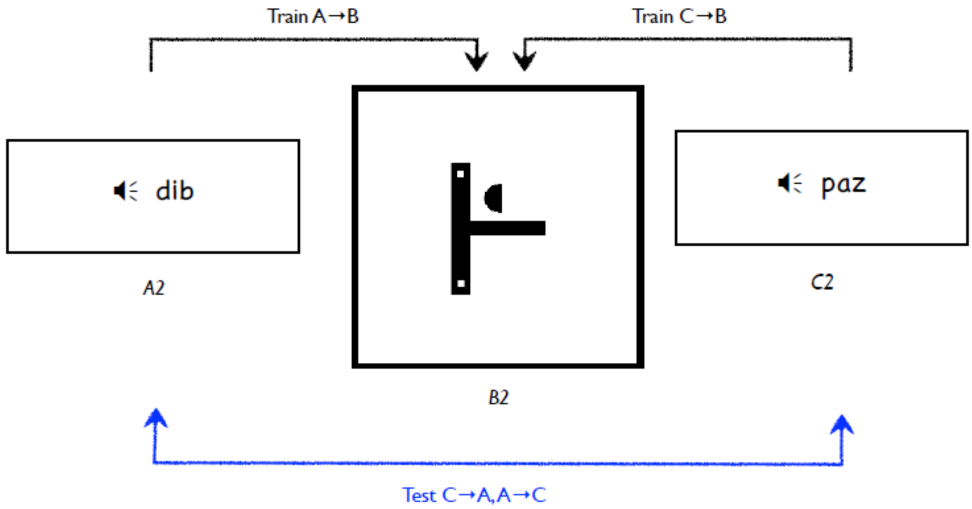
Auditory-Visual 1



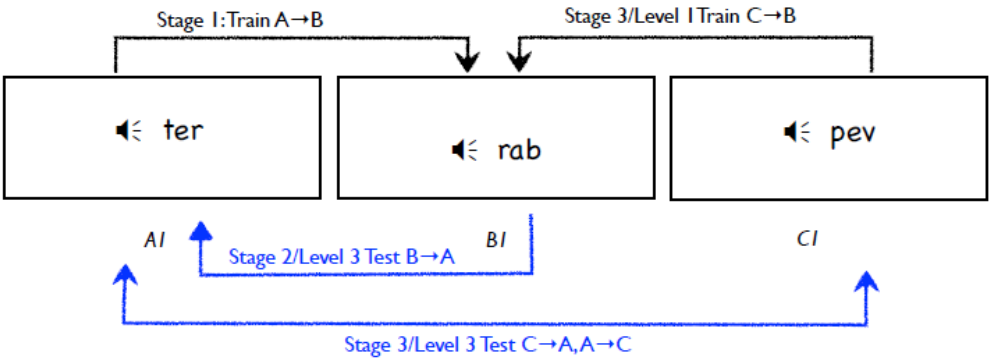


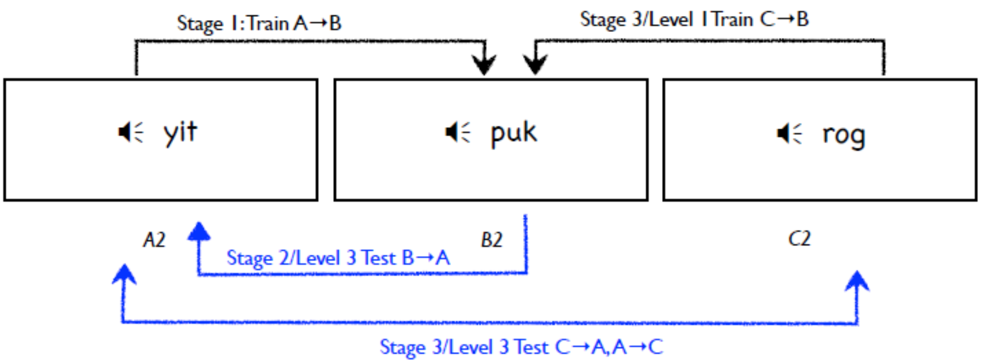
Auditory-Visual 2





Auditory-Auditory





Appendix B

TARPA Scoring System.

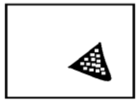
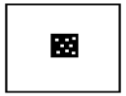
Stage	Track	Level	Points	Pass Criterion	Fail Criterion
1	V	1 (1 stimulus)	0.33	5/6 CC	6 CI
	V	2 (2 stimuli)	0.33	5/6 CC	6 CI
	V	3 (3 stimuli)	0.33	8/9 CC	9 CI
	A	1 (1 stimulus)	0.33	5/6 CC	6 CI
	A	2 (2 stimuli)	0.33	5/6 CC	6 CI
	A	3 (3 stimuli)	0.33	8/9 CC	9 CI
2	V	1 (2 comps)	0.5	5/6 CC	6 CI
	V	2 (3 comps)	0.5	5/6 CC	6 CI
	V	3 (GID)	1	5/5 CC	6 CI
	A	1 (2 comps)	0.5	5/6 CC	6 CI
	A	2 (3 comps)	0.5	5/6 CC	6 CI
	A	3 (GID)	1	5/5 CC	6 CI
3	VV	CD	3	5/6 CC	6 CI
	VV	ME	4	7/8 Correct	-
	VV	CE	4	7/8 Correct	-
	VV	TOF	3	7/8 Correct	-
	AV1	CD	3	5/6 CC	6 CI
	AV 1	ME	4	7/8 Correct	-
	AV 1	CE	4	7/8 Correct	-
	AV 1	TOF	3	7/8 Correct	-
	AV 2	CD	3	5/6 CC	6 CI
	AV 2	ME	4	7/8 Correct	-
	AV 2	CE	4	7/8 Correct	-
	AV 2	TOF	3	7/8 Correct	-
	AA	CD	3	5/6 CC	6 CI
	AA	ME	4	7/8 Correct	-
	AA	CE	4	7/8 Correct	-
	AA	TOF	3	7/8 Correct	-

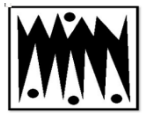
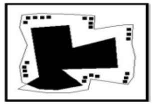
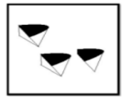
Note. GID = Generalized Identity; CD = Conditional Discrimination; ME = Mutual Entailment; CE = Combinatorial Entailment; TOF = Transfer of Function; V = Visual; A = Auditory; VV = Visual-Visual; AV 1 = Auditory-Visual 1; AV2 = Auditory-Visual 2; AA = Auditory-Auditory; CC = Consecutively Correct; CI = Cumulatively Incorrect; comps = comparisons.

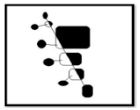
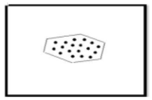
Appendix C

Test Retest Stimuli

Stage 3

Visual –Visual					
					
A1	B1	C1	A2	B2	C2

Auditory-Visual 1					
CAX			**CEV**		
A1	B1	C1	A2	B2	C2

Auditory-Visual 2					
HAN		**SAG**	**GIM**		**JIP**
A1	B1	C1	A2	B2	C2

Auditory-Auditory					
TAM	**TUD**	**DUB**	**VUZ**	**SEK**	**FUM**
A1	B1	C1	A2	B2	C2

Appendix D

Chapter 4/Experiment 4

Table 33.

School	Participants
Bofield National School, Attymass, Ballina, Co. Mayo.	P8; P9; P10; P11; P13; P14; P19
Errew National School, Errew, Castlebar, Co.Mayo.	P5; P16; P16; P7; P17; P18
Glenhest National School, Glenhest, Co. Mayo	P1; P2 & P3.
Educate Together National School, Newcastle, Galway city.	P15; P20; P21; P22; P23;P24; P25 & P26
Acquaintances of the experimenter	P4 & P12

Appendix E

Chapter 5/Experiment 5

Participant, sex, age and educational setting.

Pt (Sex)	Age	Placement	Pt (Sex)	Age	Placement
1 (m)	04:02	ABA Home programme	11 (m)	05:09	ASD special classroom
2 (f)	07:02	ASD special classroom	12 (m)	07:01	ASD special classroom
3 (m)	07:07	ASD special classroom	13 (m)	06:10	ASD special classroom
4 (m)	06:08	ASD special classroom	14 (m)	04:08	ASD special classroom
5 (m)	06:03	ASD special classroom	15 (m)	09:04	ASD special classroom
6 (m)	08:11	ASD special classroom	16 (m)	07:10	ASD special classroom
7 (m)	06:01	Special school	17 (m)	05:09	ASD special classroom
8 (m)	15:00	Special school	18 (f)	09:07	ASD special classroom
9 (f)	05:05	ABA Home programme	19 (m)	08:05	ASD special classroom
10 (m)	10:02	ASD special classroom	20 (m)	03:08	ABA Home programme

Participant, sex, chronological age and scores on the VABS (standard and raw scores and overall composite score).

Pt.(Sex)	**Age**	**VABS Scores**			
		Communication SS	**Daily Skills SS**	**Socialisation SS**	**Composite**
1 (m)	04:02	57	71	61	64
2 (f)	07:02	75	76	73	73
3 (m)	07:07	75	73	61	69
4 (m)	06:08	MD	MD	MD	MD
5 (m)	06:03	81	85	83	78
6 (m)	08:11	70	74	75	72
7 (m)	06:01	76	64	63	66
8 (m)	15:00	35	45	34	36
9 (f)	05:05	54	58	59	59
10 (m)	10:02	MD	59	62	MD
11 (m)	05:09	79	93	97	85
12 (m)	07:01	96	89	75	85
13 (m)	06:10	78	66	90	74
14 (m)	04:08	83	91	100	82
15 (m)	09:04	72	79	73	81
16 (m)	07:10	77	69	75	72
17 (m)	05:09	59	60	57	55
18(f)	09:07	72	69	57	66
19 (m)	08:05	79	79	69	74
20 (m)	03:08	78	71	66	69

Appendix F

Track difficulty across Experiments 1-4

Experiment 1

ASD. Differences between tracks is examined by comparing aggregate performance on each of the four tracks (i) in terms of the number of levels passed and (ii) in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function)

Stage 3 – Differences between tracks				
	VV	AV1	AV2	AA
Total No. of Levels passed	16	9	12	9
No. of pts. who showed all patterns of DRR within track	3	1	1	1

Experiment 2.

TDC. Differences between tracks is examined by comparing aggregate performance on each of the four tracks (i) in terms of the number of levels passed and (ii) in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function)

Stage 3 – Differences between tracks				
	VV	AV1	AV2	AA
Total No. of Levels passed	26	23	25	16
No. of pts. who showed all patterns of DRR within track	3	3	3	0

Experiment 3

TDC. Differences between tracks is examined by comparing aggregate performance on each of the four tracks (i) in terms of the number of levels passed and (ii) in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function)

Stage 3 – Differences between tracks				
	VV	AV1	AV2	AA
Total No. of Levels passed	19	28	17	10
No. of pts. who showed all patterns of DRR within track	2	5	0	0

Experiments 4A and 4B

TDC. Differences between tracks is examined by comparing aggregate performance on each of the four tracks (i) in terms of the number of levels passed and (ii) in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function)

Stage 3 – Differences between tracks				
	VV	AV1	AV2	AA
Total No. of Levels passed	53	41	41	31
No. of pts. who showed all patterns of DRR within track	9	3	4	3

Experiments 2, 3 & 4A/4B

TDC. Differences between tracks is examined by comparing aggregate performance on each of the four tracks (i) in terms of the number of levels passed and (ii) in terms of the number of participants who showed all three properties of DRR (i.e., mutual entailment, combinatorial entailment and transformation of function)

Stage 3 – Differences between tracks				
	VV	AV1	AV2	AA
Total No. of Levels passed	98	92	83	57
No. of pts. who showed all patterns of DRR within track	14	9	7	3